Roberto Tuberosa · Andreas Graner
Emile Frison *Editors*

Genomics of Plant Genetic Resources

Volume 2. Crop Productivity, Food Security and Nutritional Quality

Springer

Genomics of Plant Genetic Resources

Roberto Tuberosa • Andreas Graner • Emile Frison
Editors

Genomics of Plant Genetic Resources

Volume 2. Crop Productivity, Food Security and Nutritional Quality

Springer

Editors
Roberto Tuberosa
Agricultural Sciences
University of Bologna
Bologna
Italy

Andreas Graner
Genebank
IPK, Gatersleben
Sachsen-Anhalt
Germany

Emile Frison
Bioversity International
Rome
Italy

ISBN 978-94-007-7574-9 ISBN 978-94-007-7575-6 (eBook)
DOI 10.1007/978-94-007-7575-6
Springer Dordrecht Heidelberg New York London

Library of Congress Control Number: 2013954265

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Foreword

Plant genetic resources constitute the feedstock for the biotechnology and genetic engineering enterprises. Year 2013 marks the 60th anniversary of the discovery of the double helix structure of the DNA molecule. This discovery led to the birth of the new genetics based on genomics. The new genetics is helping to revolutionize plant breeding through both marker-assisted selection and recombinant DNA technology. It is in this context that this informative two-volume book entitled "Genomics of Plant Genetic Resources" edited by Prof. Roberto Tuberosa, Prof. Andrea Graner and Dr. Emile Frison is very timely and welcome.

The book deals with managing plant genetic resources, developing genomics platforms and approaches to investigate plant genetic resources, genome sequencing and crop domestication and mining allelic diversity. The different chapters written by eminent authorities shed much light on problems relating to both theoretical and applied genomics. We owe a deep debt of gratitude to the Editors for this labor of love in the cause of conservation and sustainable use of plant genetic resources. This book shows the pathway for achieving an ever-green revolution in agriculture based on enhancement of productivity in perpetuity without associated ecological harm.

M. S. Swaminathan

Foreword

Who would have believed only two decades ago that plant scientists would have access to nearly the complete genetic code of numerous plant species, including major crop species. The idea of having ready access to whole-genome sequences encompassing 140 million bases seemed like science fiction, let alone having available even larger genomes such as rice at 430 Mb or maize at 2500 Mb. And then proceeding to identify variation at the DNA level well beyond what was anticipated, such as the 2.6 SNPs (Single Nucleotide Polymorphisms) per kb in rice. Also produced at an unprecedented rate were literally hundreds of thousands of insert strains, allowing the association of sequences and traits. Who would have believed only a decade ago that we would be capable of analyzing the expression of genes across the whole genome and matching that profile with traits of interest. And now the area of metabolomics is allowing even more meaningful explanations of the biochemical and genetic pathways underlying important traits.

This book brings all of these advances in genomics to the forefront and prepares the plant scientist for the decade ahead. Important technologies are discussed such as association mapping, simulation modeling, and development of appropriate populations including advanced backcrosses and introgression-lines for incorporating traits into useful genetic materials. Such approaches are facilitating the identification of traits that are not obvious simply from observing the plant phenotype, and they provide ways to extract new and useful traits from wild related species. Comparing the genomic information across broadly-related species has generated important evolutionary information. In addition, the common occurrence of duplicated segments recognized in such studies may lead to information fundamental to plant performance.

Methods for the identification of genes underlying traits are improving every day. The association between allelic variation in a candidate gene and a trait is leading to a much greater understanding of the genetic control of traits. Numerous transcription factors and even non-coding sequences are being implicated as the basis of important genetic variation. Forward and reverse genetics are both found to be very useful in making these gene-trait associations.

The tremendous expansion of genomic analytical approaches along with efforts to reduce the cost, together with appropriate statistical designs and analyses, is making it easier and easier to use the ever-increasing sequence information to identify useful genes and gene families. This body of knowledge in plant genomics and its myriad of applications are nicely reflected in this book.

Ronald L. Phillips

Preface

This two-volume book collects 48 manuscripts that present a timely state-of-the-art view on how genomics of plant genetic resources contributes to improve our capacity to characterize and harness natural and artificially induced variation in order to select better cultivars while providing consumers with high-quality and nutritious food. In the past decade, the appreciation of the value of biodiversity has grown steadily, mainly due to the increased awareness of the pivotal role of plant genetic resources for securing the future supply of plant-derived products in the quantity required to meet the burgeoning needs of mankind. The remarkable progress made possible with the deployment of genomics and sequencing platforms has considerably accelerated the pace of gene discovery, the identification of novel, valuable alleles at target loci and their exploitation in breeding programs via marker-assisted selection or other molecular means. Clearly, a better understanding of the genetic make-up and functional variability underpinning the productivity of crops and their adaptation to abiotic and biotic constraints offers unprecedented opportunities for highly targeted approaches while shedding light on the molecular functions that govern such variability.

Meeting the challenges posed by climate change and the future needs of mankind for plant-derived products will require a quantum leap in productivity of the handful of species that provide the staple for our diet and existence. This quantum leap will only be possible through a more effective integration of genomics research with extant breeding programs. As we anticipate a further reduction in the cost of genotyping/sequencing, the exploitation of still largely untapped samples of wild germplasm stored in gene banks will become instrumental for the success of breeding programs. Importantly, the new selection paradigm ushered in by genomics greatly facilitates mining the genetic richness present in orphan crops and underutilized species, previously less readily accessible via conventional approaches.

The unifying picture that emerges from this book unequivocally shows the pivotal role played by genomics to characterize germplasm collections, mine genebanks, elucidate gene function, identify agronomically superior alleles and, ultimately, release improved cultivars. For each of these objectives, the book presents compelling case studies and examples; additional case studies are provided by the references of each chapter.

We hope that this book will provide a helpful reference to students, young researchers, crop specialists and breeders interested in a more effective characterization and utilization of plant genetic resources. In particular, we hope that reading of this book will encourage students and young scientists to pursue a career focused on the study of plant genetic resources and join forces with those already engaged in this challenging and equally fascinating field of science.

We wish to thank all the authors for their timely contributions that have made this book possible. We also thank all those who have contributed to the editing of this book. Last but not least, we wish to thank the policy makers and funding agencies that provide the funds required to collect, conserve, characterize and harness the allelic richness of plant genetic resources.

Roberto Tuberosa
Andreas Graner
Emile Frison

Contents

Contributors

Meike S. Andersson HarvestPlus, c/o CIAT, Cali, Colombia

Raman Babu International Maize and Wheat Improvement Center (CIMMYT), Mexico D.F., Mexico

P. Stephen Baenziger Formerly at Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, NE, USA

Ali Bakhsh College of Agriculture Dera Ghazi Khan Sub-campus University of Agriculture Faisalabad, Pakistan Formerly at Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, NE, USA

Elsa Ballini UMR BGPI INRA/CIRAD/SupAgro, 34398 Montpellier, France

Tobias Baum Fraunhofer-Institute for Factory Operation and Automation (IFF), Biosystems Engineering, Magdeburg, Germany

E. Botticella Department of Agriculture, Forests, Nature and Energy, University of Tuscia, Viterbo, Italy

Carla Ceoloni Department of Agriculture, Forestry, Nature and Energy (DAFNE), University of Tuscia, Viterbo, Italy

David Chagné The New Zealand Institute for Plant & Food Research Limited, Palmerston North, New Zealand

Ronnie Coffman College of Agriculture and Life Sciences, Cornell University, Ithaca, NY, USA

Dimitar Douchkov Leibniz-Institute of Plant Genetics and Crop Plant Research (IPK)

Sarah Davidson Evanega College of Agriculture and Life Sciences, Cornell University, Ithaca, NY, USA

Paola Forte Department of Agriculture, Forestry, Nature and Energy (DAFNE), University of Tuscia, Viterbo, Italy

E. Francia Department of Life Sciences, University of Modena and Reggio Emilia, Reggio Emilia, Italy

Jackie Freeman Rothamsted Research, Harpenden, Hertfordshire, UK

G. Galiba Agricultural Institute, Centre for Agricultural Research, Hungarian Academy of Sciences, Martonvásár, Hungary

Research Institute of Chemical and Process Engineering, Faculty of Information Technology, University of Pannonia, Veszprém, Hungary

Susan Gardiner The New Zealand Institute for Plant & Food Research Limited, Palmerston North, New Zealand

Christiane Gebhardt MPI for Plant Breeding Research, Germany

Kurt Gebruers Laboratory of Food Chemistry and Biochemistry, and Leuven Food Science and Nutrition Research Centre (LFoRCe), Catholic University Leuven, Leuven, Belgium

Andrea Gennaro Department of Agriculture, Forestry, Nature and Energy (DAFNE), University of Tuscia, Viterbo, Italy

Present address: GMO Unit, European Food Safety Authority (EFSA)

Debora Giorgi Department of Agriculture, Forestry, Nature and Energy (DAFNE), University of Tuscia, Viterbo, Italy

Present address: Italian National Agency for New Technologies, Energy and the Environment (ENEA), Casaccia Research Center, Rome, Italy

Silvana Grandillo Italian National Research Council, Institute of Biosciences and BioResources (CNR-IBBR), Research Division Portici, Portici, Naples, Italy

Maria Rosaria Grossi Department of Agriculture, Forestry, Nature and Energy (DAFNE), University of Tuscia, Viterbo, Italy

Z. Gulyás Agricultural Institute, Centre for Agricultural Research, Hungarian Academy of Sciences, Martonvásár, Hungary

Perry Gustafson Division of Plant Sciences, University of Missouri, Columbia, Missouri, USA

Owen A. Hoekenga RobertW. Holley Center for Agriculture and Health, United States Department of Agriculture, Agricultural Research Service, Ithaca, NY 14853, USA

Maria Hrmova Australian Centre for Plant Functional Genomics, University of Adelaide, Glen Osmond, SA, Australia

Alexander Ihlow Ilmenau University of Technology, Institute for Information Technology, Ilmenau, Germany

Kwang-Ryong Jo Wageningen UR Plant Breeding,Wageningen University & Research Centre, Wageningen, The Netherlands

Research Institute of Agrobiology, Academy of Agricultural Sciences, Pyongyang, Democratic People's Republic of Korea

Christian Jung Plant Breeding Institute, Christian-Albrechts-University of Kiel, Kiel, Germany

Alexander L. Kahler Department of Agronomy and Plant Genetics, University of Minnesota, Minneapolis, MN, USA

Beat Keller Institute of Plant Biology, University of Zurich, Zurich, Switzerland

Anthony J. Kern Department of Agronomy and Plant Genetics, University of Minnesota, Minneapolis, MN, USA

G. Kocsy Agricultural Institute, Centre for Agricultural Research, Hungarian Academy of Sciences, Martonvásár, Hungary

K. Kosová Department of Genetics and Plant Breeding, Crop Research Institute, Prague, the Czech Republic

Satish Kumar The New Zealand Institute for Plant & Food Research Limited, Havelock North, New Zealand

Ljiljana Kuzmanović Department of Agriculture, Forestry, Nature and Energy (DAFNE), University of Tuscia, Viterbo, Italy

D. Lafiandra Department of Agriculture, Forests, Nature and Energy, University of Tuscia, Viterbo, Italy

Sergiy Lopato Australian Centre for Plant Functional Genomics, University of Adelaide, Glen Osmond, SA, Australia

Aaron Lorenz Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, NE, USA

Siegbert Melzer Plant Breeding Institute, Christian-Albrechts-University of Kiel, Kiel, Germany

J.A. Milc Department of Life Sciences, University of Modena and Reggio Emilia, Reggio Emilia, Italy

Rowan A.C. Mitchell Rothamsted Research, Harpenden, Hertfordshire , UK

Jean-Benoit Morel UMR BGPI INRA/CIRAD/SupAgro, 34398 Montpellier, France

Andreas E. Müller Plant Breeding Institute, Christian-Albrechts-University of Kiel, Kiel, Germany

Strube Research GmbH & Co. KG, Hauptstraße 1, 38387 Söllingen, Germany

Csilla Nemeth Rothamsted Research, Harpenden, Hertfordshire , UK

School of Biosciences, University of Nottingham, Sutton Bonnington Campus, Sutton Bonnington, UK

N. Pecchioni Department of Life Sciences, University of Modena and Reggio Emilia, Reggio Emilia, Italy

Wolfgang H. Pfeiffer HarvestPlus, c/o CIAT, Cali, Colombia

Ronald L. Phillips Department of Agronomy and Plant Genetics, University of Minnesota, Minneapolis, MN, USA

Raymond A. Porter Department of Agronomy and Plant Genetics, University of Minnesota, Minneapolis, MN, USA

B. M. Prasanna CIMMYT, ICRAF House, Nairobi, Kenya

I.T. Prášil Department of Genetics and Plant Breeding, Crop Research Institute, Prague, the Czech Republic

Michael O. Pumphrey Washington State University, Pullman,WA, USA

Harsh Raman Graham Centre for Agricultural Innovation (an alliance between Charles Sturt University and NSW Department of Department of Primary Industries), WaggaWagga Agricultural Institute,WaggaWagga, NSW Australia

Luc Saulnier INRA UR1268 Biopolymers, Interactions Assemblies, Nantes, France

Patrick Schweizer Leibniz-Institute of Plant Genetics and Crop Plant Research (IPK)

Udo Seiffert Fraunhofer-Institute for Factory Operation and Automation (IFF), Biosystems Engineering, Magdeburg, Germany

Liselotte L. Selter Institute of Plant Biology, University of Zurich, Zurich, Switzerland

F. Sestili Department of Agriculture, Forests, Nature and Energy, University of Tuscia, Viterbo, Italy

Margarita Shatalina Institute of Plant Biology, University of Zurich, Zurich, Switzerland

Peter R. Shewry Rothamsted Research, Harpenden, Hertfordshire, UK

Ravi P. Singh International Maize and Wheat Improvement Center (CIMMYT), Mexico, DF 06600, USA

Jyoti Singla Institute of Plant Biology, University of Zurich, Zurich, Switzerland

Benjamin Stich MPI for Plant Breeding Research, Germany

Joe Tohme CIAT, Cali, Colombia

Yusaku Uga National Institute of Agrobiological Sciences, Tsukuba, Ibaraki, Japan

Claude Urbany MPI for Plant Breeding Research, Germany

P. Vítámvás Department of Genetics and Plant Breeding, Crop Research Institute, Prague, the Czech Republic

Richard K. Volz The New Zealand Institute for Plant & Food Research Limited, Havelock North, New Zealand

Ben Vosman Wageningen UR Plant Breeding,Wageningen University & Research Centre, Wageningen, The Netherlands

Jack H. Vossen Wageningen UR Plant Breeding,Wageningen University & Research Centre, Wageningen, The Netherlands

Harkamal Walia Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, NE, USA

Jane L.Ward Rothamsted Research, Harpenden, Hertfordshire, UK

Toshio Yamamoto National Institute of Agrobiological Sciences, Tsukuba, Ibaraki, Japan

Masahiro Yano National Institute of Agrobiological Sciences, Tsukuba, Ibaraki, Japan

Part I
Harnessing Plant Genetic Diversity for Enhancing Crop Production and Its Sustainability

Chapter 1
Genetics and Genomics of Flowering Time Regulation in Sugar Beet

Siegbert Melzer, Andreas E. Müller and Christian Jung

Abstract Leaves from beets have been used since ancient times for nutrition and the swollen roots were one of the first sweeteners in the Middle Ages that could be stored through the winter. Breeding of beets to increase sugar content began only in the 18th century, after it was uncovered that the nature of the sweet taste of sugar cane and that of beet roots relies on the same sugar molecule. The major breakthrough in breeding sugar beet was the selection of beet progenies that, unlike their wild ancestors, did not flower in the first year of growth, correlating to a high root and thus sugar yield. This was the birth of the sugar beet that became a major crop in Europe and later on worldwide. Genetics has shown that the switch from annual to biennial beets relies mainly on one gene: the 'bolting gene' *B*. However, research from model plants has shown that the regulation of flowering is complex and involves many regulatory pathways, which perceive, transduce and integrate both endogenous and environmental cues for the fine tuning of flowering. Therefore, broad approaches to study flowering time in beets have been initiated, including both forward and reverse genetic studies to elucidate the molecular nature of *B* as well as other components of what is likely an intricate regulatory network also in beet. This chapter will give a short history of beet use and breeding as well as strategies and results from recent and current efforts to understand the regulation of flowering time in sugar beet.

1.1 The Sugar Beet Crop and Its Cultivated and Wild Relatives

Sugar beet (*B. vulgaris* L. ssp. *vulgaris* Sugar Beet Group) is the only sucrose-storing species of moderate climates. It belongs to the genus *Beta* that is now grouped in the Amaranthaceae (formerly Chenopodiaceae) subfamily Chenopodiaceae (The Angiosperm Phylogeny Group 2009).

C. Jung (✉) · S. Melzer · A. E. Müller
Plant Breeding Institute, Christian-Albrechts-University of Kiel,
Olshausenstr. 40, 24098 Kiel, Germany
e-mail: c.jung@plantbreeding.uni-kiel.de

A. E. Müller
Strube Research GmbH & Co. KG, Hauptstraße 1, 38387 Söllingen, Germany

R. Tuberosa et al. (eds.), *Genomics of Plant Genetic Resources,*
DOI 10.1007/978-94-007-7575-6_1,

Species of the genus *Beta* show great morphological variation of leaves and roots, including colors that vary from red to yellow due to the production of two betalains, betacyanine (red-violett) or betaxanthine (yellow). Domesticated forms of *B. vulgaris* have been used since antiquity. Leaf beets (syn. mangold, swiss chard) (*B. vulgaris* L. ssp. *vulgaris* Leaf Beet Group) form fleshy leaves and have a long tradition as a vegetable. Red table beets (beetroot) (*B. vulgaris* L. ssp. *vulgaris* Garden Beet Group) form a thickened root and hypocotyl with an intense dark red color. They have a high content of free folic acid (vitamin B12), and are used as a vegetable and for production of natural colors for food additives. Fodder beets (*B. vulgaris* L. ssp. *vulgaris* Fodder Beet Group) form a thickened root and hypocotyl and are traditionally used as animal feed, mainly for dairy cattle.

Sugar beets form a thickened root, which like other cultivated forms store substantial amounts of sucrose. While sucrose concentration in fodder beets ranges between 4 and 10 %, sugar beet roots may contain more than 20 % sucrose. Under central European growing conditions the sucrose concentration typically is 17–18 % (Biancardi et al. 2005).

The development of sugar beet as a cultivated species began in 1747 when the German chemist A. S. Marggraf detected cane sugar within the roots of garden beets, frequently grown as a vegetable at that time. Cultivation started on a very small scale at the end of the 18th century in Germany when F. C. Achard grew 'sugar beet' near Berlin. The first beet root processing sugar factory was constructed in 1801 in Silesia, a Prussian province at that time. That year is regarded as the beginning of sugar beet cultivation.

The sucrose content of beet roots at the beginning of beet cultivation was estimated to be around 4 %. By mass selection, sucrose content was raised to 16 % by the end of the 19th century. At that time, the market share of sugar beet sucrose was 62 %, the rest coming from sugar cane. After World War I, it dropped dramatically to 23 %. In 2010, the world beet harvest reached 227 Mt fresh weight giving rise to 32.3 Mt of sugar that is 19 % of the total world production (166.8 Mt of sugar (http://www.zuckerverbaende.de/zuckermarkt/zahlen-und-fakten/weltzuckermarkt/erzeugung-verbrauch.html). Sugar beets are grown in many countries of the Northern hemisphere. The total sugar beet area harvested was 4.3 Mha (2009). Major producers were the Russian Federation (770,000 ha), the United States of America (465,000 ha), Germany (384,000 ha), and France (374,000 ha) http://faostat3.fao.org/faostat-gateway/go/to/download/Q/QC/E

The species of the genus *Beta* are divided into two sections (Kadereit et al. 2006) (Fig. 1.1). All cultivated forms belong to the same subspecies *B. vulgaris* ssp. *vulgaris*. Together with their wild progenitor *B. vulgaris* ssp. *maritima* (L.) Arcang, they belong to section I (Beta). *B. vulgaris* ssp. *maritima* and the related wild species ssp. *adanensis* (*Pamuk*) *Ford-Lloyd & Will., B. patula* Ait. and *B. macrocarpa* Guss. grow around the Mediterranean and the coasts of northwest Europe up to Scandinavia and between the Capverdian Islands and Bangladesh. There are only wild species in the section II (Corollinae). Apart from diploid species, tetraploids, pentaploids and hexaploids also exist within this section. Those species grow on the hilly and mountainous regions in Turkey and adjacent countries (Fig. 1.1).

Sections of the genus *Beta*	Chromosome numbers	Life history	Natural habitats
I. SectioBeta:			from southwest Norway to Capverdian Islands, from Bangla Desh to Canary Islands
B. vulgaris L.			
ssp. *vulgaris*	18	biennials (cultivated beet)	
• Cultivar group Leaf beet			
• Cultivar group Gardenbeet			
• Cultivar group Fodder beet			
• Cultivar group Sugar beet			
ssp.*maritima* (L.) Arcang.	18	annuals, biennials, iteroparous perennials	
ssp. *adanensis* (Pamukc. ex Aellen) Ford-Lloyd and J. T. Williams	18	annuals (strictly semelparous)	
B. patula Ait.	18	annuals	
B. macrocarpa Guss.	18, 36	annuals	
II. Sectio Corollinae:			hilly and mountainous regions in Turkey and adjacent countries
Base species *B. corolliflora* Zosimovic ex Buttler	36		
B. macrorhiza Steven	18		
B. lomatogona Fisch et Meyer	18		
B. nana Boisier et Heldreich	18		mountain heights of Greece
Hybrid species			
B. trigyna Waldstein et Kitabel	45, 54		
B. x *intermedia* Bunge			

Fig. 1.1 The species of the genus *Beta* (as revised by Kadereit et al. 2006)

The systematics within this genus have been disputed for a long time until a new taxonomy was proposed in 2006 (Kadereit et al. 2006) which became official in June 2009. The species *B. nana*, which formerly belonged to section III, was moved to section II and the former section IV (Procumbentes) (Lange et al. 1999) became the new genus *Patellifolia* (Kadereit et al. 2006) (Fig. 1.1).

1.2 Sugar Beet Breeding and Genetics

For 200 years, sugar beet has been cultivated for sucrose production. The main breeding aims were high sucrose yield in combination with quality traits such as low Na/K content and low α-amino acid content to reduce the amount of molasses. Mass selection for sucrose content was extremely successful in the early period of beet breeding because the dry matter, whose content was easily measurable, mainly consists of sucrose. Today sucrose is measured directly by refractrometry and α-amino acid and Na/K content are determined separately. Mass selection has been replaced by single plant selection in the 1930s and today hybrids are exclusively used in the major beet-growing areas. But also a number of other traits have been substantially improved during the past 50 years. The introduction of monogermic seeds laid the foundation for beet production at an industrial scale. The successful breeding for

rhizomania resistance was a breakthrough for the cultivation of sugar beet in many growing areas of the Northern hemisphere where soils are often contaminated with the beet necrotic yellow vein virus (BNYVV).

Recently, sugar beet became an interesting alternative as a renewable energy resource in central Europe. Biomass production in this part of the world heavily relies on maize and alternatives are urgently needed. Sugar beet has the highest dry matter production capacity under central European growth conditions. The beet with its high sucrose content is also suitable for loading fermenters to produce methane. Thus, breeding biomass beets, which do not necessarily have to exhibit the quality traits of 'sucrose beets', has become an interesting option. One means to increase biomass yields is to grow winter beets. Those beets are sown before winter (preferentially in August). They overwinter in the field and develop their shoot mass early in spring. Harvest time is expected to be earlier than for conventionally grown 'spring' beets. The yielding potential of winter beet has been estimated to be $\sim 20\,\%$ higher than that of conventional beet (Hoffmann and Kluge-Severin 2011), but requires a strict bolting control (see Sect. 1.9).

The haploid chromosome number of *Beta* species is 9 ($x = 9$). Sugar beet is a diploid species with $2n = 2x = 18$ chromosomes and a haploid genome size of 758 Mb (Arumuganathan and Earle 1991). Triploids and tetraploids exist, which have been frequently used in beet breeding. Thus, the sugar beet crop is a rare example of a seed-propagated triploid crop species. Several molecular marker-based genetic maps have been published (Barzen et al. 1992; Pillen et al. 1992; Schondelmaier and Jung 1997; Schumacher et al. 1997; Grimmer et al. 2007; McGrath et al. 2007; Schneider et al. 2007) and used for mapping major genes and polygenes of agronomic importance. Unfortunately, routine procedures for doubled haploid production such as microspore or anther culture are so far lacking. Doubled haploids can only be produced by costly and time-consuming gynogenesis. Therefore, F_2 or advanced inbred populations have been used for mapping. Apart from bolting time genes, which will be discussed in Sects. 1.6–1.8, major genes for nematode resistance (Cai et al. 1997; Kleine et al. 1998), rhizomania resistance (Barzen et al. 1992; Barzen et al. 1997; Lein et al. 2007a) and monocarpic seeds (Barzen et al. 1992) have been mapped. Also, a number of QTLs have been placed on the beet chromosomes such as *Cercospora* leaf spot resistance (Nilsson et al. 1999; Schäfer-Pregl et al. 1999; Setiawan et al. 2000), *Rhizoctonia* root rot resistance (Lein et al. 2007b), fertility restorer genes (Hjerdin-Panagopoulos et al. 2002) and quality traits (Schneider et al. 2002). The efficiency of association mapping in sugar beet was recently demonstrated by mapping a number of quantitative traits (e.g. sucrose content) in a panel of 460 elite sugar beet lines (Würschum et al. 2011). Other resources for studying the beet genome have been established in the past years and will be discussed in Sect. 1.4.

1.3 Phenology of *Beta Species*

The development of sugar beet after sowing in spring is characterized by secondary root growth and the formation of a large leaf rosette in the first year. During the vegetative phase, sugar beets develop a large harvestable organ, which is mainly

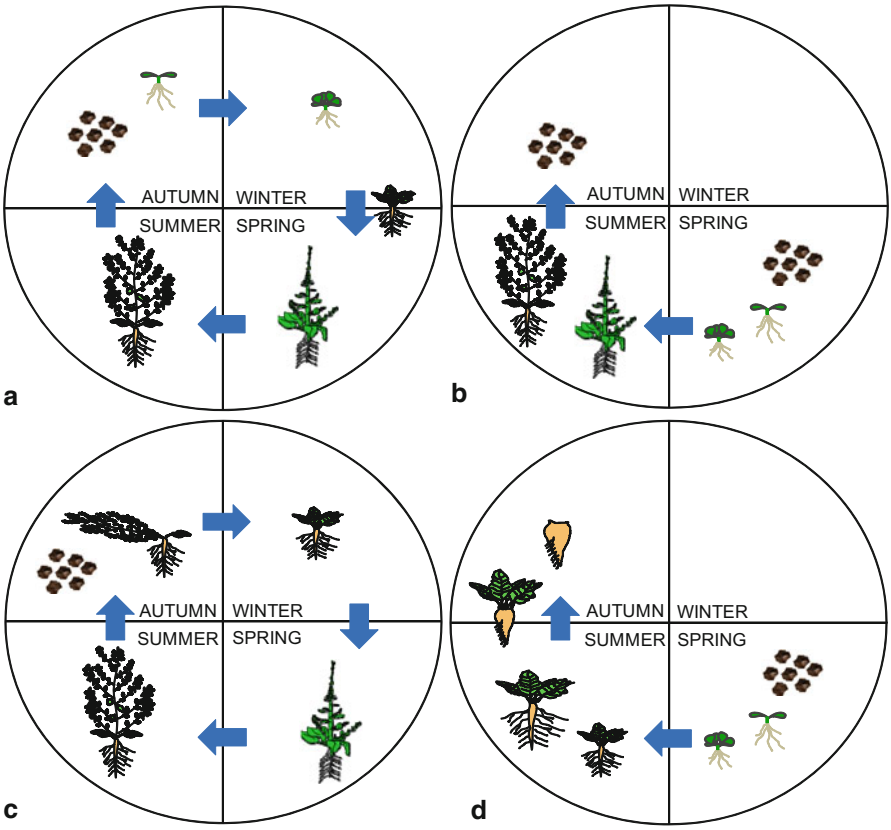

Fig. 1.2 The phenological development of sugar beet and its wild relative *Beta vulgaris* ssp. *maritima*. **a** Life cycle of biennial wild beets. **b** Life cycle of annual wild beets. **c** Life cycle of perennial wild beets. **d** Sugar beet field production ('spring beet').

formed by the root and contains only small portions of epicotyl and hypocotyl. The beet root results from secondary thickening with up to 12 successive concentric rings of cambia (Bell et al. 1996). Each cambium forms a cylindrical ring of xylem and phloem tissue and parenchyma cells in between two rings (Bhambie et al. 2000). The number of rings is much smaller in fodder beet and red beet (3–5 rings).

Sugar beets enter the generative phase only after exposure to cold temperatures typical for winter periods under central European conditions. The first visible event is the elongation of the shoot, referred to as 'bolting', usually followed by flower formation (Fig. 1.2a). A plant can have more than one flowering shoot, which are panicles and carry numerous hermaphrodite flowers (up to 10,000) that are formed in the axils of bracts. In wild beets, 2–4 flowers are merged and develop a multigerm seed ball. *B. vulgaris* is an allogamous species due to a gametophytic self-incompatibility system controlled by two series of sterility alleles (*S1-Sn*, *Z1-Zn*). Thus, self-pollination is avoided leading to highly heterozygous and heterogeneous wild populations. However, a self-fertility locus with a self-fertility allele *SF* exists, which is frequently used for selfing sugar beets to produce inbred lines (Biancardi et al. 2005).

Fig. 1.3 Cultivated beets and the related wild species *Beta vulgaris* ssp. *maritima*. **a** Ready-to-harvest beets, bolting (right) and non-bolting, with and without leaves (left). **b** Beet volunteers in a beet production field. **c** Flowering beets for seed production. **d** Annual wild beet *Beta vulgaris* ssp. *maritima*

Wild beets from the Mediterranean area are annuals, flower early without vernalization and finish their life cycle within the first year (Fig. 1.2b). By contrast, wild beets growing in the northern regions are biennials with a marked requirement for cold temperatures for flowering (Fig. 1.2a). Furthermore, long-lived, iteroparous perennials exist in the subspecies *maritima*, which produce offspring in successive cycles (Hautekeete et al. 2002) (Fig. 1.2c). However, all *Beta* species are strict long-day (LD) plants.

The onset of bolting is of greatest importance for the cultivation of sugar beet as well as of root and leaf beet. High root yield is only guaranteed if beets do not flower (Fig. 1.2d). The storage root of bolting and flowering beets is much smaller, thus sucrose yield is drastically reduced after bolting (Bürcky 1986, Fig. 1.3a) and bolting during beet production must be completely avoided. Consequently, breeders have

selected against early bolting since the beginning of beet breeding. This has been quite successful because early bolters can be easily identified and eradicated during mass selection. However, when seed production was moved to southern Europe, where annual wild beets are abundant there was an increased risk of cross pollination between wild beets and male sterile sugar beet seed parents in seed production fields. Since early bolting is controlled by a single dominant allele (see Sect. 1.5) heterozygous beets resulting from cross pollination would bolt early, creating a need for rigorous elimination of wild beets and strict isolation of seed production fields. Today, molecular markers (Gaafar et al. 2005, see Sect. 1.5) are employed for testing seed lots for the presence of early bolters. However, even biennial beets can have a tendency towards early bolting under certain environmental conditions such as exposure to cold temperatures in spring (Fig. 1.3b). Therefore, sowing time is delayed in some areas with a risk of cold temperatures late in spring (Milford and Burks 2010). On the other hand, for seed production beets must bolt and flower readily after winter (Fig. 1.3c). Therefore, breeders have selected for early flowering after winter, and completely bolting-resistant beets (that will never bolt) are not found among cultivated beets.

1.4 Genomic Resources for Beet

Genetic mapping has been used in model and crop plants to map and clone many flowering time genes in recent years (Turck et al. 2008). In contrast to model species, no collection of defined flowering time mutants is available for sugar beet. However, phenotypic variation for flowering time is easily observable among natural accessions and in structured populations derived from crosses between annual wild beets, beet cultivars and/or breeding lines. QTL analyses have been performed and linkage maps are available, but efforts to construct high density molecular marker maps thus far are rare (Lange et al. 2010). Therefore, cloning of flowering time genes from beet by mapping procedures is still challenging and time consuming (McGrath et al. 2007).

Other resources for studying the beet genome have been established in the past years. Large insert libraries exist for several beet genotypes which representatively cover the whole beet genome (Hohmann et al. 2003; Hagihara et al. 2005; Schulte et al. 2006; Lange et al. 2008). A sugar beet EST database can be found at Michigan State University (http://genomics.msu.edu/sugarbeet/index.html) and approximately 30,000 *B. vulgaris* ESTs are listed in GenBank (http://www.ncbi.nlm.nih.gov/sites/entrez?db=nucest&cmd=DetailsSearch&term=(beta+vulgaris)+AND+%22Beta+vulgaris%22%5Bporgn%5D&save_search=true). From The GenBank EST database, 315 ESTs have been placed on a sugar beet map (Schneider et al. 2007) and 2,752 were used to produce macroarrays for expression analyses (Pestsova et al. 2008). An Agilent 15 K oligonucleotide microarray has been established, which was used for mapping 392 BAC-end derived sequences and 119 ESTs (Lange et al. 2010). A beet genome mapping and sequencing consortium has started working in 2004 with the aim to physically map (GABI—The German

Plant Genome Research Program Progress Report 2004–2007; http://www.gabi.de/client/media/3/gabi_progrep_ii_web.pdf) and sequence the whole beet genome using second generation sequencing technology (http://www.gabi.de/projekte-alle-projekte-neue-seite-144.php). Approximately 67,000 genomic survey sequence fragments including BAC end sequences have been deposited in GenBank (http://www.ncbi.nlm.nih.gov/nucgss/?term=%22Beta%20vulgaris%22) and the preliminary annotation yielded approximately 28,000 gene models (Weisshaar et al. 2011). To characterize sugar beet seed vigor, a proteomic analysis was performed and 759 proteins with specific root, cotyledon and perisperm expression profiles were identified (Catusse et al. 2008). Finally, at least two transcriptomics projects are underway to generate genome-wide expression profiles and/or transcript sequences for specific developmental processes (vernalization) or tissues (shoot apex) that are relevant for the study of flowering time control in beet (http://www.gabi.de/projekte-alle-projekte-neue-seite-171.php; http://www.ncbi.nlm.nih.gov/bioproject/73561).

1.5 Genetics of Bolting Time in Beet

While beets in natural environments require long days for bolting to occur, there is considerable intraspecific variation in vernalization requirement, which follows a latitudinal cline. Wild beets from the southern part of the species distribution area (the Mediterranean) bolt in the first year without experiencing prolonged periods of cold temperatures and generally behave as annuals, but may live and flower for up to three consecutive years (Van Dijk 2007). Beets from northern latitudes including the Atlantic and North Sea coasts have a longer lifespan of approximately 4–17 years or more (Hautekeete et al. 2002; Van Dijk 2007) and commonly require vernalization, but there is quantitative variation among natural populations of different geographic origins in the extent of cold exposure required (Van Dijk et al. 1997; Boudry et al. 2002). Owen et al. (1940) coined the term 'photothermal induction' to describe the inductive effects of low temperatures and long photoperiods on bolting in *B. vulgaris* and showed that, although not required in their natural habitats, exposure to cold also promotes and accelerates bolting in annuals. As a result of strict selection during the breeding process against the annual character, which is associated with poor root yield and interferes with harvest operations, sugar beet and other cultivated beet forms require vernalization to bolt and, for seed production, are grown as biennials.

Genetically, the annual growth habit is under the control of a major dominant gene that has long been referred to as the 'bolting gene' or '*B*' (Abegg 1936). Plants which are derived from crosses between homozygous annual (*BB*) and biennial beets (*bb*) and are heterozygous at the *B* locus (*Bb*) behave as annuals under favorable conditions but may bolt several days later than homozygous annuals (Munerati 1931; Abegg 1936; Mutasa-Göttgens et al. 2010). Heterozygotes may also fail to bolt in the first year under suboptimal photothermal conditions as they are present e.g. in late spring, summer or autumn sowings (Owen 1954; Boudry et al. 1994; Abe et al. 1997). In

addition to appropriate environmental conditions, the manifestation of the annual character is also influenced by additional, modifying genes (Abe et al. 1997; Büttner et al. 2010; Abou-Elwafa et al. 2012). Furthermore, Owen et al. (1940) defined a locus for easy-bolting tendency (*B*') in a biennial beet accession which does not bolt without prior vernalization under field conditions, but bolts easily and early without vernalization under relatively low temperatures and long photoperiods in the greenhouse. On the basis of linkage data between the *B* locus and the *R* locus for hypocotyl color, and between *B*' and *R*, the authors concluded that *B*' is allelic to *B*.

Following the original observation by Rimpau (1876, 1880) that the annual habit is dominantly inherited in beet, and later work by Munerati (1931), who on the basis of phenotypic data for large F2 populations segregating for annuality suggested further that this trait follows a monogenic mode of inheritance, Abegg (1936) was able to show that the *B* gene (or 'factor' in the language of his time) in Munerati's annual accessions was linked to the hypocotyl color factor '*R*'. Abegg calculated a cross-over value of 15.5 % between *B* and *R*, and by considering the previously identified linkage relationship between *R* and *Y*, another locus affecting pigmentation, defined the first linkage group with three morphological markers in beet. The *Y-R-B* linkage group together with additional markers was later assigned to chromosome II according to the standard nomenclature for beet chromosomes suggested by Schondelmaier and Jung (1997). Using a backcross population derived from a biennial parent and a different annual accession than had been analyzed in Munerati's and Abegg's early studies, Boudry et al. (1994) confirmed linkage of a locus for annuality, which was presumed to be *B*, to *R* in their population, and were able to identify several restriction fragment length polymorphism (RFLP) markers more closely linked to *B*. The *B* locus was further fine-mapped by anonymous fragment length polymorphism (AFLP) mapping to a 0.37 cM interval of chromosome II using another unrelated mapping population (El-Mezawy et al. 2002). BAC library screening and bulked segregant analysis identified, respectively, several sequence-based markers which flank the *B* locus on either side (Hohmann et al. 2003; Gaafar et al. 2005) or completely co-segregate with *B* (Büttner et al. 2010).

Besides the major bolting locus *B*, two recent studies identified additional, previously unknown loci, which contribute to annual bolting in wild beets (Büttner et al. 2010; Abou-Elwafa et al. 2012). A screen for bolting mutants derived from ethyl methanesulfonate (EMS) mutagenesis of the same annual beet accession that was used for fine-mapping of *B* (El-Mezawy et al. 2002) identified five M3 families which bolted only after vernalization and thus behaved as biennials (Hohmann et al. 2005). Somewhat surprisingly at the time, in two out of several F2 mapping populations derived from crosses between these biennial genotypes and an annual wild beet accession, the annual bolting phenotype did not co-segregate with the *B* locus, but instead was mapped to a new locus on chromosome IX which was named *B2* (Büttner et al. 2010). Because all plants in these populations carried the dominant allele for annuality at the *B* locus, but approximately one quarter of plants failed to bolt without vernalization, the authors concluded that *B2* acts epistatically with *B* to co-regulate vernalization-independent bolting. The genetic and phenotypic data further indicated that *B2*, similar to *B*, harbors a major gene which is inherited in a

dominant-recessive fashion. Co-segregation analyses of the remaining populations segregating for annuality revealed that the natural accession used as annual parent carries at least one additional locus which also promotes bolting, but in contrast to *B2* appears to act independently of *B*. A QTL analysis of annual bolting in two populations showed that the *B* locus and the newly identified locus, termed *B4*, contributed equally to phenotypic variation in bolting behavior, and that *B4* also exhibited a dominant gene action (Abou-Elwafa et al. 2012). The *B4* locus is genetically linked to the *B* locus and was mapped to chromosome II at a genetic distance of 11 cM from *B*.

1.6 Flowering Time Genes and Their Regulation in Beet

1.6.1 Beet Homologs of the FLC Gene and Putative Regulators

Vernalization, a prolonged exposure of plants to cold temperatures over winter, is a prerequisite for many plants to flower in the following spring or summer. For Arabidopsis and other Brassicaceae, it has been shown that the MADS box gene *FLC* is the main regulator of the vernalization response. *FLC* acts as a flowering time repressor showing a characteristic expression pattern: before vernalization *FLC* mRNA accumulates to high levels, but during vernalization expression declines and remains low post-vernalization. In winter-annual Arabidopsis accessions, *FLC* is repressed by an epigenetic mechanism and is only de-repressed in the next generation. There are five paralogs of *FLC* (*MAF1–MAF5*) that are also reported to be regulated by vernalization in Arabidopsis, but these show only a mild response compared to *FLC* (Ratcliffe et al. 2003).

In cereals, wheat and barley winter varieties also exhibit a clear vernalization requirement for flowering. However, the vernalization response involves different major players, often without clear homologs in dicot species, indicating that the regulation of vernalization response evolved independently in dicots and monocots (Kim et al. 2009). One exception is *VRN3*, an *FT* ortholog from Arabidopsis, as both *FT* and *VRN3* integrate signals from various regulatory pathways and promote flowering.

Recently, in beet a homolog of *FLC* that was named *BvFL1* has been identified in EST libraries by using a phylogenetic approach (Reeves et al. 2007). The authors showed that four splice variants of *BvFL1* RNA were present in beets, which they constitutively expressed in an Arabidopsis *flc3* null mutant. All splice variants caused later flowering (Reeves et al. 2007) but to a much lesser extent than transgenic plants overexpressing the endogenous Arabidopsis *FLC* gene (Michaels and Amasino 1999). Nevertheless, *BvFL1* also acts as a repressor for flowering in transgenic Arabidopsis and two of the four splice variants are also down-regulated in beet leaves in response to a vernalization treatment of 90 days. However, after vernalization, expression of these splice variants was not stably repressed and the expression recovered to pre-vernalization levels. In addition, *BvFL1* is expressed at equal levels in annual and vernalization-requiring biennial beets, suggesting that the difference

in vernalization requirement cannot be attributed to differences in the abundance of *BvFL1* transcripts (Reeves et al. 2007). Therefore, it seems unlikely that *BvFL1* is the primary target for the vernalization response in biennial sugar beets.

In Arabidopsis, *FLC* is also regulated by a number of genes assigned to the 'autonomous pathway' of flowering time control (Simpson 2004). Beet homologs of several pathway members, namely *BvFLK*, *BvFVE1*, *BvLD* and *BvLDL1*, have recently been identified by Abou-Elwafa et al. (2011). It was shown that *BvFLK* overexpression leads to earlier flowering and can complement an Arabidopsis *flk* mutant. *BvFLK* also repressed the endogenous *FLC* gene in transgenic Arabidopsis, suggesting that gene function is at least conserved to some extent between Arabidopsis and beets. However, the authors also found indications for evolutionary divergence of autonomous pathway gene homologs in Arabidopsis and beets. Overexpression of *BvFVE1* in an Arabidopsis *fve* mutant did not rescue the late flowering mutant phenotype. Furthermore, in apparent contrast to its homolog in Arabidopsis, *BvFVE1* is under circadian clock control. Since beet carries a second closely related *FVE* homolog (*BvFVE2*), it is conceivable that *BvFVE1* and *BvFVE2* underwent sub-functionalization and that *BvFVE2* is a functional *FVE* ortholog (Abou-Elwafa et al. 2011).

1.6.2 Photoperiodic Pathway and CO Homologs

While the key regulators of vernalization requirement and response differ between distantly related species such as *A. thaliana* and cereals, a core component of the photoperiodic regulation of flowering appears to be largely conserved among angiosperms. The central regulator of the photoperiod pathway in Arabidopsis is the CONSTANS, CONSTANS-LIKE and TIMING OF CAB EXPRESSION 1 (CCT) domain transcription factor gene *CONSTANS* (*CO*), which promotes flowering in response to LD conditions (Suarez-Lopez et al. 2001). CO activity is diurnally regulated both at the transcriptional level and post-translationally, and is highest at the end of the light phase in long days when high levels of *CO* transcription and high CO protein stability coincide (Turck et al. 2008; Srikanth and Schmid 2011). The concurrent effects of both exogenous and endogenous factors on *CO* activity at critical times of the day, which involve circadian clock-regulation of transcriptional *CO* regulators and light-regulated stabilization of the protein, suggest that regulation of *CO* can account for much of the molecular basis of the 'external' and 'internal coincidence' models proposed by Bünning (1936), Pittendrigh and Minis (1964) and Pittendrigh (1972) for the induction of flowering (and other biological processes) by photoperiod (Turck et al. 2008; Srikanth and Schmid 2011). Once stably expressed under inductive LD conditions, CO transcriptionally activates *FT* in the leaf vasculature. Although the exact mode of this activation is not well understood, several co-regulatory proteins which interact with CO and contribute to the regulation of *FT* have been identified (Wenkel et al. 2006; Song et al. 2012).

CO homologs have been identified in numerous dicotyledonous and monocotyledonous species, including species which flower in response to different photoperiodic conditions such as short-day and day-neutral plants (reviewed in Turck et al. 2008). For several of these species, including both LD and short-day (SD) plants among monocots, a functional role of *CO* homologs in photoperiodic regulation of flowering has been demonstrated. However, the mode of action of *CO* genes differs to some extent between species and may be modified by various interactions with co-regulatory genes and/or light-induced changes of the protein, as has been suggested for rice (Turck et al. 2008; Ishikawa et al. 2011). Furthermore, there is also increasing evidence for *CO*-independent photoperiodic regulation in monocots including rice, where a regulatory mechanism involving the species-specific transcriptional regulators *Grain number, plant height and heading date 7* (*Ghd7*) and *Early heading date 1* (*Ehd1*) enables expression of the *FT* homolog *Heading date 3a* (*Hd3a*) under SD conditions irrespective of *Heading date 1* (*Hd1*), the rice ortholog of *CO* (Doi et al. 2004; Itoh et al. 2010). In barley, the major determinant of LD response was identified as *Photoperiod-H1* (*Ppd-H1*), which also carries a CCT domain but is otherwise unrelated to *CO* (Turner et al. 2005), whereas the function of *CO* homologs in barley is less understood and allelic variants have not been identified. A recent study of transgenic plants over-expressing *HvCO1*, the closest barley homolog of *CO* and the putative ortholog of *Hd1* in rice (Griffiths et al. 2003), showed that *HvCO1* indeed also promotes flowering in barley, in a process involving activation of the *FT* homolog *HvFT1* (Campoli et al. 2011). Interestingly, natural variation at the *Ppd-H1* locus affected flowering time irrespective of high transgenic expression of *HvCO1*, leading the authors to suggest that *Ppd-H1* may 'bypass' the regulatory *CO-FT* interaction (Campoli et al. 2011) and raising the possibility that, while *HvCO1* is a functional regulator of flowering time, the photoperiod response in barley may also involve an *HvCO1*-independent pathway.

Like in other species, a family of *CO*-like genes has also been identified in beet (Chia et al. 2008). However, none of the closest *CO* homologs in beet identified thus far (*BvCOL1* and *BvCOL2*) appear to be true orthologs of *CO*, but instead are more closely related to *CO-LIKE 1* (*COL1*) and *COL2* in Arabidopsis. Consistently, the diurnal expression profile of *BvCOL1* more closely resembled the profiles of *COL1* and *COL2*, and showed that *BvCOL1*, in contrast to *CO*, was not or only very weakly expressed at the end of the light phase in LDs (Chia et al. 2008). Nevertheless, over-expression of *BvCOL1* in Arabidopsis rescued the late-flowering phenotype of the loss-of-function *co-2* mutant and activated *FT* expression, suggesting at least a certain degree of functional conservation of the *BvCOL1* gene product (Chia et al. 2008). Perhaps noteworthy, the over-expression of *HvCO1* failed to complement the same mutant, which was suggested to result from sequence variation at conserved positions in a B-Box-type zinc finger domain (B-Box2) (Campoli et al. 2011), whereas this domain is highly conserved between *CO* in Arabidopsis and *BvCOL1*. Like the *B* and *B4* loci, *BvCOL1* was mapped to chromosome II, but at large genetic distances of approximately 22–24 cM upstream of *B* and 35–38 cM upstream of *B4* (Chia et al. 2008; Abou-Elwafa et al. 2012).

1.6.3 Two Copies of FT Homologs with Different Function in Beet

FT is a member of a protein family with structural similarities to mammalian phosphatidylethanolamine-binding protein (PEBP) domains (Kardailsky et al. 1999; Kobayashi et al. 1999) and a hitherto unknown biochemical function. In Arabidopsis, the PEBP protein family consists of three phylogenetically distinct groups represented by FT, TFL1 and MFT. FT and TSF are components of the long-sought florigen signals that promote flowering in Arabidopsis under LDs, but also integrate signals from other flowering time pathways, whereas TFL1 acts antagonistically to prevent flowering. It has been shown in other species (including both LD and SD plants) that expression of *FT* orthologs rises in response to inductive photoperiods, and that constitutive expression induces early flowering whereas mutations in *FT* orthologs delay flowering. As had been expected for the long elusive florigen, the FT protein moves from the leaves to the apex where it establishes flowering (Turck et al. 2008).

In sugar beet, Pin et al. (2010) identified two paralogous *FT* genes. Surprisingly, these genes, which were termed *BvFT1* and *BvFT2*, have antagonistic functions. While *BvFT1* acts as a repressor, *BvFT2* promotes flowering. After vernalization, biennial sugar beets are competent to flower, but the vernalized plants still require long days for floral induction. Diurnal expression studies in annual, biennial and vernalized biennial plants in different photoperiods showed that under non-inductive SD conditions *BvFT2* expression was hardly detectable, whereas *BvFT1* showed a distinct morning expression peak. Without vernalization, only annual beets bolt in LDs which was found to be coincident with very low *BvFT1* expression, whereas *BvFT2* expression peaked after 12 hours of illumination in an 18 hour photoperiod. However, unvernalized biennial beets exhibited a very different LD expression profile. Here, *BvFT2* was not expressed, whereas *BvFT1* showed a peak of expression in the morning not dissimilar to that also observed in non-inductive SD. During vernalization, *BvFT1* was down-regulated and *BvFT2* was up-regulated, indicating that *BvFT2* may be repressed by *BvFT1*. Moreover, when vernalized biennial plants were transferred to SD conditions, which lead to de-vernalization and suppression of bolting, *BvFT1* expression was induced and *BvFT2* was repressed. Finally, the observed correlation of *BvFT2* expression with the initiation of flowering in both annual and biennial beet and a complementation analysis in Arabidopsis *ft* mutants suggested that *BvFT2* is the functional *FT* ortholog in sugar beet.

In transgenic approaches overexpressing *BvFT2* under the control of the 35S promoter or down-regulating *BvFT2* expression by RNA interference (RNAi), it was demonstrated that *BvFT2* is essential for flowering in sugar beet. 35S::*BvFT2* biennial plants flowered prematurely in tissue culture without vernalization and annual *BvFT2* RNAi plants failed to bolt and continued to produce leaves in LD for more than 400 days, which was correlated with strongly reduced *BvFT2* transcript levels. Since in these transformants *BvFT1* expression was not altered, the results indicate that the modulation of flowering time is directly regulated by *BvFT2* and

not indirectly via modulation of *BvFT1* expression by *BvFT2*. Transgenic annual beets constitutively expressing *BvFT1* also failed to bolt and showed a strong down-regulation of *BvFT2*, indicating that *BvFT1* represses *BvFT2* expression. Vernalized transgenic biennial beets over-expressing *BvFT1* showed a non-bolting phenotype for more than 6 months under inductive LD and a strong down-regulation of *BvFT2*. Together, these results indicate that *BvFT1* suppresses flowering under non-inductive conditions, which at least in part may be mediated by repression of *BvFT2* expression. *BvFT1* is down-regulated during vernalization, which eliminates the repressive effect of *BvFT1* on *BvFT2* and allows sugar beet plants to respond to inductive LD conditions (Pin et al. 2010). Interestingly, *Chenopodium rubrum*, a SD plant and close relative of *B. vulgaris*, also carries two close homologs of *FT* that show different expression profiles, suggesting different roles of the encoded proteins (Cháb et al. 2008). However, a functional study in *Chenopodium* has not been performed yet. It would be highly interesting to compare the functional roles of the *FT* genes in these closely related species and their possible contributions to the differentiation of these species into SD and LD plants.

Since *BvFT1* and *BvFT2* are paralogs and show a high degree of sequence similarity, it was interesting to evaluate the encoded proteins with respect to activating or repressing functions. From previous studies in Arabidopsis, it was known that a swap of the fourth exon from *FT* (i.e. a promoter of flowering) to *TFL* (i.e. a repressor) converted FT to TFL function and vice versa (Ahn et al. 2006). Strikingly, this could also be achieved by swapping just a single amino acid of an external loop of exon 4 (Hanzawa et al. 2005). The external loops encoded by exon 4 of *BvFT1* and *BvFT2* only differ by three amino acids out of 14. Exchanging exon 4 or only the external loop regions between *BvFT1* and *BvFT2* also resulted in opposite protein functions, indicating that the functional domains for the repression or promotion of flowering through PEBP proteins are well conserved in different eudicot clades (Pin et al. 2010).

1.6.4 GA Metabolism and Early Bolting

It has been shown that treatments with gibberellins (GAs) can induce bolting and flowering especially in LD plants that form rosettes under non-inductive conditions (Zeevaart 1983). In some LD plants, the amount of endogenous GAs is increased when these plants are transferred into inductive environments (Metzger 1995), but GA-deficient and GA-insensitive mutants of Arabidopsis have also provided strong evidence that GAs are required for floral induction in SD and play only a minor role in LD (Wilson et al. 1992). In Arabidopsis, the response to GA with regard to flowering time regulation is integrated by *LFY* and *SOC1*. *SOC1* is activated by GAs and overexpression of *SOC1* can rescue the non-flowering phenotype of *ga1-3* mutants in SD, whereas *soc1* mutants show a reduced sensitivity to GAs. The expression of *LFY* is directly regulated by the GA-activated GAMYB protein, while SOC1 also activates *LFY* by directly binding to its promoter. Therefore, GAs regulate *LFY* transcription by both *SOC1*-dependent and -independent pathways (Lee and Lee 2010).

Previous studies in sugar beet have shown that GAs play a role during reproductive growth (Lenton et al. 1975), and that GA contents after vernalization and during bolting increases significantly (Sorce et al. 2002). Exogenously applied GAs can stimulate cell growth in a genotype specific manner (Sadeghian et al. 1993b), but endogenous levels of GAs are not limiting for bolting and the application of exogenous GAs failed to induce bolting and flowering in non-vernalized sugar beets (Mutasa-Göttgens et al. 2009). Mutasa-Göttgens et al. (2010) studied the interaction of GAs, the bolting gene *B* and vernalization under controlled physiological conditions by analyzing the effect of GA_4 addition in inductive LDs and non-inductive SDs in sugar beet populations segregating for the *B* allele and grown with or without vernalization. Under SD conditions, GA_4 promoted bolt initiation independently of the *B* allele, but stem elongation did not occur without prior vernalization. Moreover, vernalization and addition of GA_4 could not substitute for LD conditions, indicating that long days are absolutely required for flowering in sugar beets. Finally, the authors showed that exogenous application of GA_4 had no significant effect on bolting and flowering in LD, suggesting that GAs are not a limiting factor for bolting and flowering under these conditions.

1.7 A Model for Bolting Time Regulation in Beet

While the nature and regulatory roles of the major bolting loci in beet that were detected genetically await identification and further characterization, the presence of *FT* genes in beet as well as of homologs of photoperiod, vernalization and autonomous pathway genes suggest that important components of flowering time regulation are also conserved in beet. However, the pioneering study by Pin et al. (2010) on the antagonistic action of the *FT* genes in beet and the finding that neither *B* nor the other two mapped bolting loci, *B2* and *B4*, correspond to any of the closest beet homologs of the major flowering time regulators *FT*, *CO* and *FLC*, indicate that in beet different or additional components may have evolved as central regulators of floral transition. According to the current data, *BvFT2* is a functional ortholog of *FT* in Arabidopsis and may also in beet act as a central 'executive' component of florigen which under favorable conditions transduces the promotive signals perceived by the plant in LD or under vernalizing conditions (Pin et al. 2010). However, upstream of *BvFT2*, the *FT* paralog *BvFT1*, which does not have a counterpart in Arabidopsis, acts as a repressor of *BvFT2* in SD (in both annuals and biennials) and, in biennials, in the absence of vernalization regardless of day-length. Similar to *FLC* in Arabidopsis, *BvFT1* expression in biennials is gradually down-regulated during vernalization, resulting in the release of *BvFT2* repression and enabling its activation in the long days that follow winter. In contrast to *FLC*, however, *BvFT1* is also down-regulated in LD, but only in annuals the LD signal suffices for stable repression of *BvFT1* throughout the course of the day and activation of *BvFT2* (Pin et al. 2010). Because annual beets, carrying the dominant *B* or *B4* alleles, bolt readily in long days, whereas biennial beets require vernalization as an additional stimulus for bolting to occur, it is conceivable

that the recessive alleles may be impaired in sensing the inductive LD signal, and that *B* and/or *B4* are part of a (*CO*-dependent or independent) photoperiod pathway in beet which mediates the effect of day-length on the two *FT* genes. Because *B2* acts epistatically to *B*, it may constitute another component of the same floral induction pathway. Large-scale mapping and sequencing projects are underway to identify the underlying genes (Müller, Jung and colleagues, unpublished) and are expected to help elucidating the molecular basis of the long-postulated interaction of photoperiod and vernalization response mechanisms in beet (Owen et al. 1940). However, the currently available data for beet suggest that not only monocots and dicots have evolved different molecular mechanisms to adapt the timing of reproduction to the climatic conditions of their natural habitats, but that different strategies have also evolved within the dicot clade (Jarillo and Pieiro 2011).

1.8 Exploiting Natural Variation for Bolting Time

All cultivated beets are non-bolting under non-inductive environmental conditions, typically after sowing in (late) spring. Beet varieties grown as winter beets in some regions of southern Europe, where mild winters are predominant, have a high bolting resistance. On the other hand, plants with an extremely high bolting resistance would create problems in sugar beet seed production and have thus been avoided by breeders (Sadeghian and Johansson 1993).

Although undesirable, beets can have a tendency for early bolting under central European growing conditions and differences can be found among breeding lines (Sadeghian et al. 1993a). Longer periods of cold temperatures (5–8 °C) in spring can result in a significant increase of bolting beets in the field (Jolliffe and Arthur 1993) (Fig. 1.3b). Several studies have also measured the genetic effects underlying early bolting in biennial beets grown under flowering inductive conditions, e.g. by sowing early in spring (Jolliffe and Arthur 1993; Sadeghian and Johansson 1993). Under these conditions the frequency of bolters in July increased to 75 % (Jolliffe and Arthur 1993). Early bolting in biennial cultivated beets was found to be a quantitative character and high additive genetic effects were calculated (Sadeghian and Johansson 1993). In most studies bolting resistance was dominant over bolting susceptibility (Jolliffe and Arthur 1993).

Rich phenotypic variation for bolting time has been found in recent studies from our institute. In the course of a project to breed winter beets (see Sect. 1.9) a worldwide collection of cultivated and wild *B. vulgaris* (Fig. 1.3d) accessions was grown over winter at different locations across Europe. While annual flowering was abundant among wild beets, all cultivated types including chards and red table beets were biennial. Interestingly, all sugar beets bolted early after winter whereas wild beets showed a larger variation for bolting behavior even after winter (unpublished data). This demonstrates the potential to breed beets with altered bolting characters from the primary gene pool of *B. vulgaris*. With the sequences of major flowering time regulators at hand, it has also become possible to analyze phenotypic variation in

bolting time for associations with sequence variation in candidate genes, and a corresponding project has been initiated. Finally, we are also using TILLING (Targeting Induced Local Lesion in Genomes; McCallum et al. 2000) to detect new sequence variation in flowering time regulators. In brief, candidate gene sequences are amplified by PCR and subjected to endonuclease digests to detect mismatches between a reference wild type sequence and variant alleles, which arose either as a result of natural variation or after EMS-induced mutagenesis. While the frequency of induced mutations was found to be very low in two independent EMS-mutagenized beet populations, a number of 'natural' sequence variants have been found which in some cases could be correlated with altered bolting behaviour (Frerichman et al., in preparation).

1.9 Strategies to Breed Winter Beets by Manipulating Major Bolting Time Regulators

Sugar beets are sown in spring under central European growth conditions (temperate climates). Winter beets, i.e. beets sown before winter, are only grown on a limited scale in some parts of southern Europe. There has been increasing evidence in the past years that winter beets cultivated under temperate climates could be superior in yield to traditionally grown beets (Hoffmann and Kluge-Severin 2010). Under these conditions beets would be sown in August and stay in the field over winter. Thus, winter hardiness is a prerequisite for winter beet cultivation, and experimental data indicate that there is sufficient genetic variation for winter hardiness within the primary gene pool of *B. vulgaris*. A total of 396 cultivated and wild *B. vulgaris* accessions were grown at five different locations across central Europe. As expected for a character which has never been targeted by breeding, survival rates varied widely across the test locations and ranged from 0.7 to 86.3 % (Kirchhoff et al. 2012). The sugar beet accessions showed the highest winter hardiness and had a survival rate of 39.7 % across all environments, suggesting that there is sufficient genetic variation for breeding winter beets.

Winter beet field trials have been performed on a limited scale in Europe. Due to the lack of bolting resistant beets traditional spring beet cultivars were used, which, when sown before winter, started bolting in spring and set flowers a few weeks later. Thus, their root yield was considerably reduced, and bolting beets have a lower sugar but a higher marc content which makes them unsuitable for sugar recovery (Hoffmann and Kluge-Severin 2011). Because root mass is highly correlated with the season and duration of radiation absorption, the yield potential of winter beets was estimated by the absorbed light and the radiation use efficiency. An increase of up to 26 % was estimated independently by two different research groups (Jaggard and Werker 1999; Hoffmann and Kluge-Severin 2010) which makes autumn sown beets an interesting alternative. Moreover, winter beets will reach harvest time earlier. It has been speculated that winter beets will reach the same yield as spring beets 6–8 weeks earlier (Hoffmann and Kluge-Severin 2010). This alone could be a strong

advantage over spring beet cultivation because sugar factories would be able to make better use of their capacities if the beet harvest period will be extended. However, a substantial increase in beet yield requires a higher sink capacity (Hoffmann and Kluge-Severin 2010), and it remains to be seen whether this will be the case in winter beets. Furthermore, for the yield potential to be realized in the field, high disease pressure before winter has to be taken into account, which requires new plant protection management strategies.

The breeding of beets which are bolting-resistant after winter is a specific challenge for beet breeders. No genetic variation for this character is expected among the elite germplasm because breeders have selected for relatively early bolting and rapid flowering after winter, as otherwise seed production would not be economical. Field trials at our institute confirmed that most beet cultivars bolt readily after winter. However, we also found some variation in *B. vulgaris* for bolting behaviour, with some accessions bolting late or comprising high ratios of non-bolting plants after winter. Some accessions repeatedly did not flower until the end of summer, which is the desired phenotype of a winter beet (unpublished data). The genetic causes underlying this phenotype are presently being investigated. It remains to be seen whether the causative alleles can be useful in winter beet breeding because flowering is necessary for seed production.

Our increasing understanding of flowering time regulators in beet (see Sect. 1.6) offers a chance for genetic modification of bolting and flowering time. In principle, they rely on either repression of floral promoters or over-expression of floral repressors. While disruption of gene function can be obtained by mutagenesis (e.g. TILLING) the only realistic option currently available for gene over-expression is genetic modification by transformation. Thus, the resulting plants will be transgenic which causes some legal constraints. Examples of floral promoters as targets for mutagenesis or RNAi are *BvFT2*, as demonstrated by Pin et al. (2010) using the RNAi approach, or the bolting gene *B*, whereas the repressor gene *BvFT1* or homologs of floral repressors in other species are potential candidates for gene over-expression. For *BvFT1*, Pin et al. (2010) demonstrated that constitutive expression from a strong promoter can indeed delay bolting after vernalization for at least 6 months.

Reverting the non-bolting phenotype to induce bolting and flowering for seed production is another challenge (Fig. 1.4a). Different strategies have been suggested (Jung and Müller 2009). The expression of a transgene can be controlled if it is under transcriptional regulation of an inducible promoter, e.g. ethanol- or acetaldehyde-inducible promoters (Fig. 1.4b). Alternatively, a floral repressor such as *BvFT1* is brought under the transcriptional regulation of a constitutive or any other promoter only after cleavage of a spacer fragment separating both sequences (hybrid/recombinase approach). For site-directed cleavage, recombinase target sites flanking the spacer are introduced into a construct (parent 1). The second parent is transformed with a recombinase gene. Both parents would be expected to flower normally after vernalization, and hybrid seeds can be produced as usual. The hybrids, however, would not bolt after vernalization because the spacer fragment is removed as a result of recombinase activity, thus allowing the floral repressor gene to be transcribed from the promoter. Transgenic plants carrying both constructs have

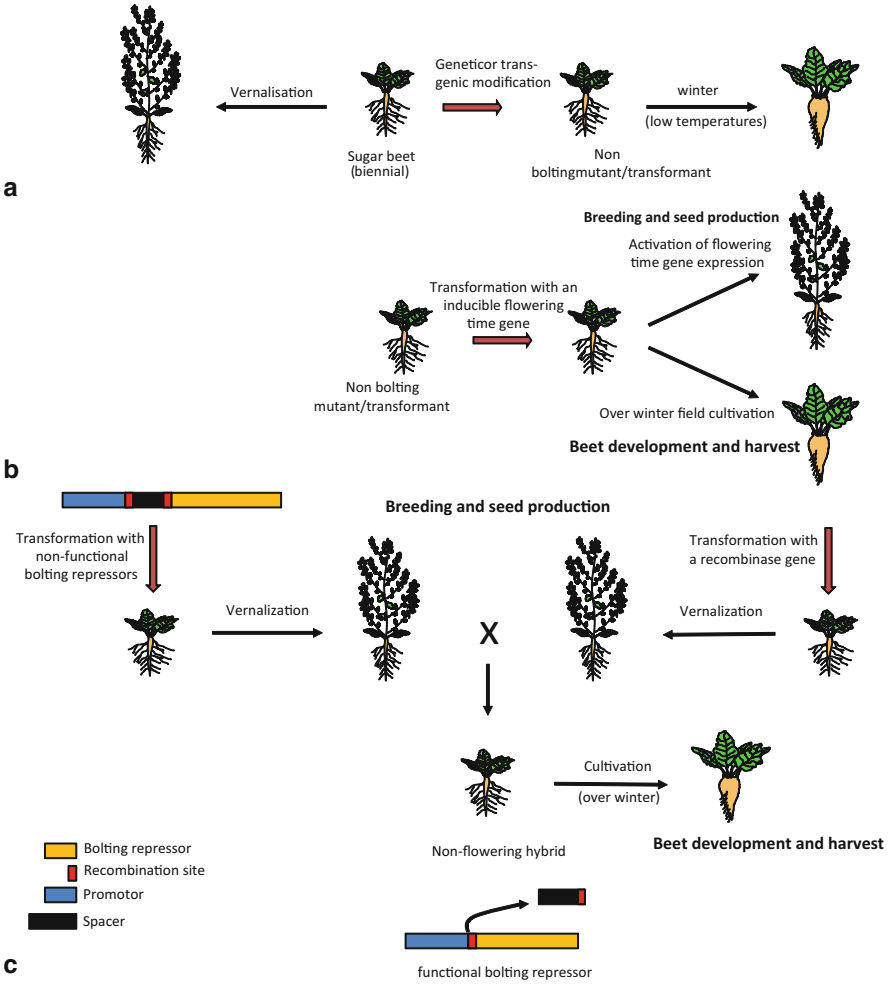

Fig. 1.4 Strategies to breed winter beets, which are bolting resistant after winter. **a** Selection of non-bolting beets after winter. **b** Bolting induction after transformation of non-bolting beets with an inducible flowering time control gene (after Jung and Müller 2009). **c** A transgenic hybrid approach to breed non-bolting F_1 hybrids by activating a flowering time repressor. Parental lines containing either a FLP recombinase or different inactive repressors for flowering time control could be used. After a genetic cross the FLP recombinase activates the repressor in F1 hybrids by depleting a spacer between promoter and coding sequence

been obtained and are presently under investigation (Fig. 1.4c). A third alternative could be the use of two late flowering parents whose phenotypes are caused by different gene actions. Depending on the nature and genetic interactions of these genes, the hybrid may not flower, or at least flower later than the parents, due to additive effects of the parental alleles. If bolting can be delayed until the time of harvest, this

approach could be an interesting alternative because it avoids, in principal, a need for genetic modification. Late bolting accessions which have been already identified from the primary gene pool of *B. vulgaris* (unpublished data) can be used as donors of late flowering alleles. A caveat of this approach is that breeding with late flowering lines is slow as a result of prolonged generation cycles. In general, hybrid approaches as described above seem a realistic option for winter beet breeding because all beet cultivars presently on the market are hybrids. Genetic modification of flowering time genes to breed non-bolting beets would be a rare (and to date unique) example of transgenic plants with improved yield potential.

Note Added in Proof Since the writing of this chapter the identification of the *B* locus by map-based cloning was described (Pin et al. 2012). In this report it was shown that the pseudo-response regulator gene *BOLTING TIME CONTROL 1* (*BvBTC1*) at the *B* locus is required for bolting in the absence of vernalization. Surprisingly, the recessive *Bvbtc1* allele in biennials facilitates bolting, too, but does so only after vernalization. Induction of bolting by *BvBTC1* or *Bvbtc1* is mediated through the regulation of *BvFT1* and *BvFT2* in both non-vernalized (annual) and vernalized (biennial) beets. According to the proposed model, expression of the dominant and fully functional *BvBTC1* allele in annual beets is enhanced in the late afternoon hours of long days, which leads to repression of *BvFT1* and activation of *BvFT2*, and in turn bolting and flowering. Biennial sugar beet cultivars carry a partial loss-of-function allele which is insufficiently active in response to long days without prior vernalization. The prolonged exposure to cold over winter enhances *Bvbtc1* activity, and possibly that of co-regulatory genes, thus restoring the plant's responsiveness to long days in the following spring. Under climate chamber conditions, down-regulation of *Bvbtc1* by RNAi resulted in suppression of bolting after vernalization, suggesting that this gene can be part of strategies towards the development of winter beets. The central role of *BvBTC1* and the *BvFT1/BvFT2* module in the regulation of vernalization requirement and photoperiod response also distinguishes beet from the well-studied flowering time control systems in Arabidopsis and cereals and illustrates the evolution of diverse floral regulatory mechanisms among angiosperms.

References

Abe J, Guan GP, Shimamoto Y (1997) A gene complex for annual habit in sugar beet (*Beta vulgaris* L.). Euphytica 94:129–135

Abegg FA (1936) A genetic factor for the annual habit in beets and linkage relationship. J Agr Res 53:493–511

Abou-Elwafa SF, Büttner B, Chia T et al (2011) Conservation and divergence of autonomous pathway genes in the flowering regulatory network of *Beta vulgaris*. J Exp Bot 62:3359–3374

Abou-Elwafa SF, Büttner B, Kopisch-Obuch FJ et al (2012) Genetic identification of a novel bolting locus in *Beta vulgaris* which promotes annuality independently of the bolting gene *B*. Mol Breed 29:989–998

Ahn JH, Miller D, Winter VJ et al (2006) A divergent external loop confers antagonistic activity on floral regulators FT and TFL1. EMBO J 25:605–614

Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. Plant Mol Biol Rep 9:208–218

Barzen E, Mechelke W, Ritter E et al (1992) RFLP markers for sugar beet breeding—chromosomal linkage maps and location of major genes for rhizomania resistance, monogermy and hypocotyl colour. Plant J 2:601–611

Barzen E, Stahl R, Fuchs E et al (1997) Development of coupling-repulsion phase SCAR markers diagnostic for the sugarbeet *Rr1* allele conferring resistance to rhizomania. Mol Breed 3:231–238

Bell CI, Milford GFJ, Leigh RA (1996) Sugar beet. In: Zamski E, Schaffer AA (eds) Marcel, New York, pp 691–707

Bhambie S, Goyal R, Gupta S (2000) Ontogeny of cambium in two genera of Chenopodiaceae. Acta Bot Indica 18:252–255

Biancardi E, Campbell LG, Skaracis GN, De Biaggi M (2005) Genetics and breeding of sugar beet. Science, Enfield.

Boudry P, McCombie H, Van Dijk H (2002) Vernalization requirement of wild beet *Beta vulgaris* ssp *maritima*: among population variation and its adaptive significance. J Ecol 90:693–703

Boudry P, Wieber R, Saumitou-Laprade P et al (1994) Identification of RFLP markers closely linked to the bolting gene *B* and their significance for the study of the annual habit in beets (*Beta vulgaris* L.). Theor Appl Genet 88:852–858

Bünning E (1936) Die endonome tagesrhythmik als grundlage der photoperiodischen reaktion. Ber Deut Bot Ges 54:590–607

Bürcky K (1986) Ertrag und qualität von zuckerrüben bei unterschiedlichem anteil an schossern im bestand. Zuckerindustrie 111:862–867

Büttner B, Abou-Elwafa SF, Zhang W et al (2010) A survey of EMS-induced biennial *Beta vulgaris* mutants reveals a novel bolting locus which is unlinked to the bolting gene *B*. Theor Appl Genet 121:1117–1131

Cai D, Kleine M, Kifle S et al (1997) Positional cloning of a gene for nematode resistance in sugar beet. Science 275:832–834

Campoli C, Drosse B, Searle I et al (2011) Functional characterisation of *HvCO1*, the barley (*Hordeum vulgare*) flowering time ortholog of *CONSTANS*. Plant J 69:868–880

Catusse J, Strub JM, Job C et al (2008) Proteome-wide characterization of sugarbeet seed vigor and its tissue specific expression. Proc Natl Acad Sci U S A 105:10262–10267

Cháb D, Kolář J, Olson M, Štorchová H (2008) Two *FLOWERING LOCUS T* (*FT*) homologs in *Chenopodium rubrum* differ in expression patterns. Planta 228:929–940

Chia T, Müller A, Jung C, Mutasa-Göttgens E (2008) Sugar beet contains a large *CONSTANS-LIKE* gene family including a putative *CO* homologue that is independent of the early-bolting (*B*) gene locus. J Exp Bot 59:2735–2748

Doi K, Izawa T, Fuse T et al (2004) *Ehd1*, a B-type response regulator in rice, confers short-day promotion of flowering and controls *FT-like* gene expression independently of *Hd1*. Genes Dev 18:926–936

El-Mezawy A, Dreyer F, Jacobs G, Jung C (2002) High resolution mapping of the bolting gene *B* of sugar beet. Theor Appl Genet 105:100–105

Gaafar RM, Hohmann U, Jung C (2005) BAC-derived molecular markers for early bolting in sugar beet. Theor Appl Genet 110:1027–1037

Griffiths S, Dunford RP, Coupland G, Laurie DA (2003) The evolution of *CONSTANS*-like gene families in barley, rice, and Arabidopsis. Plant Physiol 131:1855–1867

Grimmer MK, Trybush S, Hanley S et al (2007) An anchored linkage map for sugar beet based on AFLP, SNP anbd RAPD markers and QTL mapping of a new source of resistance to beet necrotic yellow vein virus. Theor Appl Genet 114:1151–1160

Hagihara E, Matsuhira H, Ueda M et al (2005) Sugar beet BAC library construction and assembly of a contig spanning *Rf1*, a restorer-of-fertility gene for Owen cytoplasmic male sterility. Mol Genet Genomics 274:316–323

Hanzawa Y, Money T, Bradley D (2005) A single amino acid converts a repressor to an activator of flowering. Proc Natl Acad Sci U S A 102:7748–7753

Hautekeete NC, Piquot Y, Van Dijk H (2002) Life span in *Beta vulgaris* ssp *maritima*: the effects of age at first reproduction and disturbance. J Ecol 90:508–516
Hjerdin-Panagopoulos A, Kraft T, Rading IM et al (2002) Three QTL regions for restoration of Owen CMS in sugar beet. Crop Sci 42:540–544
Hoffmann CM, Kluge-Severin S (2010) Light absorption and radiation use efficiency of autumn and spring sown sugar beets. Field Crops Res 119:238–244
Hoffmann CM, Kluge-Severin S (2011) Growth analysis of autumn and spring sown sugar beet. Eur J Agron 34:1–9
Hohmann U, Jacobs G, Jung C (2005) An EMS mutagenesis protocol for sugar beet and isolation of non-bolting mutants. Plant Breed 124:317–321
Hohmann U, Jacobs G, Telgmann A et al (2003) A bacterial artificial chromosome (BAC) library of sugar beet and a physical map comprising the bolting gene *B*. Mol Genet Genomics 269:126–136
Ishikawa R, Aoki M, Kurotani K et al (2011) Phytochrome B regulates Heading date 1 (Hd1)-mediated expression of rice florigen *Hd3a* and critical day length in rice. Mol Genet Genomics 285:461–470
Itoh H, Nonoue Y, Yano M, Izawa T (2010) A pair of floral regulators sets critical day length for Hd3a florigen expression in rice. Nat Genet 42:635–638
Jaggard KW, Werker AR (1999) An evaluation of the potential benefits and costs of autumn-sown sugarbeet in NW Europe. J Agri Sci 132:91–102
Jarillo JA, Piñeiro M (2011) Timing is everything in plant development. The central role of floral repressors. Plant Sci 181:364–378
Jolliffe TH, Arthur AE (1993) Diallel analysis of bolting in sugarbeet. J Agric Sci 121:327–332
Jung C, Müller A (2009) Flowering time control and applications in plant breeding. Trends Plant Sci 14:563–573
Kadereit G, Hohmann S, Kadereit JW (2006) A synopsis of *Chenopodiaceae* subfam. *Betoidae* and notes on the taxonomy of *Beta*. Willdenowia 36:9–19
Kardailsky I, Shukla VK, Ahn JH et al (1999) Activation tagging of the floral inducer *FT*. Science 286:1962–1965
Kim D-H, Doyle MR, Sung S, Amasino RM (2009) Vernalization: winter and the timing of flowering in plants. Annu Rev Cell Dev Biol 25:277–299
Kirchhoff M, Svirshchevskaya A, Hoffmann C et al (2012) High degree of genetic variation of winter hardiness in a panel of *Beta vulgaris* L. Crop Sci 52:179–188
Kleine M, Voss H, Cai D, Jung C (1998) Evaluation of nematode resistant sugar beet (*Beta vulgaris* L.) lines by molecular analysis. Theor Appl Genet 97:896–904
Kobayashi Y, Kaya H, Goto K et al (1999) A pair of related genes with antagonistic roles in mediating flowering signals. Science 286:1960–1962
Lange C, Holtgräwe D, Schulz B et al (2008) Construction and characterization of a sugar beet (*Beta vulgaris*) fosmid library. Genome 51:948–951
Lange C, Mittermayr L, Dohm J et al (2010) High-throughput identification of genetic markers using representational oligonucleotide microarray analysis. Theor Appl Genet 121:549–565
Lange W, Brandenburg WA, De Bock TSM (1999) Taxonomy and cultonomy of beet (*Beta vulgaris* L.). Bot J Linn Soc 130:81–96
Lee J, Lee I (2010) Regulation and function of *SOC1*, a flowering pathway integrator. J Exp Bot 61:2247–2254
Lein JC, Asbach K, Tian YY et al (2007a) Resistance gene analogues are clustered on chromosome 3 of sugar beet and co-segregate with QTL for rhizomania resistance. Genome 50:61–71
Lein JC, Sagstetter C, Schulte D et al (2007b) Mapping of rhizoctonia root rot resistance genes in sugar beet. Plant Breed 127:602–611
Lenton JR,Pocock TO, Radley ME (1975) Endogenous gibberellins and bolting of sugar beet. vol 1. UK: Rothamsted Experimental Station, 44–46
McCallum CM, Comai L, Greene EA, Henikoff S (2000) Targeting Induced Local Lesions IN Genomes (TILLING) for plant functional genomics. Plant Phys 123:439–442

McGrath JM, Trebbi D, Fenwick A et al (2007) An open-source first-generation molecular genetic map from a sugarbeet × table beet cross and its extension to physical mapping. Crop Sci 47:S27–S47

Metzger J (ed) (1995) Hormones and reproductive development. Kluwer, Dordrecht

Michaels SD, Amasino RM (1999) FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. Plant Cell 11:949–956

Milford G, Burks E (2010) Managing the risk of bolters. Brit Sugar Beet Rev 78:32–35

Munerati O (1931) L'eredita della tendenza alla annualita nella comune barbabietola coltivata. Ztschr Züchtung, Reihe A, Pflanzenzüchtung 17:84–89

Mutasa-Göttgens E, Qi A, Mathews A, Thomas S et al (2009) Modification of gibberellin signalling (metabolism & signal transduction) in sugar beet: analysis of potential targets for crop improvement. Transgenic Res 18:301–308

Mutasa-Göttgens ES, Qi A, Zhang W et al (2010) Bolting and flowering control in sugar beet: Relationships and effects of gibberellin, the bolting gene *B* and vernalization. AoB Plants. doi:10.1093/aobpla/plq012

Nilsson NO, Hansen M, Panagopoulos AH et al (1999) QTL analysis of *Cercospora* leaf spot resistance in sugar beet. Plant Breed 118:327–334

Owen FV (1954) The significance of single gene reactions in sugar beets. Proc Amer Soc Sugar Beet Technol 8:392–398

Owen FV, Carsner E, Stout M (1940) Photothermal induction of flowering in sugar beets. J Agric Sci 61:101–124

Pestsova E, Meinhard J, Menze A et al (2008) Transcript profiles uncover temporal and stress-induced changes of metabolic pathways in germinating sugar beet seeds. BMC Plant Biol 8:122

Pillen K, Steinrücken G, Wricke G et al (1992) A linkage map of sugar beet (*Beta vulgaris* L.). Theor Appl Genet 84:129–135

Pin PA, Benlloch R, Bonnet D et al (2010) An antagonistic pair of *FT* homologs mediates the control of flowering time in sugar beet. Science 330:1397–1400

Pin PA, Zhang W, Vogt SH et al (2012) The role of a pseudo-response regulator gene in life cycle adaptation and domestication of beet. Curr Biol 22:1095–1101

Pittendrigh CS (1972) Circadian surfaces and the diversity of possible roles of circadian organization in photoperiodic induction. Proc Natl Acad Sci U S A 69:2734–2737

Pittendrigh CS, Minis DH (1964) The entrainment of circadian oscillations by light and their role as photoperiodic clocks. Amer Natur 98:261–294

Ratcliffe OJ, Kumimoto RW, Wong BJ, Riechmann JL (2003) Analysis of the Arabidopsis MADS AFFECTING FLOWERING gene family: MAF2 prevents vernalization by short periods of cold. Plant Cell 15:1159–1169

Reeves PA, He Y, Schmitz RJ et al (2007) Evolutionary conservation of the FLOWERING LOCUS C-mediated vernalization response: evidence from the sugar beet (*Beta vulgaris*). Genetics 176:295–307

Rimpau W (1876) Das Aufschiessen der Runkelrüben. Landw Jahrb 5:31–45

Rimpau W (1880) Das Aufschiessen der Runkelrüben. Landw Jahrb 9:191–203

Sadeghian SY, Johansson E (1993) Genetic study of bolting and stem length in sugar beet (*Beta vulgaris* L.) using a factorial cross design. Euphytica 65:177–185

Sadeghian SY, Becker HC, Johansson E (1993a) Inheritance of bolting in three sugar beet crosses after different periods of vernalization. Plant Breed 110:328–333

Sadeghian SY, Johansson E, Lexander K (1993b) A genetic analysis of the numer of cells, length of cell, and gibberellic acid sensitivity in sugar beet and their relation to bolting mechanism. Euphytica 68:59–67

Schäfer-Pregl R, Borchardt D, Barzen E et al (1999) Localization of QTLs for tolerance to *Cercospora beticola* on sugar beet linkage groups. Theor Appl Genet 99:829–836

Schneider K, Kulosa D, Rosleff-Soerensen T et al (2007) Analysis of DNA polymorphisms in sugar beet (*Beta vulgaris* L.) and development of an SNP-based map of expressed genes. Theor Appl Genet 115:601–615

Schneider K, Schäfer-Pregl R, Borchardt DC, Salamini F (2002) Mapping QTLs for sucrose content, yield and quality in a sugar beet population fingerprinted by EST-related markers. Theor Appl Genet 104:1107–1113

Schondelmaier J, Jung C (1997) Chromosomal assignment of the nine linkage groups of sugar beet (*Beta vulgaris* L.) using primary trisomics. Theor Appl Genet 95:590–596

Schulte D, Cai D, Kleine M et al (2006) A complete physical map of a wild beet (*Beta procumbens*) translocation in sugar beet. Mol Genet Genomics 275:504–511

Schumacher K, Schondelmaier J, Barzen E et al (1997) Combining different linkage maps in sugar beet (*Beta vulgaris* L.) to make one map. Plant Breed 116:23–38

Setiawan A, Koch G, Barnes SR, Jung C (2000) Mapping quantitative trait loci (QTLs) for resistance to *Cercospora* leaf spot disease (*Cercospora beticola* Sacc.) in sugar beet (*Beta vulgaris* L.). Theor Appl Genet 100:1176–1182

Simpson GG (2004) The autonomous pathway: epigenetic and post-transcriptional gene regulation in the control of *Arabidopsis* flowering time. Curr Opin Plant Biol 7:570–574

Song YH, Lee I, Lee SY et al (2012) CONSTANS and ASYMMETRIC LEAVES 1 complex is involved in the induction of FLOWERING LOCUS T in photoperiodic flowering in *Arabidopsis*. Plant J 69:332–342

Sorce C, Stevanato P, Biancardi E, Lorenzi R (2002) Physiological mechanisms of floral stem elongation (bolting) control in sugar beet (*Beta vulgaris* ssp. *vulgaris* L.). Agroindustria 1:87–91

Srikanth A, Schmid M (2011) Regulation of flowering time: all roads lead to Rome. Cell Mol Life Sci 68:2013–2037

Suarez-Lopez P, Wheatley K, Robson F et al (2001) CONSTANS mediates between the circadian clock and the control of flowering in *Arabidopsis*. Nature 410:1116–1120

The Angiosperm Phylogeny Group (2009) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. Bot J Linn Soc 161:105–121

Turck F, Fornara F, Coupland G (2008) Regulation and identity of florigen: FLOWERING LOCUS T moves center stage. Annu Rev Plant Biol 59:573–594

Turner A, Beales J, Faure S et al (2005) The pseudo-response regulator Ppd-H1 provides adaptation to photoperiod in barley. Science 310:1031–1034

Van Dijk H (2007) Artificial selection for flowering date in the iteroparous perennial plant *Beta vulgaris* subsp. *maritima* (sea beet): an analysis of the underlying mechanisms. In: Conference of the European Society for Evolutionary Biology, Uppsala

Van Dijk H, Boudry P, McCombie H, Vernet P (1997) Flowering time in wild beet (*Beta vulgaris* ssp. *maritima*) along a latitudinal cline. Acta Oecol 18:47–60

Weisshaar B, Dohm JC, Minoche A et al (2011) The draft genome sequence of sugar beet (*Beta vulgaris* L.). In: Plant & Animal Genome Conference XIX, San Diego

Wenkel S, Turck F, Singer K et al (2006) CONSTANS and the CCAAT box binding complex share a functionally important domain and interact to regulate flowering of *Arabidopsis*. Plant Cell 18:2971–2984

Wilson RN, Heckman JW, Somerville CR (1992) Gibberellin is required for flowering in *Arabidopsis thaliana* under short days. Plant Phys 100:403–408

Würschum T, Maurer HP, Schulz B et al (2011) Genome-wide association mapping reveals epistasis and genetic interaction networks in sugar beet. Theor Appl Genet 123:109–118

Zeevaart JAD (1983) Gibberellins and flowering. Praeger, New York

Chapter 2
Mining the Genus *Solanum* for Increasing Disease Resistance

Jack H. Vossen, Kwang-Ryong Jo and Ben Vosman

Abstract Plant Breeding is the art of selecting and discarding genetic material to achieve crop improvement. Favourable alleles resulting in quality improvement or disease resistance must be added, while unfavourable alleles must be removed. The source for novel alleles can be other varieties, landraces or crop wild relatives. The identification of allelic variation is referred to as allele mining. Before allelic variation can be used for breeding purposes several steps need to be taken. First of all an inventory is needed of the available genetic resources. Phenotypic screens are needed to uncover potential expected and even unanticipated alleles. Next, using genetic and molecular tools, the alleles responsible for the identified traits must be traced and distinguished in order to be introgressed into new varieties.

In this review we focus on the identification of novel disease resistance traits in the agronomically important genus *Solanum*. The fact that *R* genes are present in multigene clusters within the genome, which often include many paralogs necessitates thorough discussion on the distinction between alleles and paralogs. Often such a distinction cannot easily be made. An overview is given of how natural resources can be tapped, e.g. how germplasm can be most efficiently screened. Techniques are presented by which alleles and paralogs can be distinguished in functional and/or genetic screens, including also a specific tagging of alleles and paralogs. Several examples are given in which allele and paralog mining was successfully applied. Also examples are presented as to how allele mining supported our understanding about the evolution of *R* gene clusters. Finally an outlook is provided how the research field of allele mining might develop in the near future.

B. Vosman (✉) · J. H. Vossen · K.-R. Jo
Wageningen UR Plant Breeding, Wageningen University & Research Centre,
P.O. Box 16, 6700 AA Wageningen, The Netherlands
e-mail: ben.vosman@wur.nl

K.-R. Jo
Research Institute of Agrobiology,
Academy of Agricultural Sciences, Pyongyang,
Democratic People's Republic of Korea

R. Tuberosa et al. (eds.), *Genomics of Plant Genetic Resources,*
DOI 10.1007/978-94-007-7575-6_2,

2.1 Introduction

2.1.1 General Introduction

Alleles are different forms of a gene and affect a particular process in different ways. Different combinations of alleles may result in different phenotypes. Plant breeders try to improve varieties by introducing new alleles, resulting in higher yields and better quality or resistance characteristics. Identifying new, promising alleles is not an easy task. In the post-genomics era, mining of a crop's (wild) gene pool for novel and superior alleles for agronomically important traits is becoming more and more feasible. Genebanks all over the world contain huge untapped resources of distinct alleles that may have potential application in crop breeding programs. This hidden diversity, which can consist of naturally occurring sequence variation in coding or regulatory regions of genes, can be explored by allele mining (Ramkumar et al. 2010; Varshney et al. 2005, 2009). The variation includes single nucleotide polymorphisms (SNPs) as well as insertions and deletions (InDels), which have the possibility to change the resulting phenotype. This may be by altering the amount of protein or its structure and/or function (Ramkumar et al. 2010). The recent rapid advancements in the field of genomics leads to the accumulation of enormous amounts of sequence information and fast evolving bioinformatic tools which pave the road for identifying, characterizing, isolating, and deploying previously unknown or under-utilized sources of genetic variation.

In this chapter we consider allele mining as the research field that aims at unlocking the genetic diversity existing in genetic resource collections (genebanks) and artificially created mutant populations by identifying allelic variants of genes and loci. Since resistance genes occur in clusters, where allelic relationships are often not clear (Sanchez et al. 2006; Millett et al. 2007) and because paralogs in a cluster can have different functions, the scope of this chapter is broader than allele mining alone. To deal with this we introduce the concept of paralog mining. Paralog mining is the identification of a gene within a cluster of highly homologues genes with different, often unknown, functions. Paralog mining can be used as a tool to generate molecular markers and in combination with functional screens it can be used to identify new genes conferring resistance to a particular pathogen. In this review we discuss how allele and paralog mining can help to improve disease resistance in *Solanum* crops.

2.1.2 Solanaceae Resources

The family of *Solanaceae* is of high economic importance and is composed of more than 3,000 species which include important crop and model plants such as potato (*Solanum tuberosum*), tomato (*Solanum lycopersicum*) and eggplant (*Solanum melongena*) (Knapp 2002), but also wild species occurring in very different habitats

(Spooner and Hijmans 2001; Spooner et al. 2004). About 15,000 wild potato accessions are being maintained in large collections worldwide and the establishment of core and mini collections that enable an effective use of the existing variation in gene banks while maintaining the variability, as has been proposed before (Frankel and Brown 1984; Hoekstra 2009). Allele mining requires the assembly of a reasonably sized core germplasm collection usually comprising ~1,000 accessions representative of genetic diversity existing in the global population (Hofinger et al. 2009). Such collections can effectively be constructed using the Focussed Identification of Germplasm Strategy (FIGS) approach (Mackay et al. 2004; Bhullar et al. 2009). About 15,000 wild potato accessions are being maintained in large collections worldwide and the establishment of core and mini collections that enable an effective use of the existing variation in gene banks while maintaining the variability, as has been proposed before (Frankel and Brown 1984; Hoekstra 2009).

The genome sequence of potato (Potato Genome Sequencing Consortium et al. 2011) and tomato (The Tomato Genome Consortium et al. 2012) will facilitate mining for novel alleles or paralogs of resistance® genes. These may be found in the largely untapped resources of crossable species within the genus *Solanum* allowing their exploitation in breeding programs. Also, insight into sequence diversity at the *R* gene loci in wild *Solanum* species with different resistance response against economically important diseases will result in a better understanding of the mechanism of *R* gene functionality and evolution but can also help to identify new alleles or paralogs with different race specificities, and develop allele-specific diagnostic markers for marker assisted breeding.

2.1.3 Resistance Genes

If a gene is responsible for the resistance of a particular plant to a particular pathogen, this gene is called a resistance ® gene. To date, more than 100 *R* genes which confer resistance to a diversity of pathogens including bacteria, fungi, oomycetes, viruses, insects and nematodes have been identified and/or cloned from various plants, by a wide variety of methods including map-based cloning, transposon tagging, and similarity based DNA library screening (Sanchez et al. 2006; Ingvardsen et al. 2008; Vleeshouwers et al. 2011a). An overview of mapped and cloned *R* genes from Solanaceae is given in Fig. 2.1.

R genes often encode receptors for pathogen derived ligands and they are classified based on the combination of different domains (e.g. CC = coiled coil, TIR = toll interleukin receptor, Protein Kinase, NBS = nucleotide binding site, Lec (lectin), and LRRs = leucine rich repeats). Five classes can be identified, transmembrane proteins with extracelular LRRs (receptor like proteins, RLPs), transmembrane proteins with extracellular LRRs and intracellular protein kinase (receptor like kinases, RLKs), transmembrane proteins with extracellular "lectin like" domain and intracellular protein kinase (lectin receptor kinases, LecRKs), and intracellular NBS-LRR proteins which can be divided in CC-NBS-LRR and TIR-NBS-LRR (Dubery et al.

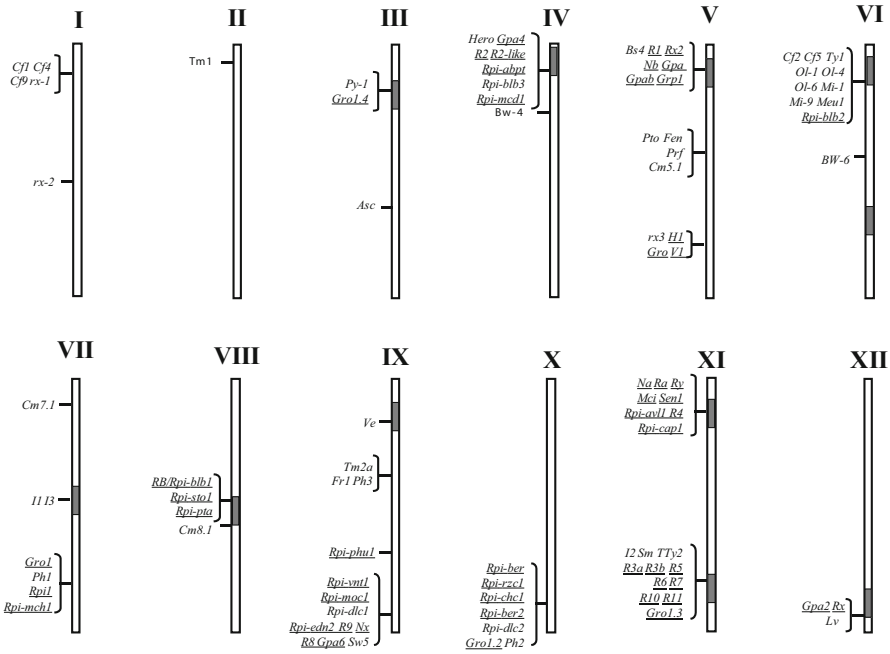

Fig. 2.1 Genetic locations of disease resistance traits in Solanaceae. Twelve linkage groups are shown and the position of *R* genes is indicated. The *R* genes for potato are underlined and those for other species, mainly tomato, are not underlined. Map segments having QTL for resistance to *Phytophthora infestans* in potato are shown in black color

2012). The NBS-LRR class is the most abundant and has been extensively studied (Hulbert et al. 2001). Although NBS-LRR genes are assumed to cause pathogen race specific (or vertical) resistance, it has also been suggested that members of the NBS-LRR gene family are candidates for quantitative trait loci (QTL) that are responsible for horizontal resistance (Rietman et al. 2012; Sanz et al. 2012; Gebhardt and Valkonen 2001). Most characterized plant NBS-LRR genes are physically clustered in the plant genome. The homologous sequences in such a cluster are referred to as paralogs (Gebhardt and Valkonen 2001) and paralogs can confer resistance to different isolates of the same pathogen (Dodds et al. 2001; Li et al. 2011; Lokossou 2010) or to different pathogens (Dodds et al. 2001; van der Vossen et al. 2000). Some paralogs may also be considered as molecular fossils of evolution, whose activity is unclear or even absent, e.g. many pseudogenes have been found. In most *R* gene clusters the number of paralogs is very high and often an allelic relationship is hard to determine (Kuang et al. 2004). However, as the genome structure between species in the *Solanaceae* family is highly conserved, positional conservation of *R* gene clusters (synteny) is observed across *Solanaceous* species (Grube et al. 2000; Park et al. 2009, Fig. 2.1). Therefore, even when relatively unknown genetic sources are used, it is likely that the genes conferring resistance are linked to syntenic clusters of *R* genes known from well-studied species like potato and tomato.

Not just the 2006; Millett et al. identification of new alleles is important, also the functional characterisation of the identified alleles is extremely important to assess the added value of the new allele over alleles that are already present in crop plants. Many approaches have already been used and especially the currently booming research field of effector genomics, through which the identification of *Avr* genes is greatly accelerated, offers fast functional assays to distinguish the activity of newly identified *R* gene alleles and paralogs. So, allele mining approaches coupled with effector profiling enable the discovery of novel *R* genes at an unprecedented rate (Vleeshouwers et al. 2008, 2011a).

2.2 Functional Resistance Screens

2.2.1 Screening for Disease Resistant Accessions in Gene Bank Material

Several methods are available to carry out phenotypic screens for disease resistance in gene bank collections. Here we use the evaluation of potato germplasm for late blight resistance as an example. Inoculation of entire *in vitro* plantlets or inoculation of detached leaves can be used as high throughput screening methods (Vleeshouwers 2011b). In case of race specific resistance, the selection of the pathogen isolates is an important issue in the identification of major *R* genes. If the selected isolate happens to be compatible with the *R* gene (s) in a particular accession, these *R* genes may be overlooked. Multiple isolates can be used to distinguish the different *R* genes in a particular resistant accession (Huang et al. 2005; Verzaux 2010). Complementary to working with the entire pathogen, effector responsiveness can be used to identify and classify *R* gene alleles in a germplasm core collection (Rietman et al. 2010). In such an effectoromics approach, effectors or potential Avr genes from the pathogen, are expressed in the plant using agro-infiltration or through inoculation with recombinant Potato Virus X, referred to as agro-infection. Upon recognition of the effector by the *R* gene expressed by the plant, a defence reaction, referred to as hypersensitive response (HR), is initiated which is visible as a necrotic lesions in the infiltrated leaf (Fig. 2.2). The agro-infiltration test appears to be applicable and reliable for many genotypes and the variation in the HR in different genetic backgrounds is limited. The use of specific pathogen isolates and the use of specific effectors can also be employed to identify functional groups of *R* genes in germplasm of crop wild relatives. Within groups of functionally similar *R* genes, a true allele mining approach can be pursued in order to identify (sequence) variation. The functional grouping of *R* genes can also be employed to reduce the redundancy that is inevitably present in germplasm collections. Another virtue of the effectoromics approach was shown recently. Potato plants which have shown durable resistance to late blight contained stacks of different *R* genes (Verzaux 2010; Kim et al. 2012). The polygenic nature of the resistances could easily be characterised using the segregation patterns of the

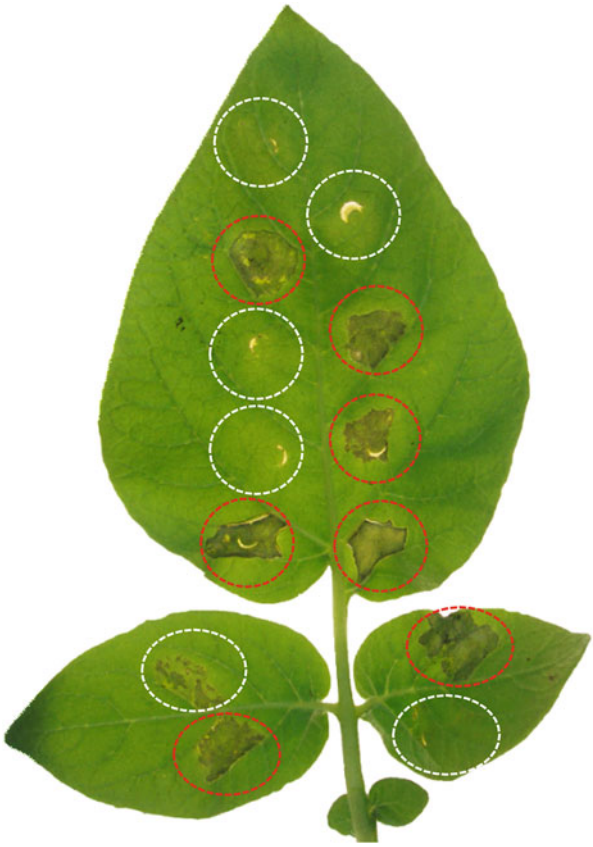

Fig. 2.2 Effector induced hypersensitive response (HR) in *Solanum tuberosum* genotype Ma*R8*. Available effectors from *Phytophthora infestans* were applied using agroinfiltration in leaves of the resistant plant Ma*R8*. The dotted circles surround the infiltrated leaf area. The red dotted circles surround the effectors that elicited an HR. These effectors are selected for validation of *R-Avr* interactions by additional genetic analysis

different effector responses. Effectors which displayed HR response in germplasm screens are potential *Avr* gene(s) recognized by the cognate *R* gene. These potential *R-Avr* interactions should be validated by additional genetic studies. Ideally, by cosegregation of responses to the effector with resistance to *P. infestans* isolates in segregating populations.

2.2.2 QTL Mapping/LD Mapping

Plant pathogen resistance, at the phenotypic level, often does not behave as a single *R* gene but as a quantitative trait that is controlled by multiple genetic and environmental factors (Trognitz et al. 2002; Bai 2003). Understanding the molecular basis for quantitative traits will facilitate diagnosis and facilitate the combination of superior alleles in crop improvement programs. The possible approaches to mapping genes that underlie quantitative traits fall broadly into two categories: candidate gene studies, which use either association or resequencing approaches, and linkage studies,

which include both QTL mapping and genome-wide association studies (GWAS). In this review, we do not discuss GWAS further because of the extensive review by Hirschhorn and Daly (2005). Linkage disequilibrium (LD) mapping, or association analysis based on candidate genes is also considered as an allele mining approach (Malosetti et al. 2007).

Some cases of close linkage between an *R* gene and quantitative trait loci (QTL) for pathogen resistance supports the hypothesis that qualitative and quantitative resistance have a similar molecular basis (Leonards-Schippers et al. 1994), thereby suggesting that genes showing sequence similarity to *R* genes are candidates for being factors underlying quantitative resistance (Rickert et al. 2003; Rietman et al. 2010). Candidate genes participating in the control of the quantitative resistance to pathogens are those involved in the disease response network; (i) *R* genes which recognize the pathogen and trigger the resistance response, (ii) genes which are involved in signal transduction pathways and (iii) the large group of pathogensis related (PR) genes which are expressed in response to pathogen attack and are involved in the execution phase of the defence response (reviewed by Gebhardt and Valkonen 2001).

The genetic dissection of complex plant traits in QTLs first became possible with the advent of DNA-based markers (Osborn et al. 1987). The first genes and their allelic variants underlying plant QTLs have been identified by positional cloning (reviewed in Salvi and Tuberosa 2005). Positional QTL cloning is a labor- and time-consuming process which requires the generation and analysis of large experimental mapping populations. An alternative to positional cloning of QTLs may be the allele mining approach, which is based on the knowledge of a gene's function in controlling a characteristic of interest on the one hand, and genetic co-localization of a functional candidate gene with QTL of interest on the other (Pflieger et al. 2001; Faino et al. 2011). However, in this approach substantial *a priori* knowledge is required. DNA variation for genes fulfilling these criteria has been examined in natural populations of accessions related by descent for associations with positive or negative characteristic values (Li et al. 2005; Gonzalez-Martınez et al. 2007). Finding such associations indicates that DNA variation either at the candidate locus itself or at a physically linked locus is causal for the phenotypic variation, but defined prove for the involvement of the gene is still circumstantial.

2.3 Techniques for Allele Mining

Dependent on the research question but also dependent on genetic, genomic and financial resources available, several techniques can be used for allele mining, ranging from a rapid and inexpensive polymerase chain reaction (PCR) to next gen sequencing and everything in between. For some applications (partial) sequence information or only molecular polymorphism of the alleles is sufficient. For other applications actual cloning of the entire allele is required. Generally, all DNA based tools require the careful selection of target genes. The target gene model might require verification, and successively, a careful design of selective primers will allow the

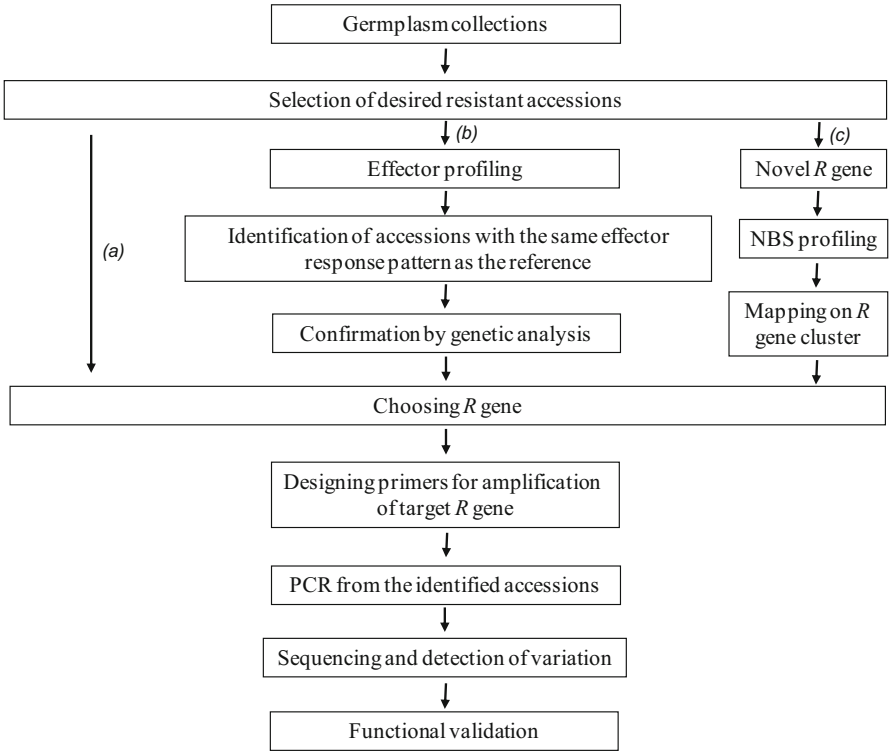

Fig. 2.3 Pipeline of allele mining of *R* genes. **a** Indicates allele mining in case that effector tools are not available for pathogen under study, **b** joint effector profiling-allele mining approach, and **c** novel *R* gene discovery in the combination of NBS profiling and allele mining

identification of novel alleles at candidate loci in the entire or core germplasm collection. In Fig. 2.3 a pipeline for novel allele discovery from germplasm collections is presented, including a combination of different approaches.

2.3.1 Molecular Tools for Allele Tagging

All molecular marker techniques include a PCR amplification of one or multiple alleles or paralogs. In order to identify polymorphisms between amplified alleles, single-strand specific nucleases could be applied. Using this technique, that is often used in TILLING approaches, nicking of heteroduplexes of PCR products can be easily detected. A recent development is the use of high resolution melting point analysis in order to screen for mismatches between amplified alleles in a high throughput fashion. Especially suitable for the highly polymorphic and duplicated *R* genes, is the NBS profiling technique (van der Linden et al. 2004). It is a powerful tool to

identify specific fragments of candidate *R* genes or *R* gene homologs throughout the genome by using degenerated primers that anneal to conserved sequences in the NBS domain of the NBS-LRR class of *R* genes. A high throughput application of this technique is to study fragment length polymorphisms as molecular markers. Also, PCR amplification of specific *R* genes is possible if primers are located in unique regions in order to target the specific paralog. The results may be visible as a DNA fragment of a specific size on an agarose gel. However, when gene specific markers are used in different germplasm material, often a-specific annealing of the primers can occur and therefore it will always be necessary to sequence the resulting PCR fragments to confirm their identity and homogeniety.

2.3.1.1 NBS Profiling

Many plant *R* genes are a member of a multigene cluster composed of multiple copies with high sequence similarity (Song et al. 2003). The NBS region of (NBS-LRR) *R* genes and their analogs (RGAs) contain highly conserved common motifs like the P-loop, the kinase −2 motif and the GLPL motif (Meyers et al. 2003; Monosi et al. 2004). These conserved motifs within the NBS-LRR genes have been used successfully to sequence (parts of) NBS regions from various plant species (Collins et al. 1998; Pflieger et al. 1999; Zhang and Gassmann 2007). NBS profiling uses the conserved motifs for efficient tagging of NBS-LRR type of *R* genes and their analogs (Van der Linden et al. 2004, 2005). The technique involves three different steps. (1) Digestion of genomic DNA with a restriction enzyme and ligation of adaptors to compatible restriction ends. (2) PCR amplification of NBS containing fragments using an NBS primer and an adaptor primer. (3) Separation of amplified fragments by polyacrylamide gel electrophoresis. The technique produces a multilocus profile of the genome.

NBS profiling can easily be adapted to target other conserved gene families, which is referred to as motif-directed profiling (Van der Linden et al. 2004, 2005). Also NBS profiling can be adapted to target particular *R* gene clusters. *R* genes from the same cluster usually have similarities in their sequences not shared with other *R* genes (McDowell and Simon 2006; Meyers et al. 2005), allowing the design of specific primers for a particular *R* gene cluster. NBS profiling could therefore also be adapted to reach high fragment saturation in an *R* gene cluster of interest (Verzaux et al. 2011, 2012; Jo et al. 2011). This technique is referred to as cluster directed profiling.

2.3.1.2 (Eco)-tilling

Eco tilling is a molecular method to screen germplasm core and mini collections. This technique is distinct from the TILLING approach since TILLING screens identify novel alleles that are induced by mutagenesis (Till et al. 2003; Barkley and Wang 2008) whereas eco-tilling identifies naturally occurring alleles in germplasm

(Barone et al. 2009). Both approaches employ a similar screening method to identify variation in alleles. Polymorphisms in PCR amplified DNA fragments are detected in hetroduplexes of the amplicons using single strand specific nucleases, high resolution melting point analysis or deep sequencing in next generation sequencing.

2.3.1.3 Amplification of Specific Allelic Variants

Family members with very similar sequences may have dispersed around the genome into non synthenous loci or may have remained within a genetic locus but has multiplicated resulting in tandem or inverted repeats. In general, sequences in coding regions will be more conserved than primers in flanking sequences. Dependent of the downstream application (sequence comparison, in plant expression), primers are chosen in- or outside coding sequence to amplify the entire gene, or part of the gene or only the open reading frame. Because even single nucleotide polymorphisms can be relevant differences between alleles, preferably the DNA polymerase will contain proofreading activity. Also, because often long stretches of the target gene are amplified, a long range polymerase chain reaction (LR-PCR) polymerase is preferred. Examples of enzymes that harbour both characteristics are *Pfu*-Turbo from Invitrogen or Phusion from Fermentas. One approach is to amplify the entire coding sequence of the *R* gene of interest using primers annealing to start and stop codon regions. Subsequently, the amplicon is sequenced and for expression studies it can be cloned in a vector that harbours heterologous regulatory sequences. For some accessions a possible lack of amplification can be expected due to absence of a coding gene or to low sequence homology at the primer annealing sites. A drawback of this approach is that the promotor and terminator regions of the novel alleles are missing, so variation in these regulatory regions are neglected. For 'true' allele mining, the use of primers matching the promotor and terminator regions is feasible when sequence conservation is sufficient. Accessions may also first be screened for the presence of the known *R* gene with a diagnostic molecular marker obtained from haplotype studies at the *R* gene locus and next for the presence of new alleles of the known *R* gene (Bhullar et al. 2009) to identify stronger alleles. Song et al. (2003) showed that allele mining could be used to clone the functional *RB* allele from a cluster with two highly similar paralogs. Also Wang et al. (2008) and Lokossou et al. (2010) could specifically amplify the target allele rather than paralogous genes in a *Rpi-blb1* allele mining study. Latha et al. (2004) exploited allele mining to identify stress tolerance genes in *Oryza* species and related germplasm. A common feature of the three genes investigated was that they were members of multigene families. Primers based on the 5' and 3' untranslated region of genes were found to be sufficiently conserved over the entire range of germplasm in rice to which the concept of allelism is applicable, while the primers based on the start and stop codon amplified sequences from additional loci (Latha et al. 2004).

is the cloning of the *Rpi-vnt1.1* gene (Pel et al. 2009). NBS profiling revealed a fragment that was co-segregating with resistance in a F1 population. The sequence of this NBS profiling band was similar to a known *R* gene (*Tm-2*^{2}). The mined allele

had a different genetic position on the same chromosome as the *Tm-2*^{2} gene. The entire coding sequence of the *Rpi-vnt1.1* allele was found after sequence analysis of a BAC clone derived from the genomic locus (Foster et al. 2009).

2.3.2 Next Generation Sequencing in Allele Mining

Currently, most genome and transcriptome sequencing projects, which used Sanger sequencing methodology in the past, are being replaced by next generation sequencing (NGS) technologies. These NGS technologies are able to generate data inexpensively and at a rate that is several orders of magnitude faster than that of traditional technologies (reviewed in Ercolano et al. 2012). At present there are several next generation sequencers on the market (Voelkerding et al. 2009). Most of these systems have different underlying biochemistries but all of these technologies sequence populations of PCR-amplified DNA molecules. The Heliscope and the PacBio, which sequence single molecules, are the exceptions. The amount of sequence data and the length of the reads are increasing with the continued development of the technology. Now resequencing and *de novo* sequencing of transcriptomes and genomes is becoming more and more accessible for individual labs (Varshney et al. 2009). This will lead to the discovery of novel useful variation which has been limiting the application of sequence-based selection in plants in the pre NGS era (Henry 2011). The availability of large numbers of genetic markers that can facilitate linkage mapping and whole genome scanning (WGS)-based association genetics that are of practical use for MAS in marker-deficient crops (Varshney et al. 2009). Resequencing of several genomes (Cao et al. 2011) followed by the comparison of all candidate *R* genes is now feasible in Arabidopsis (Guo et al. 2011). Soon this type of analysis will also be applied for crops and their wild relatives. Resequencing of parts of the genome with duplicated sequences, like *R* gene clusters, will remain a challenge, especially in heterozygous species like potato (Potato Genome Sequencing Consortium 2012). Single molecule sequencing will offer great opportunities for this research field (Koren et al. 2012).

2.3.3 Functional Analysis of Newly Identified Alleles

If the candidate genes have been identified in allele mining studies, it is required to confirm their functionality. Transient and stable transformations are valid for that purpose. Agroinfiltration, an *Agrobacterium tumefaciens*-based method, is currently the best developed and most reliable method for transient expression in plants (Vleeshouwers et al. 2011). Using this method *R* gene alleles and the *Avr* genes can be coexpressed in *N. benthamiana* (Bos et al. 2006) or other plant species (Rietman et al. 2010). Consequently, a hypersensitive response (HR) occurs in the infiltrated leaf area. This approach is only applicable in cases where the cognate *Avr* gene is

available. Agroinfiltration can also be carried out by expressing only the *R* gene in a host plant, followed by pathogen challenge inoculations (Lokossou et al. 2009; Pel et al. 2009). Another type of transient expression, which allows high throughput screenings is agroinfection. A gene of interest is cloned into a viral genome. Successively, the viral genome is introduced into plant cells through *A. tumefaciens*. Only a few cells need to be infected after which viral particles are formed that spread through the plant. Along with the virus, the gene of interest is expressed in the plant. There is, however, a limitation to the size of the gene to be expressed. Fragments over 500 bp in size will not express sufficiently.

The stable transformation into plant is still considered functional analysis which provides the most clear and definitive evidence. Transgenic plants can be tested for resistance in different developmental stages. For example, an *in vitro* inoculation assay was developed for routine high-throughput disease testing of *Phytophthora infestans* in potato (Huang et al. 2005).

2.4 Examples of Allele Mining in Solanum

As described in the previous section the technique of choice is highly dependent on the research question and application. Applications can be very diverse, ranging from very practical, like *R* gene mapping and cloning, to the identification of novel genetic resources, to more scientific applications like *R* gene geographic distribution and evolution. In this section examples of applications using the different techniques are presented.

2.4.1 Genetic Mapping

Mapping of *R* genes is strongly facilitated by allele mining through NBS profiling. Typical examples of *R* gene mapping approach are provided by Pel et al. (2009), Jacobs et al. (2010), Jo et al. (2011) and Verzaux et al. (2011, 2012). The first step in the approach consists of producing small ($n = 20$–100) populations segregating for *P. infestans* resistance, phenotyping the populations for resistance, and composing bulks of resistant and susceptible individuals. Then, the bulks are genotyped using NBS profiling to obtain markers that co-segregate with resistance. Next sequencing of co-segregating NBS fragments and BLAST analysis to identify the fragments is performed. Combining this information with literature and genome sequence data on mapping of resistance genes will suggest a putative map position. Finally, the map positions are confirmed using known flanking markers.

Jo et al. (2011) used NBS profiling and successive marker sequence comparison to the potato and tomato genome draft sequences to identify the genetic position of the late blight resistance gene *R8*. According to this work, *R8* was located on the long arm of chromosome IX and not on the short arm of chromosome XI as was suggested

previously by Huang et al. (2005). This is a first example where NBS markers could be directly landed in the sequenced (draft) genomes of potato and tomato. Through comparison of known markers in the tomato genetic map to the draft sequence, scaffolds were anchored to the tomato genetic map (anchored scaffold approach). Very recently, the *R9* mediated-late blight resistance was also mapped near the *R8* locus on chromosome IX using *R* gene cluster directed profiling approaches (Jo et al. in preparation).

2.4.2 Cloning Functional Alleles

Several late blight *R* genes have been cloned from potato wild relatives using allele- and paralog mining (for reviews see: Vleeshouwers et al. 2011, Rietman et al. 2010). Sometimes there is no clear distinction between allele- and paralog mining because of the high similarity among genes. An example of true allele mining was shown by Vleeshouwers et al. (2008) who isolated the functional alleles of *Rpi-blb1* present in *S. stoloniferum, Rpi-sto1* and *Rpi-pta1*. The entire genes were isolated by long range PCR using primers up and downstream of the coding regions. Specificity of the cloned genes was shown with different *P. infestans* isolates and with effector IpiO −1 and 2, which is recognized by *Rpi-blb1*, *Rpi-sto1* and *Rpi-pta1*. An allelic relationship between the three genes was also shown using marker (CT88) segregation studies (Wang et al. 2008). Sequence analyses showed that the putative functional homologs *Rpi-sto1* and *Rpi-pta1* are nearly identical to *Rpi-blb1*, with only 3 and 5 non-synonymous nucleotide substitutions inside the coding sequence, respectively.

A slightly different example of allele mining was provided by Lokossou et al. (2009), who described the map based cloning and functional characterization of *Rpi-blb3* and *Rpi-abpt,* which are allelic variants *R2* and *R2*-like. An allele mining strategy was employed using a start stopcodon approach. In this study a major technological improvement was made. The Gateway™ technology was used to clone the entire amplified coding sequences in a destination vector under the control of the *Rpi-blb3* promotor and terminator. A combination of efficient cloning of candidate alleles was combined with transient complementation assays in *Nicotiana benthamiana* and allowed for the rapid cloning and identification of *R2* and *R2*-like alleles.

Champouret used a similar technical approach to mine for *R3a* and *R2* alleles. A start stopcodon approach was pursued and the *R3a* screen revealed alleles with identical activity in distantly related species. This is considered as true allele mining. Also the *R2* screen revealed many genes with identical activity, however, also a few genes were identified which had slightly different recognition specificities, suggesting that not only alleles but also paralogs were mined. This is an example where allele mining and paralog mining are overlapping (Champouret 2010).

Paralog mining strategies can be pursued in order to facilitate map based cloning of novel *R* gene variants. An example of successful paralog mining came available using an *R2* mining approach applied on *S microdontum*. This resulted in the isolation of *Rpi-mcd1* which is functionally distinct from *R2* since the *Avr2* gene was not

recognised (Lokossou 2010). Another example is the cloning of the *Rpi-vnt1.1* gene (Pel et al. 2009). NBS profiling revealed a fragment that was co-segregating with resistance in a F1 population. The sequence of this NBS profiling band was similar to a known *R* gene (*Tm-2*^{2}). PCR amplification of *Tm-2*^{2} homologs identified the functional *Rpi-vnt1.1* gene. The mined allele had a different genetic position on the same chromosome as the *Tm-2*^{2} gene. Also the biological activity was different and therefore this study followed a typical paralog mining approach. This study also illustrates a risk associated with paralog mining in multigene families (Pel et al. 2009). Using the start stop codon primer pair derived from *Tm-2*^{2} only a part of the coding sequence was identified and a N-terminal extension, specific for the *Rpi-vnt1* alleles were overlooked. The entire coding sequence of the *Rpi-vnt1.1* allele was found after sequence analysis of a BAC clone derived from the genomic locus (Foster et al. 2009).

2.4.3 Uncovering Allelic Variation for Specific Genes

Allele mining can also be used to uncover genetic variation for a particular *R* gene and identify germplasm containing functional alleles from the same or different species. Nunziata et al. (2007) studied the variability of one cluster of genes at the *Gro1* locus responsible for resistance to *Globodera rostochiensis* race Ro1 in several potato species. The cluster is known to comprise 10 different paralogs, among which only the *Gro1–4* gene has been demonstrated to confer resistance against *Globodera rostochiensis* race Ro1. Using available sequence information, three primer pairs were designed that target different regions of the *Gro1* sequence. The first was designed in a highly conserved region and allowed the presence of at least one member of the gene cluster to be identified in 16 wild species analysed. The second primer pair was designed on a *Gro1–4* specific region and its use demonstrated that no gene identical to *Gro1–4* was present in any wild potato species analysed. Finally, the major part of the LRR coding sequence of the *Gro1* gene was amplified and sequenced in 16 wild species. In total, 409 SNPs were identified, varying between species from 12 SNPs in *S. demissum* to 35 in *S. stoloniferum*. These data could be used to identify evolutionary selection pressure since the non-synonymous/synonymous ratio (Ka/Ks) in most species was different from 1.

A similar type of screen was performed by Wang et al. (2008) and Lokossou et al. (2010). They analyzed the presence and allelic diversity of the late blight *R* genes *Rpi-blb1*, *Rpi-blb2* and *Rpi-blb3* in 196 different taxa of tuber-bearing *Solanum* species. The *Rpi-blb1* gene is part of a resistance gene analog (RGA) cluster of four members on chromosome VIII, *Rpi-blb2* resides in a locus harbouring at least 15 tomato *Mi* gene homologs on chromosome VI and *Rpi-blb3* originates from a cluster on chromosome IV. For all genes primers were design that would allow amplification of a specific fragment of the gene. The genes were only present in some Mexican diploid as well as polyploid species closely related to *S. bulbocastanum*, although differences in the distribution existed among the 3 genes. The *Rpi-blb1*

gene was only found in *S. bulbocastanum*, *S.cardiophyllum* subsp. *cardiophyllum*, and *S. stoloniferum*, the *Rpi-blb2* only in *S. bulbocastanum,* and the *Rpi-blb3* gene in *S. pinnatisectum*, *S. bulbocastanum* (including some subspecies), *S. hjertingii, S. nayaritense, S. brachistotrichum*, and *S. stoloniferum*. Sequence analysis of part of the *Rpi-blb1* and *Rpi-blb3* gene suggests an evolution through recombination and point mutations. For *Rpi-blb2* only sequences identical to the cloned gene were found, suggesting that it has emerged recently. The three *R* genes occurred in different combinations and frequencies in *S. bulbocastanum* accessions and their spread is confined to Central America (Lokossou et al. 2010). A practical outcome of the allele mining study by Wang et al. (2008) was the discovery of conserved homologues of *Rpi-blb1* in an EBN 2 tetraploid potato species, e.g. *S. stoloniferum*. The *Rpi-blb1* is present in the diploid tuber-bearing *S. bulbocastanum*, which is not directly crossable with the tetraploid *S. tuberosum*. *Solanum stoloniferum* can be directly crossed to cultivated potato, thus facilitating an easy transfer of a gene with exactly the same specificity and functionality as *Rpi-blb1*.

An allele mining approach to identify variation in the *Avr9* recognizing *Cf-9* alleles provided evidence for the presumed evolutionary mechanism driving *R* gene diversification. Subsequent intra- and intergenic unequal recombination events were held responsible for the sequence diversification of *Cf-9* alleles. However, this diversification was not accompanied by a functional diversification since the Avr9 effector could still be recognized (Van der Hoorn et al. 2001; Kruijt et al. 2004).

2.4.4 Alleles in Natural Populations of Solanum

Knowledge on the evolution and distribution of disease resistance genes is important for a better understanding of the dynamics of these genes in nature. Caicedo (2008) studied geographic diversity cline of *R* gene homologs in natural populations of *Solanum pimpinellifolium* L., a wild relative of cultivated tomato, to determine the possible roles of demography and selection on *R* gene evolution. The patterns of diversity at the multigenic *Cf-2* gene family were investigated which consisted of 26 closely related homologs, referred to as the Hcr2-p family (Caicedo and Schaal 2004). The 26 Hcr2-p homologs display length variation due primarily to variation in the number of LRR-coding units within each gene and can be classified into nine different size classes according to length; within size-classes, homologs differ from each other by one or a few single nucleotide polymorphisms (SNPs). *Solanum pimpinellifolium* individuals vary extensively in the number of Hcr2-p homologs they carry, with Southern blots results suggesting 1–5 genes per individual (Caicedo and Schaal 2004). Species-wide analyses of Hcr2-p sequence diversity suggest that selection has played a role in the evolution of the gene family. Patterns of amino acid substitution are consistent with purifying selection in the 5′ LRR-coding portion of the genes and positive selection on some amino acid residues in the 3′ region. Evolutionary relationships among homologs also suggest that balancing selection has shaped species-wide patterns of diversity. Studies on patterns of diversity at the

multigenic *Cf-2* gene family in *S. pimpinellifolium* populations along the northern coast of Peru showed that population diversity levels of *Cf-2* homologs follow a latitudinal cline, consistent with the species' history of gradual colonization of the Peruvian coast and population variation in outcrossing.

In another approach the wild tomato germplasm was screened for responsiveness to the Avr4 and Avr9 effectors from *C. fulvum*, which are recognized by the Cf-4 and Cf-9 R proteins respectively. Recognition and the presence of the matching *R* genes was ubiquitous throughout the screened germplasm. This allele mining approach showed that *C. fulvum* is an ancient pathogen of the genus *Lycopersicon* (Kruijt et al. 2005).

Several studies have now clearly shown the potential of allele mining in *Solanum* for the improvement of disease resistance. A large number of allelic variants of known disease resistance genes have been discovered and in several cases also functionality of the variants was shown. As was shown, allele mining can strongly facilitate the cloning of *R* genes by using comparative genomics approaches. Allele mining has also been shown to be useful for the identification of orthologous sequences in species that are more easily crossable with the cultivated material then the species the gene was originally discovered in, thus facilitating a more rapid deployment of genes in breeding programs. Furthermore allele mining was useful for the identification of novel, yet unknown *R* genes and shed light on evolutionary processes related to these genes. As more and more *R* genes are identified and cloned, the chances increase that new *R* genes reside at known and well-characterized loci, enabling the use of comparative genomics and, thus, the development of efficient allele mining strategies.

The availability of the potato and tomato genome sequences, together with a constant drop in the sequencing cost will boost allele mining even more. The fast (r)evolution in the high throughput sequencing technologies, especially the increase in read lengths expected from the third generation of single molecule sequencing platforms, will provide a complete survey of the distribution of *R* gene clusters in the *Solanaceae* family, enabling a dramatic acceleration in the process of identifying agronomically important genes like novel *R* genes.

We envisage that novel and efficient 'mining' strategies can give direct access to disease resistance genes of interest using next generation sequencing in combination with effector genomics. However, as not all effectors are known yet more effort should be made in that area. Another interesting research area relates to the durability of the *R* genes. At present it is unknown whether all allelic variants discovered for a particular gene are equally easily overcome by the pathogen. If not, this may be a way to identify more durable genes.

Acknowledgements JV was supported by the DuRPh program funded by the Ministry of Agriculture in the Netherlands (now Ministry of EL&I). KRJ was financially supported by the international program BO −10–010-112 program of the Ministry of EL&I and the EuropeAid program 128275/C/ACT/KP2 project DCI-FOOD/2009/218–671.

References

Bai Y, Huang CC, van der Hulst R et al (2003) QTLs for tomato powdery mildew resistance (*Oidium lycopersici*) in *Lycopersicon parviflorum* G1.1601 co-localize with two qualitative powdery mildew resistance genes. Mol Plant Microbe Interact 16:169–176

Barkley NA, Wang ML (2008) Application of TILLING and EcoTILLING as reverse genetic approaches to elucidate the function of genes in plants and animals. Curr Genomics 9:212–226

Barone A, Di Matteo A, Carputo D, Frusciante L (2009) High-throughput genomics enhances tomato breeding efficiency. Curr Genomics 10:1–9

Bhullar NK, Street K, Mackay M et al (2009) Unlocking wheat genetic resources for the molecular identification of previously undescribed functional alleles at the *Pm3* resistance locus. Proc Natl Acad Sci U S A 106:9519–9524

Bos JIB, Kanneganti TD, Young C et al (2006) The C-terminal half of *Phytophthora infestans* RXLR effector *AVR3*a is sufficient to trigger *R3a*-mediated hypersensitivity and suppress *INF1*-induced cell death in *Nicotiana benthamiana*. Plant J 48:165–176

Caicedo AL (2008) Geographic diversity cline of *R* gene homologs in natural populations of *Solanum pimpinellifolium* (Solanaceae). Am J Bot 95:393–398

Caicedo AL, Schaal BA (2004) Heterogeneous evolutionary processes affect *R* gene diversity in natural populations of *Solanum pimpinellifolium*. Proc Natl Acad Sci U S A 101:17444–17449

Cao J, Schneeberger K, Ossowski S et al (2011) Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. Nat Genet 43:956–963

Champouret (2010) Functional genomics of *Phytophthora infestans* effectors and *Solanum* resistance genes. PhD thesis, Wageningen University

Dodds PN, Lawrence GJ, Ellis JG (2001) Six amino acid changes confined to the leucine-rich repeat β-strand/β-turn motif determine the difference between the *P* and *P2* rust resistance specificities in flax. Plant Cell 13:495–506

Dubery IA, Sanabria NM, Huang JC (2012) Self and nonself. Advances in experimental medicine biology 738:79–107. doi:10.1007/978-1-4614-1680-7_6

Ercolano MR, Sanseverino W, Carli P et al (2012) Genetic and genomic approaches for R-gene mediated disease resistance in tomato: retrospects and prospects. Plant Cell Rep 31:973–985

Faino L, Azizinia S, Hassanzadeh BH et al (2012) Fine mapping of two major QTLs conferring resistance to powdery mildew in tomato. Euphytica 184:223–234

Foster SJ, Park TH, Pel MA et al (2009) *Rpi-vnt1.1*, a *Tm-2* homolog from *Solanum venturii* confers resistance to potato late blight. Mol Plant Microbe Interact 22:589–600

Frankel OH, Brown ADH (1984) Plant genetic resources today: a critical appraisal. In: Holden JHW, Williams JH (eds) Crop genetic resources: conservation & evaluation. G. Allen, London, pp. 249–257

Gebhardt C, Valkonen JPT (2001) Organization of genes controlling disease resistance in the potato genome. Annu Rev Phytopathol 39:79–102

Grube RC, Radwanski ER, Jahn M (2000) Comparative genetics of disease resistance within the *Solanaceae*. Genetics 155:873–887

Guo YL, Fitz J, Schneeberger K et al (2011) Genome-wide comparison of nucleotide-binding site-leucine-rich repeat-encoding genes in *Arabidopsis*. Plant Physiol 157:757–769

Gupta PK (2008) Single-molecule DNA sequencing technologies for future genomics research. Trends Biotechnol 26:602–611

Henry RJ (2011) Next-generation sequencing for understanding and accelerating crop domestication. Brief Funct Genomics [Epub. PMID:22025450]

Hoekstra R (2009) Exploring the natural biodiversity of potato late blight resistance. Potato Res 52:237–244

Hofinger BJ, Jing HC, Kim EHK, Kanyuka K (2009) High-resolution melting analysis of cDNA-derived PCR amplicons for rapid and cost-effective identification of novel alleles in barley. Theor Appl Genet 119:851–865

Huang S (2005) Discovery and characterization of the major late blight resistance complex in potato. PhD thesis, Wageningen University, Wageningen

Hulbert SH, Webb CA, Smith SM, Sun Q (2001) Resistance gene complexes: evolution and utilization. Annu Rev Phytopathol 9:285–312

Imelfort M, Batley J, Grimmond S, Edwards D (2009) Genome sequencing approaches and successes. In: Somers DJ et al (eds) Methods in molecular biology, plant genomics, vol 513. Humana Press, pp. 345–358

Jacobs MMJ, Vosman B, Vleeshouwers VGAA et al (2010) A novel approach to locate *Phytophthora infestans* resistance genes on the potato genetic map. Theor Appl Genet 120:785–796

Jo KR, Arens M, Kim TY et al (2011) Mapping of the *S. demissum* late blight resistance gene *R8* to a new locus on chromosome IX. Theor Appl Genet 123:1331–1340

Kaur N, Street K, Mackay M et al (2008) Molecular approaches for characterization and use of natural disease resistance in wheat. Eur J Plant Pathol 121:387–397

Kim HJ, Lee HR, Jo KR et al (2012) Broad spectrum late blight resistance in potato differential set plants Ma*R8* and Ma*R9* is conferred by multiple stacked *R* genes. Theor Appl Genet 124:5

Knapp S (2002) Tobacco to tomatoes: a phylogenetic perspective on fruit diversity in the *Solanaceae*. J Exp Bot 53:2001–2022

Koren S, Schatz MC, Walenz BP et al (2012) Hybrid error correction and de novo assembly of single-molecule sequencing reads. Nat Biotechnol 30:693–700

Kuang H, Woo SS, Meyers BC et al (2004) Multiple genetic processes result in heterogeneous rates of evolution within the major cluster disease resistance genes in lettuce. Plant Cell 16:2870–2894

Kruijt M, Brandwagt BF, de Wit PJ (2004) Rearrangements in the *Cf-9* disease resistance gene cluster of wild tomato have resulted in three genes that mediate Avr9 responsiveness. Genetics 168:1655–1663

Kruijt M, Kip DJ, Joosten MH et al (2005) The *Cf-4* and *Cf-9* resistance genes against *Cladosporium fulvum* are conserved in wild tomato species. Mol Plant Microbe Interact 18:1011–1021

Latha R, Rubia L, Bennett J, Swaminathan MS (2004) Allele mining for stress tolerance genes in *Oryza* species and related germplasm. Mol Biotech 27:101–108

Li G, Huang S, Guo X et al (2011) Cloning and characterization of *R3b*; members of the *R3* superfamily of late blight resistance genes show sequence and functional divergence. Mol Plant Microbe Interact 24:1132–1142

Lokossou (2010) Dissection of the major late blight resistance cluster on potato linkage group IV. PhD thesis, Wageningen University

Lokossou AA, Park TH, van Arkel G et al (2009) Exploiting knowledge of *R/Avr* genes to rapidly clone a new LZ-NBS-LRR family of late blight resistance genes from potato linkage group IV. Mol Plant Microbe Interact 22:630–641

Lokossou AA, Rietman H, Wang M et al (2010) Diversity, distribution and evolution of *Solanum bulbocastanum* late blight resistance genes. Mol Plant Microbe Interact 23:1206–1216

McDowell JM, Simon SA (2006) Recent insights into *R* gene evolution. Mol Plant Pathol 7:437–448

Meyers BC, Kaushik S, Nandety RS (2005) Evolving disease resistance genes. Curr Opin Plant Biol 8:129–134

Meyers BC, Kozik A, Griego A et al (2003) Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. Plant Cell 15:809–834

Millett BP, Bradeen JM (2007) Development of allele-specific PCR and RT-PCR assays for clustered resistance genes using a potato late blight resistance transgene as a model. Theor Appl Genet 114:501–513

Nunziata A, Ruggieri V, Frusciante L, Barone A (2007) Allele mining at the locus *Gro 1* in *Solanum* wild species. In: VI International Solanaceae Conference 449–456

Park TH, Vleeshouwers VGAA, Jacobsen E et al (2009) Molecular breeding for resistance to *Phytophthora infestans* (Mont.) de Bary in potato (*Solanum tuberosum* L.): a perspective of cisgenesis. Plant Breed 128:109–117

Pel MA, Foster SJ, Park TH et al (2009) Mapping and cloning of late blight resistance genes from *Solanum venturii* using an interspecific candidate gene approach. Mol Plant Microbe Interact 22:601–615

Pflieger S, Lefebvre V, Caranta C et al (1999) Disease resistance gene analogs as candidates for QTLs involved in pepper/pathogen interactions. Genome 42:1100–1110

Pflieger S, Palloix A, Caranta C et al (2001) Defense response genes co-localize with quantitative disease resistance loci in pepper. Theor Appl Genet 103:920–929

Potato Genome Sequencing Consortium, Xu X, Pan S, Cheng S et al (2011) Genome sequence and analysis of the tuber crop potato. Nature 475:189–195

Ramkumar G, Sakthivel K, Sundaram RM et al (2010) Allele mining in crops: prospects and potentials. Biotechnol Adv 28:451–461

Rietman H, Bijsterbosch, Liliana MC et al (2012) Qualitative and quantitative late blight resistance in the potato cultivar Sarpo Mira is determined by the perception of five distinct RXLR effectors. Mol Plant Microb Interact 25:910–919

Rietman H, Champouret N, Hein I et al (2010) Plants and oomycetes, an intimate relationship: co-evolutionary principles and impact on agricultural practice. Hemming D (ed), CAB Reviews, CABI, UK. June 2011 5:1–17

Sanchez MJ, Bradeen JM (2006) Towards efficient isolation of *R* gene orthologs from multiple genotypes: optimization of long range-PCR. Mol Breed 17:137–148

Sanseverino W, Roma G, De Simone M et al (2010) PRGdb: a bioinformatics platform for plant resistance gene analysis. Nucl Acids Res 38:D814–D821

Sanz MJ, Loarce Y, Fominaya A et al (2012) Identification of RFLP and NBS/PK profiling markers for disease resistance loci in genetic maps of oats. Theor Appl Genet. doi:10.1007/s00122-012-1974-8

Sato S, Tabata S, Mueller LA et al (2012) The tomato genome sequence provides insights into fleshy fruit evolution. Nature 485:635–641

Schneeberger K, Weigel D (2011) Fast-forward genetics enabled by new sequencing technologies. Trends Plant Sci 6:282–288

Song J, Bradeen JM, Naess SK et al (2003) Gene *RB* cloned from *Solanum bulbocastanum* confers broad spectrum resistance to potato late blight. Proc Natl Acad Sci U S A 100:9128–9133

Spooner DM, Hijmans RJ (2001) Potato systematics and germplasm collecting, 1989–2000. Amer J Potato Res 78:237–268

Spooner DM, Peralta IE, Knapp S (2005) Comparison of AFLPs with other markers for phylogenetic inference in wild tomatoes [*Solanum* L. section *Lycopersicon* (Mill.) Wettst.]. Taxon 54:43–61

Till BJ, Colbert T, Tompa et al (2003) High-throughput TILLING for functional genomics. In: Grotewold E (ed) plant functional genomics: methods and protocols. Methods in molecular biology, vol. 236. Human Press, Totowa, NJ, pp 205–220

Trognitz F, Manosalva P, Gysin R et al (2002) Plant defense genes associated with quantitative resistance to potato late blight in *Solanum phureja* × dihaploid *S. tuberosum* hybrids. Mol Plant Microbe Interact 15:587–597

Upadhyaya HD, Gowda CLL, Buhariwalla HK, Crouch JH (2006) Efficient use of crop germplasm resources: identifying useful germplasm for crop improvement through core and mini-core collections and molecular marker approaches. Plant Genetic Resour: Charac Util 4:25–35

Van der Hoorn RA, Kruijt M, Roth R et al (2001) Intragenic recombination generated two distinct *Cf* genes that mediate AVR9 recognition in the natural population of *Lycopersicon pimpinellifolium*. Proc Natl Acad Sci U S A 98:10493–10498

van der LCG, Wouters D, Mihalka V et al (2004) Efficient targeting of plant disease resistance loci using NBS profiling. Theor Appl Genet 109:384–393

van der Vossen EAG, van der Voort JNAMR, Kanyuka K et al (2000) Homologues of a single resistance-gene cluster in potato confer resistance to distinct pathogens: a virus and a nematode. Plant J 23:567–576

Varshney RK, Andreas GA, Sorrells ME (2005) Genomics assisted breeding for crop improvement. Trends Plant Sci 10:621–630

Verzaux E, Budding D, de Vetten N et al (2011) High resolution mapping of a novel late blight resistance gene *Rpi-avl1*, from the wild Bolivian species *Solanum avilesii*. Am J Potato Res 88:511–519

Verzaux E, van Arkel G, Vleeshouwers VGAA et al (2012) High-resolution mapping of two broad-spectrum late blight resistance genes from two wild species of the *Solanum circaeifolium* group. Potato Res 55:109–123

Vleeshouwers VGAA, Raffaele S, Vossen JH et al (2011a) Understanding and exploiting late blight resistance in the age of effectors. Annu Rev Phytopathol 49:507–531

Vleeshouwers VGAA, Finkers R, Budding D et al (2011b) SolRgene: an online database to explore disease resistance genes in tuber-bearing *Solanum* species. BMC Plant Biol 11:116

Vleeshouwers VGAA, Rietman H, Krenek P et al (2008) Effector genomics accelerates discovery and functional profiling of potato disease resistance and *Phytophthora infestans* avirulence genes. PLoS ONE 3:e2875

Voelkerding KV, Dames SA, Durtschi JD (2009) Next-generation sequencing: from basic research to diagnostics. Clin Chem 55:641–658

Wang M, Allefs S, van den Berg R et al (2008) Allele mining in *Solanum*: conserved homologues of *Rpi-blb1* are identified in *Solanum stoloniferum*. Theor Appl Genet 116:933–943

Zhang XC, Gassmann W (2007) Alternative splicing and mRNA levels of the disease resistance gene *RPS4* are induced during defense responses. Plant Phys 145:1577–1587

Chapter 3
Dissection of Potato Complex Traits by Linkage and Association Genetics as Basis for Developing Molecular Diagnostics in Breeding Programs

Christiane Gebhardt, Claude Urbany and Benjamin Stich

Abstract Most characters relevant for the development of commercial potato varieties are complex, meaning that they are controlled by multiple genetic and environmental factors. Over the past 20 years, potato complex traits have been dissected by linkage mapping in quantitative trait loci (QTL) using DNA-based markers. QTL mapping was performed in genetically diverse diploid and in few tetraploid biparental progeny. The integration of various QTL maps revealed regions in the potato genome where genes controlling complex traits are located. This in combination with mapping genes known to functionally contribute to a given phenotype identified candidate loci for complex traits. More recently, association mapping entered the stage and resulted in the identification of the first markers diagnostic for potato complex traits in populations of tetraploid cultivars related by descent. Association mapping based on DNA polymorphisms in functional and positional candidate genes identified loci that might be causal for natural variation of resistance to late blight or tuber quality traits such as processing quality and bruising resistance.

3.1 Introduction

The cultivated potato (*Solanum tuberosum*) is with around 330 million tons produced per year the world's fourth most important crop after rice, wheat and maize. Currently, China, India and the Russian Federation are the largest producers and together generate 42 % of the world potato harvest (http://faostat.fao.org/site/339/default.aspx). Potatoes are traditionally used for direct consumption as staple food or vegetable (McGregor 2007), for processed products such as chips, snacks and French fries (Kirkman 2007), for industrial production of starch and alcohol and for animal feed. Unconventional uses as a source of biofuel or as bioreactor for producing pharmaceutical compounds in transgenic plants might gain importance in the future. Potato breeding programs are targeted at the selection and distribution of new cultivars with

C. Gebhardt (✉) · C. Urbany · B. Stich
MPI for Plant Breeding Research, Carl von Linné Weg 10, 50829 Cologne, Germany
e-mail: gebhardt@mpipz.mpg.de

R. Tuberosa et al. (eds.), *Genomics of Plant Genetic Resources,*
DOI 10.1007/978-94-007-7575-6_3,

optimal adaptation to different environments and end uses. This process requires the evaluation of multiple plant and tuber characteristics in multi-year and location trials. Besides characters of general importance such as tuber yield, tuber starch content, resistance to pests and diseases and plant maturity type, there are many traits considered for specific production areas, markets and end uses. For example, table ware potatoes are evaluated and selected in addition to the above traits for tuber shape, skin and flesh color, eye depth, texture, taste, cooking type, after cooking darkening and resistance to bruising. The majority of the phenotypic characters that are relevant for developing commercial cultivars are complex, which means that they are controlled by multiple genetic and environmental factors (Mackay 2001). Whereas some traits are highly heritable and easy to evaluate at phenotypic level (e.g. color traits, tuber shape, plant maturity), the assessment of other traits is difficult by various reasons, for example due to the quarantine status of a pathogen (e.g. potato wart), the ambiguity, time and labour costs of the phenotypic assay (e.g. nematode resistance, wart resistance) or the requirement for a destructive test of large numbers of tubers, which become available only after several years of vegetative multiplication (e.g. tuber bruising). The 'difficult' traits are those, for which diagnostic DNA-based markers are most welcomed by breeders. Such markers should be cost effective, easy to use and suitable for the early identification of parents and progeny carrying superior trait alleles (precision breeding). The diagnostic power of a DNA marker depends on the frequency of recombination over multiple meiotic generations between a specific marker allele and the trait allele (linkage disequilibrium, LD). The ultimate diagnostic marker originates directly from the allelic variant of the gene that is causal for the phenotypic effect, as this reduces the recombination frequency to zero. As the genes and their allelic variants responsible for complex phenotypes are at present mostly unknown, markers in strong linkage disequilibrium with specific trait alleles are the alternative. The identification of DNA-based markers diagnostic for a trait of interest requires linkage mapping in bi-parental populations and association mapping in populations of genotypes related by descent. In this chapter we summarize the approaches and achievements in dissecting complex potato traits with DNA-based markers and identifying genes underlying quantitative trait loci (QTL). We also try to assess the current status of precision breeding in potato based on published results, knowing that commercial successes of marker assisted selection hardly appear in the scientific literature.

3.2 Methods for the Genetic Dissection of Complex Traits in Potato

Linkage mapping as well as association mapping use the LD between molecular markers and functional loci to link phenotypic trait variation with molecular marker variation (Flint-Garcia et al. 2003). The difference between both approaches is that linkage mapping methods use the LD generated by the mating design involving

two parents, while association mapping uses the historical LD that is present in a germplasm set of genotypes related with each other by common ancestors.

3.2.1 Linkage Mapping

The majority of cultivated potatoes (*Solanum tuberosum* ssp *tuberosum*, *Solanum tuberosum* ssp *andigena*), are tetraploids with tetrasomic inheritance. Tetraploid potatoes possess four sets of 12 chromosomes each. Genetic variation is generated by crossing different heterozygous parents, which is then fixed by vegetative propagation of F1 genotypes via tubers. Tetraploid, heterozygous genotypes carry one to four different alleles per locus. At a bi-allelic locus, for example a single nucleotide polymorphism (SNP), one of five possible genotypes occurs: homozygous *AAAA or BBBB,* or heterozygous *ABBB, AABB* or *AAAB*. An individual with three alleles at, for example, a microsatellite locus, can have the genotypes *AABC, ABBC* and *ABCC*, whereas an individual with four alleles at a given locus is always *ABCD*. Many genetic models are therefore possible for the segregation of the alleles at two loci in the F1 progeny of a cross between heterozygous tetraploid parents, which complicates linkage analysis. Only few of these models are suitable for linkage analysis in tetraploids, which are, in principle, based on the segregation of dominant alleles: *A000* × *0000* (simplex by nulliplex, expected segregation ratio 1:1), *AA00* × *0000* (duplex by nulliplex, expected segregation ratio 5:1) and *A000* × *A000*(simplex by simplex, expected segregation ratio 3:1). Methods for linkage and QTL mapping in tetraploids have been developed based on these models and incorporated in the TetraploidMap software (Hackett et al. 2003, 2001, 2007; Luo et al. 2001). Full coverage of the 96 chromosomes, which segregate in a tetraploid F1 progeny requires essentially the construction of 96 linkage groups, 48 (4 × 12) for each parent. This coverage can only be achieved by genotyping with a very large number of markers. Amplified fragment length polymorphisms (AFLP) and microsatellite markers were used to construct a linkage map in one tetraploid mapping population (12601ab1 × 'Sterling'). Linkage to QTL was estimated by single marker analysis (Bradshaw et al. 1998; Meyer et al. 1998) or interval mapping (Bradshaw et al. 2004, 2008; Bryan et al. 2004). A simplified approach to QTL mapping in tetraploids is QTL tagging, which does not require linkage map construction. Phenotyped segregating populations are genotyped with markers of known map position, which are then tested for linkage with QTL using analysis of variance or the *t*-test (Bormann et al. 2004). In the future, high-throughput genotyping with SNP genotyping chips (Hamilton et al. 2011) has the potential to achieve high genome coverage and thereby will greatly facilitate linkage and QTL mapping in tetraploid populations.

The ploidy of tetraploid potatoes can be reduced to the diploid level (Hougas et al. 1964; Powell and Uhrig 1987). Due to the complexity of linkage mapping in tetraploid species, the majority of QTL studies in potato have therefore been conducted based on diploid, bi-parental mapping populations. Diploid potatoes are largely self-incompatible (gametophytic self-incompatibility) and therefore highly

heterozygous. The genetics of diploid potato is equivalent to human genetics: Partially heterozygous parents generate F1 offspring, in which two, three or four alleles segregate per locus. Alleles are linked to each other either in coupling or repulsion phase. The heterozygocity of the parents allows the construction of two linkage maps in a single F1 mapping population, based on meiotic recombination in the female and the male parent (Ritter et al. 1990). QTL are identified and mapped either based on single marker tests or on interval mapping, (e.g. Schäfer-Pregl et al. 1998). In general, highly significant marker-QTL linkages are detected similarly by both methods. The MAPMAKER/QTL software (Lincoln and Lander 1989) has been used for linkage and QTL mapping in diploid populations, (e.g. Bonierbale et al. 1994). Currently, the most popular software packages for linkage and QTL mapping in out-crossing diploid species such as potato are JoinMap (Stam 1993) and MapQTL® (Van Ooijen and Maliepaard 1996), which was used, for example, by Werij et al. (2007).

3.2.2 Association Mapping

Association mapping examines the joint inheritance of functional polymorphisms and physically linked molecular markers in a set of genotypes with unknown ancestry. As the unknown ancestry can extend across thousands of generations, the joint inheritance will only persist for very closely linked polymorphisms. Hence, association mapping exploits historical recombination events. By exploring deeper population genealogy rather than family pedigrees, association mapping offers mainly four advantages over linkage mapping: higher mapping resolution, higher number of alleles evaluated, broader reference population and less time in establishing an association. In contrast to linkage mapping populations, the LD in association mapping populations might not only be influenced by recombination but also by various other forces (Flint-Garcia et al. 2003). The forces influencing the pattern and extent of LD are: (1) mating type, (2) genetic drift, (3) selection, (4) mutation, (5) population substructure and relatedness and (6) ascertainment bias (Clark et al. 2005; Stich et al. 2005; Yu et al. 2006). For most association mapping populations, the importance of these factors is unknown or can only be roughly estimated. The extent of LD can therefore not be derived theoretically but must be inferred from empirical studies based on molecular markers in order to evaluate the applicability and resolution of association mapping approaches. A first estimate of LD in potato (Simko et al. 2006b) indicated that the r^2 value reached 0.21 within about 1 kb distance and 0.10 at a distance of about 10 cm. The DNA fragments from which the SNPs of that study were derived were end-sequences of BACs containing *R* genes from wild species. This may have led to an overestimation of LD due to the few meioses since the introgression of *R* genes (D'hoop et al. 2010). D'hoop et al. (2010) found that LD decayed on average within about 5 cm to r^2 values of 0.10. However, this study was based on AFLPs, the position of which on the physical map cannot be easily inferred. Nevertheless, these results suggested that the extent of LD in potato is considerably higher than in

various other species (reviewed by Ersoz et al. 2007) and, thus, association mapping can be performed with a relatively low number of markers.

Two different types of association mapping experiments can be distinguished: (1) family-based association mapping experiments and (2) population-based association mapping experiments. In human genetics, family-based association mapping tests, such as the transmission disequilibrium test (TDT) (Spielman et al. 1993) and for quantitative traits the quantitative TDT (QTDT) (Allison 1997) have been developed to detect marker-phenotype associations while correcting for population structure. These approaches are based on a large number of so called family samples. In these family samples, the parental individuals and their progeny are genotyped, whereas the progeny is phenotyped as well. This constellation allows performing valid tests of association in the presence of linkage. Family-based association tests have also been developed for a plant genetics context (Stich et al. 2006). In potato however, mainly population-based association mapping approaches have been applied so far, in which arbitrary germplasm sets such as collections of varieties and breeding lines are used for association analysis (e.g. Li et al. 2008; Simko et al. 2004). Therefore, only the latter is discussed in the following.

Population structure and familial relatedness potentially create LD between unlinked loci (Pritchard 2001; Stich et al. 2005; Yu et al. 2006). This increases the proportion of false-positive marker-phenotype associations to a level which is considerably higher than the nominal type I error level (Thornsberry et al. 2001). The results from LD analyses outlined above but also the information from pedigrees of potato varieties (http://www.plantbreeding.wur.nl/potatopedigree/lookup.php) indicated that familial relatedness is an important factor influencing LD in potato germplasm and thus has to be considered during population based association analyses. On the other hand, breeders usually neglect information of subgroup assignment when selecting the parents of new crosses, which leads to the absence of strong subpopulations in elite potato breeding material (Li et al. 2005a, 2008). Nevertheless, the comparison of mixed-model approaches for association mapping in rapeseed, potato, sugar beet, maize and Arabidopsis (Stich and Melchinger 2009) suggested that even in such material the consideration of population structure in addition to familial relatedness improves the adherence to the nominal type I error rate and increases the statistical power to detect marker-phenotype associations.

Population substructure and familial relatedness can be inferred from pedigree information or by genotyping the association mapping panel with a set of selectively neutral background markers distributed throughout the genome. As pedigree information is neither always available nor reliable in potato, the marker approach is considered more appropriate. In potato, pedigree information (Malosetti et al. 2007; Simko et al. 2004), AFLPs (D'hoop et al. 2010; Li et al. 2005a) and microsatellites (Pajerowska-Mukhtar et al. 2009) have been used to assess population structure. The results of Pajerowska-Mukhtar et al. (2009) suggested that for a panel of 184 tetraploid elite breeding genotypes between 25 and 30 SSR markers were sufficient to get stable results about population structure and familial relatedness. AFLP markers, as a result of their dominant inheritance demand special statistical methods if used to estimate population genetic parameters (Falush et al. 2007). Given their higher

genome density, lower mutation rate and better amenability to high-throughput detection systems, SNPs are rapidly becoming the markers of choice for complex trait dissection studies in various crops (Hamblin et al. 2007). Principal component analysis has been used for long time to analyse genetic diversity and was recently proposed as a fast and effective way to diagnose population structure in an association mapping context (Patterson et al. 2006; Price et al. 2006). Promising results, also for potato, were obtained for mixed-model approaches in which the matrix from STRUCTURE (Pritchard et al. 2000), the estimation of which is computationally demanding (Price et al. 2006), was replaced by principal components (Stich and Melchinger 2009).

The test statistics for detecting association between a molecular marker and a complex trait in tetraploid potato has quickly progressed from a simplistic *t*-test or Mann-Whitney-U test (Gebhardt et al. 2004; Simko et al. 2004) to state of the art statistical approaches using mixed models (Pajerowska-Mukhtar et al. 2009; Stich and Gebhardt 2011). The mixed-model approach takes into account population structure and familial relatedness (Yu et al. 2006), which are estimated from random markers. This approach was most appropriate when comparing various methodologies for association mapping in potato and other crops (Stich and Melchinger 2009).

3.3 Resistance to Pests and Diseases

Potato is prone to numerous diseases, which can be, in contrast to seed propagated crops, transmitted to the tubers and therefore affect not only crop yield but also the next vegetative generation (Mulder and Turkensteen 2005). Disease resistance is conferred by single dominant genes (*R* genes) and/or by multiple genes, the latter encoded at quantitative trait loci (QTL). Quantitative resistance can be considered as the natural variation of a compatible interaction between plant host and pathogen, whereas *R* genes trigger an incompatible interaction with no visible symptoms or local necrotic lesions (immunity or hypersensitive response). The molecular basis of compatible and incompatible host-pathogen interactions has been extensively studied in the model plant *Arabidopsis thaliana* and in crop plants, among others in potato. Genes functional in pathogen recognition, defense signalling and defense response have been functionally characterized (Garcia-Brugger et al. 2006; Grant and Lamb 2006; Takken and Joosten 2000; van Loon et al. 2006). These functional studies are the knowledge base for the candidate gene approach to identify genes controlling quantitative resistance. The implicit assumption here is that the same genes underlay qualitative as well as quantitative resistance, and that allelic variants of some of those genes are causal for the quantitative variation of resistance.

3.3.1 Late Blight: Phytophthora Infestans

The most important pathogen in potato cultivation world wide is the oomycete *Phytophthora infestans*, which causes the late blight disease on foliage, stems and tubers

(Kamoun and Smart 2005). The search for plant resistance began after the catastrophic late blight epidemics that destroyed the potato crop in temperate Europe in the middle of the 19th century, causing famine in Ireland and other European countries. Resistance to late blight has been and still is found in wild potato species native to Mexico and Latin America and was introgressed in the cultivated potato. Monogenic resistance (*R* genes) to late blight was however not durable under the field conditions of modern agriculture (Wastie 1991), although 'defeated *R* genes' might form one component of quantitative resistance (Stewart et al. 2003). Contemporary breeding goals emphasize the improvement of polygenic or quantitative resistance, which is considered more durable. The selection of cultivars combining good agronomic qualities with high levels of quantitative resistance to late blight is difficult, particularly under the long day growing conditions of middle and Northern Europe, where quantitative resistance is correlated with late plant maturity (Visker et al. 2004) corresponding to less adaptation to long day length, which is an undesirable agronomic character. The identification and molecular diagnosis of alleles increasing quantitative late blight resistance that is not compromised by late plant maturity can therefore make a significant contribution to the genetic improvement of the cultivated potato.

Owing to its relevance, quantitative resistance to late blight was one of the first complex traits that was subjected to genetic dissection by means of DNA-based markers (Leonards-Schippers et al. 1994). Since then, numerous QTL mapping or tagging studies have been conducted in bi-parental populations, most of them originated from crossing diploid, heterozygous parents (Bradshaw et al. 2006; Collins et al. 1999; Costanzo et al. 2005; Danan et al. 2009; Ewing et al. 2000; Ghislain et al. 2001; Leonards-Schippers et al. 1994; Oberhagemann et al. 1999; Sandbrink et al. 2000; Simko et al. 2006a; Sørensen et al. 2006; Villamon et al. 2005; Visker et al. 2003, 2005) and some from tetraploid, heterozygous cultivars (Bormann et al. 2004; Bradshaw et al. 2004; Mayton et al. 2010; Meyer et al. 1998). QTL mapping revealed the truly polygenic nature of quantitative resistance to late blight. At least one late blight QTL on each of the 12 potato chromosomes was identified. QTL effects are highly variable, depending on the genetic background. In QTL mapping studies where both resistance to late blight and plant maturity were evaluated (Bormann et al. 2004; Collins et al. 1999; Oberhagemann et al. 1999; Visker et al. 2003, 2004, 2005), most but not all QTL for late blight resistance co-localized with QTL for plant maturity. This indicates that breeding for resistance to late blight not compromised by late maturity is feasible. In most cases, the diagnostic value of the numerous DNA markers found to be linked with late blight QTL in experimental populations of highly diverse and often interspecific genetic background has not been evaluated in advanced tetraploid germplasm used for variety development. These markers are therefore primarily diagnostic for the QTL alleles inherited by the specific parents of a particular mapping population. Rather than in immediate breeding applications, the value of these QTL mapping studies lays in the fact that they allow—by comparing and integrating QTL maps constructed in different genetic backgrounds by means of shared markers—the reproducible identification of regions in the potato genome, which contain factors for resistance to late blight (Danan et al. 2011; Gebhardt and

Valkonen 2001; Simko 2002). The integration of positional information of *R* genes and QTL for resistance to late blight and various other pathogens led to the discovery of resistance 'hot spots', genome segments where genes for qualitative and quantitative resistance to different pathogens are clustered (Gebhardt and Valkonen 2001; Leister et al. 1996) (http://www.gabipd.org/database/maps.shtml). This essential information on the 'Know Where' of resistance factors on the potato molecular maps guided the selection of candidate loci for the first association mapping experiments and is also important for the future search for the causal genes for pathogen resistance in the potato genome sequence (PGSC 2011).

Positional information of late blight QTL, *R* genes and candidate genes provided the basis for the first attempts to identify associations between DNA-based markers and quantitative resistance to late blight in populations of tetraploid varieties and advanced breeding clones (Gebhardt et al. 2004; Malosetti et al. 2007; Pajerowska-Mukhtar et al. 2009). A PCR (polymerase chain reaction) marker diagnostic for the *R1* gene conferring race specific resistance to late blight (Ballvora et al. 2002) and markers tightly linked to *R1* were associated with quantitative resistance to late blight and plant maturity in a collection of 400 historical cultivars (Gebhardt et al. 2004). This association was confirmed in a different population of varieties bred in the former Soviet Union (Beketova et al. 2006). The markers for the association test had been chosen based on the fact that the *R1* locus and a major QTL for late blight resistance mapped to the same resistance 'hot spot' on potato chromosome V (Leonards-Schippers et al. 1994). Markers tagging genes containing a nucleotide binding site (NBS), were also associated with quantitative resistance to late blight (Malosetti et al. 2007). The choice of this type of marker was guided by the notion that the majority of plant *R* genes contain a NBS domain (Takken and Joosten 2000) and that NBS containing genes sometimes co-localize with resistance QTL (Leister et al. 1996). These results suggest that allelic variants of *R* genes or *R* gene homologs may be one of the causal factors for quantitative resistance to late blight. The quantitative resistance scores in both studies were derived from variety passport data, which are highly unbalanced and do not take into account the maturity effect on resistance. In a more recent association mapping experiment (Pajerowska-Mukhtar et al. 2009), balanced phenotypic data were obtained for quantitative late blight resistance and plant maturity by evaluating 184 advanced breeding clones over 2 years in replicated field trials. The trait 'maturity corrected resistance' (MCR) was calculated from the phenotypic data. The population was genotyped for single nucleotide polymorphisms (SNPs) at 21 candidate loci selected based on linkage to resistance QTL in experimental populations and/or function in pathogen recognition (*R* genes), defense signalling or defense response. SNPs at two candidate loci, *StAOS2* and *StAOS1* on chromosome XI and V, respectively, were associated with MCR, explaining simultaneously 47 % of the genetic variance in this association panel. *StAOS2* and *StAOS1* both encode allene oxide synthase, an enzyme functional in the biosynthesis of jasmonates (Wasternack and Kombrink 2009), which function in defense and stress signalling. *StAOS2* has been functionally characterized (Pajerowska-Mukhtar et al. 2008). Silencing *StAOS2* in a potato genotype with high level of quantitative resistance to late blight resulted in inhanced susceptibility. Complementation of

the *Arabidopsis thaliana* knock-out mutant *aos* with different potato *StAOS2* alleles resulted in quantitative differences between complementation phenotypes. Alleles linked to increased resistance in potato complemented *aos* mutant phenotypes significantly better than alleles linked to increased susceptibility. Taken together, the evidence suggests that *StAOS2* is one of the causal genes, natural variation of which contributes to quantitative resistance in potato by a so far unknown mechanism.

A further candidate for quantitative resistance to late blight is the complex *StKI* locus on chromosome III, which consists of mixed clusters of protease inhibitor genes that are down-regulated during the compatible interaction with *P. infestans.* Two microsatellites located at the *StKI* locus were associated with MCR (Odeny et al. 2010).

Statistic epistatic interactions were also analysed in the association panel of Pajerowska-Mukhtar et al. (2009), (Stich and Gebhardt 2011). Whereas the epistatic interaction of two loci both on chromosome V explained 16 % of the genetic variance of plant maturity, only minor interactions were found for resistance to late blight.

Probing the potato genome for marker trait associations at such a small number of loci as used in the association mapping experiments summarized above, is comparable with finding a needle in the haystack upon the first attempt. The reason for the surprising fact that testing a very small number of loci identified already several markers associated with complex disease resistance, can be the presence of large haplotype blocks in the genomes of tetraploid potato cultivars. The number of meiotic generations during the past 100 years of breeding is rather limited, the consequence of which could be large linkage disequilibrium (LD) between physically linked loci. Evidence for LD extending over several Centimorgans has been found (D'hoop et al. 2010; Li et al. 2008; Simko et al. 2006b). A counteracting factor that decreases LD is the breeding system in potato, which is based on out-crossing (Gebhardt et al. 2004; Simko et al. 2006b). On the other hand, the knowledge biased selection of candidate loci should have increased the chance of finding marker trait associations.

3.3.2 Root Cyst Nematodes: Globodera Pallida

Two closely related root cyst nematode species, *Globodera rostochiensis* and *Globodera pallida,* are the most damaging parasitic nematodes in potato cultivation (Evans and Trudgill 1992). Chemical control is not a valid option, due to the high environmental toxicity of nematicides. Moreover the long survival rate of the cysts in the soil prevents control by crop rotation. *Globodera* species are therefore quarantined and resistance to root cyst nematodes is obligatory for new potato varieties bred for the European market. Nematode resistance is assessed by the inoculation of plantlets with a defined number of cysts or by planting in nematode infested soil. The newly formed cysts are collected and counted 4–6 weeks post inoculation. This phenotypic test is laborious and time consuming. Replacing it early in the selection cycle by diagnostic markers is therefore advantageous. Whereas resistance to *Globodera rostochiensis* is controlled by major genes of high durability, which have

been introgressed from related potato species and are widely distributed in modern varieties (Ross 1986), resistance to *G. pallida* is more difficult to select as it appears quantitative and its phenotypic assessment is complicated by the fact that it depends not only on the genotype of the potato plant tested but also on the nematode population used for inoculation.

As prerequisite for marker development, QTL for resistance to *G. pallida* have been mapped in diploid and tetraploid bi-parental populations, mostly derived from interspecific crosses involving various wild potato species (Bradshaw et al. 1998; Bryan et al. 2002, 2004; Caromel et al. 2003, 2005; Kreike et al. 1994; Rouppe van der Voort et al. 1998, 2000; Sattarzadeh et al. 2006; Tan et al. 2009). QTL mapping revealed a less complex genetic structure of resistance to *G. pallida* as compared to *P. infestans*. Most of the variation in cyst counts could be explained by one major and one or few minor QTL. In two cases, the combination of resistance alleles from the major and one minor locus was additive, leading to very high levels of resistance (Caromel et al. 2005; Rouppe van der Voort et al. 2000). Three major QTL of different origin have been identified on potato chromosomes IV, V and XI. Particularly the major QTL for resistance to *G. pallida* on chromosome V was repeatedly and independently identified using different wild potato species as resistance donor (Bryan et al. 2002; Caromel et al. 2005, 2003; Kreike et al. 1994; Rouppe van der Voort et al. 1998, 2000; Sattarzadeh et al. 2006). This QTL is part of the resistance 'hot spot' on potato chromosome V, which also includes qualitative and quantitative resistance factors to late blight as well as virus resistance genes (Gebhardt and Valkonen 2001). The high resistance against the most important *G. pallida* pathotypes present in around two dozen varieties bred in The Netherlands is based on a resistance allele at this QTL that has been introgressed from *Solanum vernei*. SNP markers closely linked with this resistance allele were converted in an allele specific PCR assay, which is easy to use and highly diagnostic in tetraploid cultivars, meaning that the PCR product was absent in all susceptible cultivars tested and present in almost all resistant cultivars (Sattarzadeh et al. 2006). The second major QTL for resistance to *G. pallida* on chromosome IV originates from *Solanum tuberosum* ssp. *andigena*. Two markers, one SNP marker and one microsatellite marker both linked to the same resistance allele were shown to be diagnostic for this alternative source of quantitative resistance to *G. pallida.* Genotyping a population of breeding clones with these markers demonstrated that clones having the markers had highly significantly lower cyst counts than genotypes lacking the markers (Moloney et al. 2010). To the best of our knowledge, the markers diagnostic for resistance to *G. pallida* described above are now being used in commercial breeding programs.

3.3.3 Fungi

One of the first examples of association mapping of a complex trait in potato is resistance to *Verticillium* wilt (Simko et al. 2004). *Verticillium* species are soil borne fungi that cause premature wilting of the host plant by blocking their vascular

system and thereby lead to considerable yield losses. Phenotypic testing for resistance is time and labour intensive and not always reliable. In the case of *Verticillium*, the co-linearity between the potato and tomato genome and the DNA sequence of *Ve*, a cloned tomato *R* gene for race specific resistance to *Verticillium* wilt (Kawchuk et al. 2001), were exploited to localize on potato chromosome IX the orthologous *StVe* locus, which confers quantitative resistance to *Verticillium* wilt. A microsatellite marker closely linked to *StVe* was then successfully tested for association with *Verticillium* wilt resistance evaluated in a population comprising most North American breeding lines and commercial cultivars. Furthermore, a CAPS (cleaved amplified polymorphic sequence) marker derived from the tomato *Ve* gene was diagnostic for *Verticillium* wilt resistance in new breeding materials (Bae et al. 2008).

The obligate biotrophic, soil-borne fungus *Synchytrium endobioticum* causes wart disease of potato, which is a serious problem in countries with moderate climates. *S. endobioticum* induces hypertrophic cell divisions in plant host tissues leading to the formation of tumor-like structures (Baayen et al. 2006). Potato wart is a quarantine disease and chemical control is not possible. New *S. endobioticum* pathotypes have appeared, to which most commercial varieties are susceptible. The phenotypic evaluation of wart resistance is laborious, time-consuming and sometimes ambiguous, which makes breeding for resistance difficult. Wart resistance is therefore a typical trait, where diagnostic markers would be highly welcomed by the potato breeding industry. In diploid mapping populations, resistance to *S. endobioticum* segregated as single dominant gene (Brugmans et al. 2006; Hehl et al. 1999), whereas quantitative resistance was observed in two half sib families derived from crossing tetraploid breeding clones (Ballvora et al. 2011). Bulked segregant analysis of siblings with the highest and lowest resistance level identified three microsatellite markers linked to wart resistance loci (*Sen*) on chromosomes I, IX and XI. Further dissection of the *Sen* loci with linked SNP markers revealed a complex allele structure with *Sen* alleles increasing or decreasing resistance to wart (Ballvora et al. 2011). The diagnostic value of these markers in multi-parental breeding populations has not been tested yet. Association genetics for resistance to *S. endobioticum* pathotype 1 is possible, as resistance alleles to pathotype 1 are sufficiently frequent in the tetraploid germplasm pool. Resistance to the more recent pathotypes is however rare and calls for linkage mapping and pedigree based introgression in progenies of specific resistance donors.

3.3.4 Viruses: PLRV

Viral infections in the foliage are transmitted to the next tuber generations, leading year after year to more severe reduction of crop productivity. This is a problem in production areas, where certified seed tubers are not in standard use (Valkonen 2007). Genes for virus resistance have been introgressed from wild potato species and are in most cases inherited as single dominant factors (Ross 1986). For the economically most relevant *Potato Virus Y* (PVY), diagnostic DNA markers for two different genes from *S. tuberosum* ssp *andigena* and *S. stoloniferum*, both conferring extreme

resistance to PVY, are available (Flis et al. 2005; Kasai et al. 2000; Song et al. 2005; Song and Schwarzfischer 2008; Witek et al. 2006). Resistance to *Potato Leaf Roll Virus* (PLRV), one of the most widespread potato viruses, appears quantitative and has been subjected to QTL analysis in two diploid, bi-parental mapping populations (Marczewski et al. 2001, 2004). Two different major QTL (*PLRV.1* and *PLRV.4*) explaining most of the phenotypic variation were identified on potato chromosome XI, one of them (*PLRV.1*) in a 'hot spot' for qualitative as well as quantitative resistance to different pathogens (Marczewski et al. 2001). The SCAR (sequence characterized amplified region) markers linked to these major QTL were not further evaluated for diagnostic power in tetraploid breeding materials.

3.3.5 Bacteria: Erwinia (Pectobacterium)

The two most important bacterial diseases of cultivated potato, blackleg of stems and tuber soft rot, are caused by *Erwinia* species. Blackleg of stems occurs early in the growing season, whereas tuber soft rot occurs during storage (Van der Wolf and De Boer 2007). Chemical control and other agricultural practices are only in part effective in confining the pathogen. Phenotypic screening for resistance is destructive, requires a high number of tubers and is hampered by environmental variation. The genetic resistance currently present in cultivars is insufficient to protect the crop. Association mapping is therefore not promising in this case. Diagnostic markers will have to be developed for new resistance sources, which can be introgressed in advanced breeding clones. New sources of polygenic resistance to *Erwinia carotovora* ssp. *atroseptica* (*Eca*) ssp. atroseptica (Eca) have been selected in diploid hybrids originating from intercrossing *S. tuberosum* with the wild species *S. chacoense* and *S. yungasense*. One progeny derived from these materials was used to map QTL for resistance of tubers and leaves to *Eca* (Zimnoch-Guzowska et al. 2000). Linkage maps were constructed using AFLP, RFLP (restriction fragment length polymorphisms) and resistance-gene-like (RGL) markers. QTL analysis revealed complex inheritance of resistance to *Eca*, without major QTL effects. Similar to QTL for late blight resistance, QTL for resistance to *Eca* were located on most potato chromosomes. Interestingly, some QTL for *Eca* co-localized with QTL for late blight resistance, several were linked to RGL loci and four QTL mapped to resistance 'hot spots'. The genetic dissection of resistance to *Eca* is an important contribution to the 'Solanaceae function map for pathogen resistance' (Gebhardt and Valkonen 2001) (http://www.gabipd.org/database/maps.shtml).

3.3.6 Insect Resistance

The Colorado potato beetle (CPB) (*Leptinotarsa decemlineata*) is widely regarded as the most important insect pest in potato cultivation (Alyokhin et al. 2008). Besides that, aphids are also important as vectors for virus transmission. Resistance to CPB

has been genetically engineered by inserting in the potato genome a *Cry* gene from *Bacillus thuringiensis* (Perlak et al. 1993). Natural insect resistance was discovered in the wild species *S. berthaultii*, which is associated with the presence of glandular trichomes and was introgressed in *S. tuberosum* (Plaisted et al. 1992). A QTL mapping experiment was performed in two inter-specific, diploid mapping populations originated from crossing an insect resistant *S. berthaultii* genotype with a susceptible *S. tuberosum* clone (Bonierbale et al. 1994). The populations were phenotyped for five quantitative traits associated with trichome morphology and biochemistry, as well as for CPB oviposition and consumption, and genotyped with RFLP markers. Between two and five QTL were identified for each of the various traits on eight potato chromosomes. Interestingly, the only QTL for both CPB oviposition and consumption coincided with three QTL for trichome morphology and biochemistry and was located on chromosome V (Bonierbale et al. 1994). The positions of these QTL overlap with the resistance 'hot spot' on potato chromosome V. The introgression of the insect resistance of *S. berthaultii* is hampered by linkage drag of unacceptable characters such as late tuberization and therefore late plant maturity (Bonierbale et al. 1994).

3.4 Tuber Traits

Besides freshly cooked or fried for the table, potato is used for dried products, starch production and processed into French fries, chips and snacks. In developed countries, up to 60 % of potato is consumed in processed form, and the increase in consumption of processed potato products is an indicator for industrialization in developing countries (Kirkman 2007). Potato quality requirements for the processing industry varied little over time and are the basis for the selection of cultivars with as many desirable characteristics as possible (Gould 1999; Talburt and Smith 1987).

In addition to yield, the essential trait for any crop, tuber morphology (e.g. shape, size, eye depth), dry matter composition, starch content and composition, and the amount of reducing sugars with and without cold storage (cold-induced sweetening) are important for tuber quality and considered as key traits by the breeding as well as processing industries. Other quality characters like tuber texture, flavour, taste, flesh colour, nutritional value, discoloration potential and susceptibility to bruising emanate from the demands of consumers and the retail industry. Precision breeding for tuber quality traits is at a very early stage, due to the phenotypic and genetic complexity. To facilitate in the future the selection of cultivars with improved tuber quality traits by molecular diagnostic tools and thereby increase the precision of breeding programs, knowledge of gene number, chromosomal location and identity, and the magnitude of allelic effects on the traits of interest is required. Progress has been made in tackling these issues not only by linkage mapping but also by association genetics and the integration of next generation sequencing techniques as well as *omics* approaches. In the following section we compile the current knowledge about the genetics of potato tuber quality traits and the genes possibly contributing to their natural variation.

3.4.1 Tuber Starch Content and Starch Yield

The net amount of starch accumulated in plastids is the result of anabolic and catabolic enzymatic reactions and carbon flux from source leaves to sink tubers. The major storage compound starch is composed of the glucose polymers amylose and amylopectin, which account for 10–25 % of the tuber fresh weight. Tuber starch content is strongly correlated but not synonymous with dry matter content. In addition to starch, dry matter content includes minor amounts of other components, such as lipids and proteins (Morrison and Karkalas 1990). Tuber dry matter and starch content are easily and in non destructive manner quantified by measuring specific gravity (Scheele et al. 1937). Tuber starch content is slightly correlated with tuber yield (Urbany et al. 2011b). Starch yield, the product of tuber starch content and yield (see below), rather than tuber starch content alone, is the selection parameter most relevant in breeding programs.

Tuber starch content has been extensively studied at the molecular and functional level, and is therefore considered as model trait for exploring the candidate gene approach for identifying the molecular basis of QTL (Pflieger et al. 2001). Transgenic approaches highlighted the impact on tuber starch content of single genes or combinations of genes functional in starch metabolism (Jobling 2004). Examples are the production of amylose free (waxy) starches in sweet potato and potato tubers by antisense down-regulation of starch synthase genes (*GBSS, SSII* and *SSIII*) (Kimura et al. 2001; Visser et al. 1991; Zheng and Sosulski 1998), high amylose starches by inhibiting starch branching enzymes (*SBE*) (Bird et al. 2000; Schwall et al. 2000) and the generation of starch with low phosphate content via antisense inhibition of α-glucan water dikinase (GWD) (Lorberth et al. 1998).

QTL for tuber starch content or dry matter have been mapped in diploid experimental populations (Freyre and Douches 1994; Gebhardt et al. 2005; Schäfer-Pregl et al. 1998) and in one tetraploid population (Bradshaw et al. 2008). QTL mapping demonstrated that tuber starch content is controlled by factors on all potato chromosomes (Gebhardt et al. 2005; Schäfer-Pregl et al. 1998). Furthermore, cloned genes functioning in the biosynthesis, degradation or transport of starch and sugars have been positioned on potato molecular maps, and candidate genes were identified on the basis of co-localization with QTL (Chen et al. 2001; Menendez et al. 2002). These studies facilitated the first association studies in populations of tetraploid cultivars based on candidate gene markers (Li et al. 2005a, 2008). A whole genome association mapping approach based on genotyping with anonymous AFLP markers was performed by (D'hoop et al. 2008).

Among many others, invertases and L-type starch phosphorylases are functional candidate genes, because they influence the balance between starch and sugars. L-type starch phosphorylases are starch degrading enzymes located in the plastids, which catalyse the reversible phosphorolytic cleavage of linear glucan chains. The gene *PHO1A (Stp23)* on potato chromosome III is expressed in amyloplasts, whereas *PHO1B (StpL)* on chromosome V is expressed mainly in chloroplasts (Albrecht et al. 2001). Associations of single strand conformation polymorphism (SSCP) markers derived from *PHO1A* and *PHO1B* with tuber starch content and starch yield support

a contribution of starch phosphorylase to the natural variation of these traits. SSCP markers originating from *PHO1A (Stp23)* were associated both with increased tuber starch content and starch yield, whereas a marker derived from *PHO1B (StpL)* showed a decreasing effect on both traits (Li et al. 2008; Urbany et al. 2011b). In contrast, SSCP marker *StpL-3e* (*PHO1B*) was associated with increased tuber starch content but decreased tuber yield and had therefore no effect on starch yield (Li et al. 2008). Effects of *PHO1A* and *PHO1B* allelic variants on tuber starch content were verified in two independent association studies using different populations in multi- year and -environment trials (Li et al. 2008; Urbany et al. 2011b).

Invertases cleave irreversibly sucrose into the reducing sugars fructose and glucose and thereby play an important role in the partitioning of carbon between source (photosynthetic leaves) and heterotrophic sink tissues (potato tubers). Four independent loci encode invertase genes in potato (Hedley et al. 1993, 1994; Liu et al. 2011; Zhou et al. 1994; Zrenner et al. 1996). The *Pain-1* locus on potato chromosome III encodes a vacuolar invertase, whereas loci Inv_{ap}-*a* and Inv_{ap}-*b* on chromosomes X and IX encode apoplastic invertases (Chen et al. 2001). The locus Inv_{ap}-a *consists of* two tandem duplicated genes *InvCD111* and *InvCD141,*which are orthologous to the similarly organized gene pair *LIN6* and *LIN8 on* tomato chromosome 10. The Inv_{ap}-*b* locus also consists of two tandem duplicated genes, *InvGE* and *InvGF*, which are orthologous to the tomato invertase genes *LIN5* and *LIN7* on chromosome 9 (Draffehn et al. 2010; Fridman and Zamir 2003). Positive as well as negative associations with tuber starch content were detected with SSCP markers derived from the gene sequences at all three invertase loci (Li et al.2005a, 2008). The most significant association was found for the SSCP marker Pain1–9a, which increased tuber starch content, explaining 12 % of the total variation (Li et al. 2008). cDNA alleles corresponding to the associated SSCP markers were identified among 55 full length cDNA clones derived from five invertase genes. The cDNA alleles were cloned from three tetraploid and three diploid genotypes. Allelic invertase sequences differed between 4 and 9 % from each other, demonstrating tremendous natural allelic variation in this gene family (Draffehn et al. 2010). SNPs diagnostic for the associated invertase alleles were identified (Draffehn et al. 2010; Li et al. 2005a), which are the basis for designing allele specific PCR assays for breeding applications (Draffehn et al. 2010; Li et al. 2005a). Recently, a sixth potato invertase gene has been identified (Liu et al. 2011), which resides on potato chromosome VIII according to the newly released potato genome sequence (PGSC 2011) accessible via the genome browser at http://potatogenomics.plantbiology.msu.edu/index.html. No data about the possible role of this candidate gene and its allelic variants for tuber starch content and other quality traits are currently available.

In addition to invertases and starch phosphorylases, other genes are associated with tuber starch content. An SSCP marker derived from the *G6pdh* locus on chromosome II showed a negative association with tuber starch content (Li et al. 2008). Glucose −6-phosphate dehydrogenase (G6pdh) catalyzes the oxidation of glucose −6-phosphate (G6P) to 6-phosphogluconolactone concomitant with reducing NADP to NADPH. G6pdh controls the flux through the irreversible branch of the oxidative pentose phosphate pathway. In higher plants, isoforms of the enzyme reside in the cytosol and in plastids (Wendt et al. 2000). Furthermore, an SSCP marker derived

from soluble starch synthase I (*SssI*) on chromosome III was positively associated with tuber starch content. *SssI* is predominantly expressed in source and sink leaves, and only to lower extend in potato tubers (Kossmann et al. 1999). Starch synthases catalyse the glucosyl transfer from ADP-glucose to the non-reducing end of an α-1,4-glucan (Preiss et al. 1991) and thereby contribute to the synthesis of starch grains. Although this starch synthase isoform represents only a minor activity in potato tubers, it might influence tuber starch content by controlling sink to source carbon flux and regulating the carbon-fraction available for storage deposit in tubers.

An intriguing recent finding demonstrated that, in addition to association of single candidate gene markers, also pair-wise epistatic interactions of such markers contribute to the variation of tuber starch content and starch yield (Li et al. 2010). When 190 DNA markers at 36 loci scored in an association mapping population were tested for pair-wise statistical epistatic interactions, 50 marker pairs were associated mainly with tuber starch content and/or starch yield. Alleles at loci encoding ribulose-bisphosphate carboxylase/oxygenase activase (*Rca*), sucrose phosphate synthase (*Sps*) and vacuolar invertase (*Pain −1*) were most frequently involved in epistatic interactions. Sps is a key enzyme in the biosynthesis of sucrose, the molecule transporting carbon from source leaves to the developing tubers, where it is converted into starch. Rubisco activase (Rca), a chloroplast-localized enzyme, belongs to the AAA + family of ATPases associated with diverse cellular activities (Portis et al. 2008), and activates ribulose −1,5-bisphosphate carboxylase/oxygenase (Rubisco). Whereas a SCAR (sequence characterized amplified region) marker derived from *Rca* did not show significant association with tuber starch content on its own, a highly significant, positive effect on tuber starch content and starch yield was observed when combined with the SSCP marker *Pain1–8c*. This epistatic interaction explained 9 and 10 % of the total variance, respectively (Li et al. 2010). Sps, Pain −1 and Rca proteins are functionally linked with each other and with many other enzymes through the metabolic flux from photosynthetic (source) to storage (sink) organs. Epistatic interactions among them might be explained by the fact, that in a metabolic network the effect of allelic variation at one locus depends on the allelic variation at a second or more loci (Alcázar et al. 2009).

3.4.2 Reducing Sugars, Cold Induced Sweetening and Chip Quality

Potato tubers remobilize a small fraction of starch during cold storage (Isherwood 1973). The degradation of amylose and amylopectins and the hydrolytic cleavage of sucrose by invertase lead to the accumulation of the reducing sugars glucose and fructose (cold-induced sweetening). Cold storage of tubers is used by the potato processing industry to delay sprouting, as an alternative to chemical sprout suppressants with a potential risk to human health. Cold-induced sweetening presents however a problem to the processing industry, especially when producing potato chips and French fries. This is mainly due to the non-enzymatic Maillard reaction that takes

place between reducing sugars and free amino acids at the high temperatures during the frying process. The Maillard reaction produces adducts that cause a brown discoloration of the processed products. Excessive discolouration and the development of off-flavours severely reduce product quality (Townsend and Hope 1960). The accumulation of reducing sugars and thereby the chip quality depends on genotype and environment, and keeping a very low content of reducing sugars during cold storage is a key trait for the selection of new processing varieties. Phenotypic selection of genotypes with low reducing sugar content is done by a frying test, which is destructive, requires multiple tubers and, thus, can be performed only after several years of tuber multiplication.

QTL mapping for tuber sugar content or chip/fry colour before and after cold storage identified a number of regions in the potato genome, where genes influencing these traits are located (Bradshaw et al. 2008; Douches and Freyre 1994; Menendez et al. 2002). Several QTL overlapped with loci for genes functional in carbohydrate metabolism and transport (Chen et al. 2001; Gebhardt et al. 2005), first of all invertases (see above), the most obvious functional candidate genes, but also ADP-glucose pyrophosphorylase subunit S (*AGPaseS* on chromosome I), sucrose phosphate synthase (*Sps* on chromosome VII), sucrose synthases (*Sus3* and *Sus4* on chromosomes VII and XII), soluble starch synthase (*SssI* on chromosome III), starch branching enzyme and amylase (*SbeI* and *AmyZ* on chromosome IV), starch phosphorylases (*Stp23* and *StpL* on chromosomes III and V), as well as sucrose translocators and sensors (*Sut1* and *Sut2* on chromosomes XI and V).

Based on function and co-localization with QTL for tuber sugar content, candidate genes were selected and tested for association with chip quality before and after cold storage in the same populations as evaluated for tuber starch content and starch yield (see above) (Draffehn et al. 2010; Li et al. 2005a, 2008). The markers most strongly associated with tuber starch content were also associated with chip quality before and after cold storage: SSCP markers derived from the invertase and starch phosphorylase loci *Pain −1*, Inv_{ap}-a, PHO1a and *PHO1b*, from *G6pdh, SssI* and *Pha2* (plasma membrane H^+-ATPase 2), as well as the non-coding SCAR marker *GP171*. Interestingly, the directions of the effects were also the same. Markers associated with increased tuber starch content were associated with increased chip quality (less reducing sugars) and vice versa. An exception was the *Rca* SCAR marker that was negatively associated with chip quality but not with tuber starch content. This suggests that both tuber starch content and chip quality/sugar content are in part controlled by the same genes. Three and six associations with the traits 'frying color' and 'chipping color', respectively, were detected with AFLP markers in an independent population of tetraploid cultivars studied by d'Hoop et al. (2008). The AFLP markers overlapped with some sugar QTL mapped in (Menendez et al. 2002). A comparison of this association study with the one of Li et al. (2008) is however difficult due to the lack of common anchor markers.

The most promising candidates for a causal relationship between allelic variants and natural variation of tuber starch and sugar content are at present invertases and starch phosphorylases. The positive association with chip quality of the SCAR maker *InvGE −6f* derived from the Inv_{ap}-*b* locus, which was discovered by Li et al. (2005a),

could be validated in the independent association mapping population described in Li et al. (2008), although there it explained only a small portion of the total variance of chip quality. Furthermore, the vacuolar invertase allele Pain1–9a explained 10 % of the variance of chip quality after cold storage. Comparative analysis of structure and function of invertase cDNA alleles identified biochemical differences between cDNA alleles (Draffehn et al. 2012). The impact of vacuolar invertase on chip quality was recently demonstrated by antisense down-regulation of the gene in potato plants (Bhaskar et al. 2010). Tubers of transgenic plants showed significant reduction in the accumulation of reducing sugars upon cold storage and better frying quality compared to wild type plants. Another interesting finding represents a possible physical interaction of a putative invertase inhibitor StInvInh2 and vacuolar invertase (Liu et al. 2010). Such interactions could result in a decrease of tuber invertase activity and therefore result in better chip quality due to reduced sugar accumulation after cold storage. The highly significant associations with chip quality and tuber starch content at two plastidic starch phosphorylase loci might be explained by 'positive' *PHO1* alleles that influence the starch-sugar balance in favour of starch by reduced expression, translation or enzyme activity, whereas 'negative' *PHO1* alleles accelerate starch breakdown and sugar accumulation by higher expression, translation or enzyme activity. Comparative analysis of structure and function of associated versus non-associated *PHO1* cDNA alleles revealed a possible mechanism for the observed phenotypic effects (unpublished data from the author's laboratory).

The candidate gene alleles found so far that associated with tuber quality traits in breeding materials, explained collectively 40–50 % of the total variance (Li et al. 2008). There are certainly more genes involved, which have not yet been identified. For example, important enzymes involved in cold-sweetening are β-amylases that catalyse the hydrolysis of α-1,4-D-glucosidic linkages in polysaccharides and thereby remove successively maltose units from the non-reducing ends of the chains (Scheidig et al. 2002). Phospho-gluco-mutase (PGM) interconverts α-D-glucose 1-phosphate to α-D-glucose 6-phosphate, providing the substrate for AGPase, a key enzyme in starch biosynthesis (Fernie et al. 2002). Transgenic approaches have shown that antisense down-regulation of these genes in potato severely affects carbohydrate metabolism and levels (Fernie et al. 2002; Scheidig et al. 2002). A further interesting candidate is UDP-glucose pyrophosphorylase (*UGPase* on chromosome XI). A correlation of UGPase alleles with frying quality as well as sugar accumulation was demonstrated in a panel of 12 cultivars (Sowokinos et al. 2004).

The candidate gene approach is necessarily biased towards known genes. Recent publications provide valuable sources for novel candidate genes by using omcis approaches in order to investigate cold-induced sweetening. In the study Bagnaresi et al. (2008), the authors took advantage of a tomato Affymetrix GeneChip to detect differentially expressed potato tuber transcripts in response to cold treatment. Strong up-regulation was observed, among others, for transcripts annotated as ethylene-associated enzymes, which have not been linked to cold induced sweetening so far. In two proteomics studies (Hoehenwarter et al. 2011; Yang et al. 2011) gel-free approaches were used to identify and quantify proteins that are differentially regulated after cold storage of tubers of cultivar Désirée (Yang et al. 2011) or show

quantitative differences between twelve cultivars (Hoehenwarter et al. 2011). These studies provide molecular information for further understanding the phenomenon of cold induced sweetening at the functional level. Association mapping will have to be performed in order to test whether indeed candidate gene variation influences phenotypic diversity of processing quality in advanced breeding material.

3.4.3 Enzymatic Discoloration, Tuber Bruising and After Cooking Darkening

Enzymatic discoloration, tuber bruising and after cooking darkening are complex traits that have in common the formation of dark pigments from phenolic precursors, a negative tuber attribute. Breeding for resistance to discoloration reactions by phenotypic selection is difficult, because testing is destructive and requires high tuber numbers that are not available early in the breeding cycle. Molecular diagnostic tools for these traits would have therefore offer a real advantage.

Enzymatic discoloration is a phenomenon not only observed in tubers but also in fruits and vegetables. Phenolic compounds are oxidized to quinones by oxidoreductases, notably polyphenol oxidases (PPOs) or tyrosinases (Friedman 1997; Ravi and Aked 1996). PPO has been identified as one of the key factors in enzymatic discoloration. Several studies, including transgenic approaches supported an impact of PPO upon the discoloration reaction (Bachem et al. 1994; Rommens et al. 2006; Thipyapong et al. 2007). Potato PPO is encoded by a gene family of at least six members, located in a gene cluster on potato chromosome VIII (Hunt et al. 1993; Thygesen et al. 1995). The genes are differentially expressed, with POT32 being the major form expressed in tubers (Thygesen et al. 1995). Recent studies identified three QTL for enzymatic discoloration in a bi-parental diploid mapping population on chromosomes I, III and VIII (Werij et al. 2007). The QTL peak on chromosome VIII coincided with the map position of *POT32,*and three different *PPO* alleles (*POT32–1,−2*, and *−3*) were discriminated at this locus. A clear correlation between allele composition and the degree of discoloration was observed. Furthermore, analysis of POT32 expression indicated a correlation between transcript level and allele composition. Genotypes having two copies of the *POT32–1* allele had both the highest average degree of discoloration and the highest average level of POT32 expression. These findings substantiate PPO's as primary functional and positional candidate genes for the natural variation of enzymatic discoloration in potato tubers. PPO substrate concentrations, notably chlorogenic acid and tyrosine have also been reported to influence the degree of discoloration. Measuring the levels of these metabolites in tubers of the mapping population and treating them as quantitative trait, Werij et al. (2007) identified a QTL for chlorogenic acid levels on chromosome II and a QTL for tyrosine levels on chromosome VIII. In contrast to earlier studies, the authors proposed the absence of genetic correlations between the metabolite QTL and enzymatic discoloration, as none of the QTL overlapped.

Mechanical impact during harvest, transport and storage of potato tubers initiates the development of an internal tissue discoloration called 'blackspot bruising'. Different cellular compartments are disrupted in damaged cells and in consequence, the enzymatic discoloration is triggered. Bruising symptoms develop under the tuber skin 3–24 h after mechanical impact. As for tissue discoloration, *PPO*s are also considered as the major candidate genes for natural variation of tuber bruising. Contrasting findings do not allow a clear statement on the impact on discoloration and tuber bruising of chlorogenic acid content, the main phenolic compound in potato tubers, of its precursor tyrosine and the partitioning of tyrosine between tuber protein and the free amino acid pool (Corsini et al. 1992; Mondy and Munshi 1993). In addition to PPO and its substrates, reducing agents such as ascorbate and carotenoids as well as redox homeostasis influence tuber bruising (el-Shimi 1993; Henze 1956). Besides enzymatic components, structural factors like cellular membrane stability, cell number and architecture also affect tuber bruising (Stevens and Davelaar 1997). In contrast to enzymatic discoloration, black-spot bruising is strongly correlated with specific gravity and thereby tuber starch content (Urbany et al. 2011b). The amount and steric properties of starch granules affect cellular stability. In consequence, tubers with higher tuber starch content are more sensitive to mechanical damage and show higher bruising susceptibility. The effects of tuber starch content and bruising can be separated by calculating the trait 'starch corrected bruising' (SCB) from the residuals of the regression of the bruising index on specific gravity (Urbany et al. 2011b).

A population of 205 tetraploid varieties and breeding clones was phenotyped for tuber bruising susceptibility, specific gravity, yield, shape and plant maturity and genotyped with SSCP markers derived from 33 functional candidate genes for tuber bruising (Urbany et al. 2011b). Among others, significant associations between PPO isoforms and tuber bruising susceptibility were identified. The SSCP marker *POT32PS1-f* derived from the main tuber PPO isoform POT32 was present in 10 % of the genotypes and was associated with decreased values for SCB (higher resistance to bruising). In contrast, the *POLOXA* marker present in 35 % of the population correlated with an increased bruising susceptibility as well as increased SCB values. The association of the *POLOXA* marker originating from PPO isoform *POTP1* with bruising susceptibility is intriguing, because expression of the *POTP1* gene is reported to be high in leaves but restricted in tubers (Thygesen et al. 1995). A possible explanation is that the PCR primers for *POLOXA* amplified more than one *PPO* gene (Urbany et al. 2011b). Nevertheless, selection against the *POLOXA* marker and enriching for the *POT32PS1-f marker* should improve the average bruising resistance in a breeding population. Furthermore, the markers *POLOXA* and *POT32PS1-f* are particularly interesting for marker-assisted-selection aiming at increased bruising resistance that is not compromised by low tuber starch content.

SSCP markers originating from the *PHO1A (Stp23)* locus on potato chromosome III and *PHO1B (StpL)* on chromosome V were associated with bruising susceptibility and with tuber starch content. *PHO1A* markers mainly increased bruising susceptibility and to less extent tuber starch content. Vice versa, *PHO1B* markers were

associated primarily with tuber starch content and secondary with bruising susceptibility (Urbany et al. 2011b). The effects of *PHO1A* and *PHO1B* allelic variants on tuber starch content were consistent with the associations previously found at these two loci in an independent association mapping experiment (Li et al. 2008). The observed concerted allele effects on bruising susceptibility and tuber starch content might be due to the fact that alleles increasing tuber starch content lead to higher amyloplastic starch load and thereby to higher susceptibility to mechanical stress and vice versa.

A novel candidate gene annotated as putative class III lipase was *de novo* discovered by comparing the tuber proteome of 12 bruising resistant and twelve susceptible cultivars (Urbany et al. 2011a). SSCP markers derived from this gene, located on chromosome II, showed significant associations with either bruising, SCB and tuber shape (*LIPIII −27–1h*) or bruising and tuber starch content (*LIPIII −27–1e*) or starch content and tuber yield (*LIPIII −27-a*). Recent work on the closest homolog of this potato class III lipase, the *CaPLA1* gene encoding a phospholipase A1 of hot pepper (Seo et al. 2008), points to a possible connection of this gene with tuber bruising. Seo et al. (2008) proposed that *CaPLA1* is involved in the regulation of cell shape and number as well as the control of carbon flux through gluconeogenesis and β-oxidation. In addition, it was postulated that the enzyme is involved in lipid signalling and thereby regulates cellular and biochemical functions in heterotrophic plant tissue. The potato *CaPLA1* homologue might fulfil analogous functions in heterotrophic tuber tissue, thereby indirectly influencing tuber bruising susceptibility and specific gravity. A putative role in the regulation of cell shape and number could also explain why the marker *LIPIII −27–1h* associates with tuber shape.

Potato tubers contain high amounts of phenolic compounds. Besides various roles in biological processes such as defense against pests and pathogens, these compounds also provoke 'after-cooking darkening' (ACD). ACD is an undesirable tuber trait, which results from the development of a complex of chlorogenic acid and ferric iron, when tubers are heated by boiling, frying, or dehydrating (Murphy et al. 2010). The subsequent oxidation of this complex produces a bluish-grey discoloration (Wang-Pruski and Nowak 2004; Wang-Pruski et al. 2007). The severity of this reaction varies among potato cultivars, and has been related to the levels of several metabolites, as well as tuber pH (Silva et al. 1991; Thomas et al. 1979). Investigations on ACD revealed a strong genetic component across different environmental conditions (Dalianis et al. 1966). Molecular mapping based on AFLP and few microsatellite markers in a bi-parental tetraploid progeny identified six QTL for 'after cooking blackening' (Bradshaw et al. 2008). Association mapping also using AFLP genotyping of a variety panel evaluated for ACD and 'after baking darkening' revealed seven and ten QTL, respectively (D'hoop et al. 2008). Five QTL mapped to the same chromosome in both studies. A more detailed comparison, including the candidate loci associated with tuber bruising (Urbany et al. 2011b), is not possible due to the lack of common anchor markers.

Candidate genes like quinate-hydroxycinnamoyl transferase (HQT) catalysing the final step of chlorogenic acid synthesis were proposed (Friedman 1997). The association of an HQT derived SSCP marker with SCB (Urbany et al. 2011b) suggests an

influence on the synthesis of polyphenols, which possibly modulates the strength of the discoloration reaction accompanying ACD. Further evidence concerning genes that have a role in ACD comes from a proteomics study, which compared quantitatively tuber protein profiles of high and low ACD potato clones (Murphy et al. 2010). Thirty proteins showed a correlation with tissue discoloration, among others PPO. Five differential proteins were further investigated by qRT-PCR. Hence, the transcript levels of an aspartic protease inhibitor and a linolelate:oxygen oxidoreductase were in agreement with the proteomic results and were more abundant in high-ACD genotypes. For 5-lipoxygenase, PPO and patatin T5 precursor, qRT results contradicted the proteomic results. An explanation for the discrepancy between transcript and protein abundance could be post translational modification as well as the multitude of isoforms and alleles present in tetraploid potato. The latter hypothesis is further substantiated, as two peptides assigned to PPO's were identified, which showed opposite trends concerning protein abundance between low and high ACD genotypes. Proteomic studies like the one of Murphy et al. (2010) and Urbany et al. (2011a) are a valuable source for novel candidate genes that contribute to trait variation. When combined with association genetics, the molecular basis of complex traits such as tissue discoloration becomes accessible and the knowledge gained can eventually be translated in breeding applications.

3.4.4 Potato Tuber Yield, Size and Tuber Initiation (Tuberization)

Potato tuber yield is a complex trait of primary importance, but the one with the least knowledge available about number and identity of the genes involved. Tuber yield is easily measured by tuber weight per area unit. Tuber yield shows fairly low heritability (H^2), due to a large environmental component influencing yield variation in the field (Urbany et al. 2011b). Not surprisingly therefore, only few QTL for tuber yield have been mapped so far and most of those appear specific for the biparental populations that were analysed (Bonierbale et al. 1993; Bradshaw et al. 2008; Schäfer-Pregl et al. 1998). Exceptions might be a yield QTL on the long arm of chromosome I that was detected in the QTL mapping studies of Schäfer-Pregl et al. (1998) and Bradshaw et al. (2008), and yield QTL on chromosomes II and V detected by Bonierbale et al. (1993) as well as by Schäfer-Pregl et al. (1998). Co-localization of yield QTL with some QTL for tuber starch and sugar content has been observed (Gebhardt et al. 2005; Schäfer-Pregl et al. 1998), one of which was validated in the association mapping experiment of Li et al. (2008). The SSCP marker *StpL −3e* derived from the starch phosphorylase gene *Pho1B* on chromosome V was associated with decreased tuber yield, increased starch content and better chip quality (Li et al. 2008).

Hardly any functional candidate genes for tuber yield are known. Anti-sense repression of plastidial adenylate kinase, which catalyzes the interconversion of ATP and AMP into ADP, in the cultivar Désireé resulted in transgenic plants with increased yield and tuber starch content compared to wild type (Regierer et al. 2002).

The ectopic expression of the Arabidopsis *PHYB* (phytochrome B) gene, resulted in increased tuber yield in field-grown transgenic potatoes (Boccalandro et al. 2003). Phytochrome B is a photoreceptor involved in detecting red to far-red light ratio associated with plant density. These genes have not been tested yet for association with yield in advanced breeding populations. The association mapping population of Urbany et al. (2011b) was evaluated for several traits (see above) including tuber yield, and two highly significant associations with yield were found. First, the SSCP marker *GLDH-h* derived from L-galactono −1,4-lactone dehydrogenase (*GLDH* on chromosome X), involved in ascorbic acid metabolism, was associated with tuber yield. This marker had a high allele frequency of 81 % in the population and its presence correlated with a yield decrease. Second, *HQT-f*, another SSCP marker originating from hydroxycinnamoyl quinate CoA transferase (HQT on chromosome VII) and functional in polyphenol synthesis, was associated with an increased tuber yield (Urbany et al. 2011b). The two genes were selected based on their possible role in tuber bruising and the association with tuber yield is only by coincidence. Whether or not these gene variants are causal for the observed effects on tuber yield awaits further studies, for example by functional analysis of the candidate gene alleles.

Tuber yield is correlated with plant maturity (Urbany et al. 2011b). Cultivars that mature later in the growing season yield on average more, probably due to longer periods of photosynthetic activity. QTL for plant maturity or 'earliness' were mapped on most potato chromosomes in diploid and tetraploid bi-parental populations (Bormann et al. 2004; Bradshaw et al. 2008; Collins et al. 1999; Oberhagemann et al. 1999; Visker et al. 2003, 2004) and in tetraploid association panels (D'hoop et al. 2008; Pajerowska-Mukhtar et al. 2009; Urbany et al. 2011b), mostly in the context of resistance to late blight (see above). However, one major QTL for plant maturity on chromosome V was consistently detected in all studies except by Urbany et al.(2011b). This QTL for plant maturity overlaps with the QTL for yield, tuber starch content and resistance to late blight as described above and in addition with a QTL for plant vigour, which was mapped in a diploid pi-parental population (Collins et al. 1999; Oberhagemann et al. 1999).

Tuber size and number per plant determine tuber yield. Tuber development certainly plays also an important role in determining tuber yield. It comprises stolon initiation, elongation, tuber initiation by stolon swelling and tuber growth (Celis-Gamboa et al. 2003; Ewing and Struik 1992). Tuberization (time until tubers are initiated at the stolon tips) depends on day length and is related, but not synonymous with plant maturity. Recently, two paralogous FT (flowering locus T)-like genes, *StSP6A* and *StSP3D,* were shown to control tuberization and flowering, respectively. The expression of *StSP6A* correlated with the contrasting maturity type of three varieties (Navarro et al. 2011). Whether polymorphisms at the *StSP6A* locus are associated with natural variation of tuberization, plant maturity or yield remains to be tested. QTL mapping of tuberization in an inter-specific, diploid, bi-parental population detected factors on eight potato chromosomes (Šimko et al. 1999; van den Berg et al. 1996b). QTL for tuberization and plant maturity appear to overlap on chromosomes V and VIII, which is consistent with the fact that both traits are physiologically related (Bormann et al. 2004; Collins et al. 1999; van den Berg et al.

1996b). Bradshaw and colleagues reported the correlation between tuber size and yield and identified two QTL for tuber size on chromosome V and an unidentified linkage group in a tetraploid, bi-parental mapping population (Bradshaw et al. 2008).

3.4.5 Tuber Shape and Eye Depth

Tuber shape varies from round to oblong, and eye depth varies from deep to shallow. Large natural variation of tuber shape exists among cultivated varieties and wild species (Bradshaw and Mackay 1994; De Jong and Burns 1993; van Eck 2007). When treated as either quantitative or qualitative (round versus oblong) trait, tuber shape mapped to a single locus *Ro* on potato chromosome X (Jacobs et al. 1995; van Eck et al. 1994). A series of multiple alleles at the *Ro* locus explained most of the variation of tuber shape between round and oblong (van Eck et al. 1994). The *Ro* locus is an example, how a single locus with multiple alleles may be disguised as quantitative trait, particularly at the tetraploid level, where more allele combinations are possible per locus than at the diploid level. The identity of the gene(s) encoded at the *Ro* locus remains to be elucidated. For eye depth a major locus was identified also on chromosome X at four Centimorgan distance from the *Ro* locus (Li et al. 2005b; Maris 1966). However, there are other loci besides *Ro* and *Eye* that influence tuber shape and eye depth. Additional QTL for tuber shape were mapped to chromosomes II, V and XI (Bradshaw et al. 2008; Śliwka et al. 2008) and for eye depth on chromosomes III and V (Śliwka et al. 2008). Furthermore, a putative class III lipase on chromosome II showed association with tuber shape (Urbany et al. 2011b).

3.4.6 Tuber Dormancy, Sprouting

After maturation and harvest, potato tubers undergo a dormancy period of variable length, during which sprouting is suppressed. The end of the dormancy is indicated by the outgrowth of new stems (sprouts) from the tuber eyes. Plant hormones, among others abscisic acid, play a role in regulating tuber dormancy (Suttle 2007). Dormancy is an important quality trait, which determines the storability of the raw material for the processing industry. Genetic dissection of dormancy or sprouting identified one to two QTL on six potato chromosomes (Bradshaw et al. 2008; Freyre et al. 1994; Simko et al. 1997; Śliwka et al. 2008; van den Berg et al. 1996a). Epistatic interactions also play a significant role (Simko et al. 1997; van den Berg et al. 1996a). Interestingly, QTL for tuber abscisic acid content and dormancy co-localized on chromosomes IV and VII, and an epistatic interaction between two markers on chromosomes II and IV was found for both tuber abscisic acid content and dormancy, suggesting that the abscisic acid content of the tuber is causally involved in the variation of tuber dormancy (Simko et al. 1997).

3.4.7 Tuber Texture and Flavour

Texture and flavour of cooked potatoes are complex quality traits that play an important role for consumer preferences and are therefore subject to selection in breeding programs. Texture refers to the differences between mealy and non-mealy/waxy tubers. The molecular basis of the variation in texture and flavour characteristics is not fully understood. Besides starch content and distribution within the tuber, cell size, cell-wall structure and composition and the breakdown of the cell wall middle lamella during cooking are factors influencing cooked potato texture. The relative contribution and importance of the different factors involved is unclear. Potato flavour likely depends on both volatile and matrix associated compounds (Taylor et al. 2007). Candidate genes for tuber texture after cooking were identified by RNA-profiling of contrasting phenotypic bulks of potato genotypes selected from a diploid bi-parental mapping population (Kloosterman et al. 2010). Seventy eight genes were identified, which were differentially expressed between groups of mealy and waxy potatoes. Among those was a candidate with homology to a tyrosine-lysine rich protein (TLRP), which mapped on chromosome IX in close proximity to a QTL for potato cooking type (Kloosterman et al. 2010). With respect to cell wall characteristics, pectin methyl esterases (PME's) are good candidate genes, with PME activity being a potential factor impacting on texture (Ross et al. 2011). PME's are cell-wall-associated enzymes that catalyse the de-esterification of pectin and thereby facilitate plant cell wall modification and subsequent breakdown. The authors showed that a higher level of total PME activity in tubers correlated with a reduced degree of cell wall pectin methylation and a less mealy potato texture. (Ducreux et al. 2008) used expression profiling in tubers of potato genotypes that differed for tuber flavour and texture in order to identify genes involved in these traits. Comparative gene expression analysis between two Phureja and two Tuberosum cultivars resulted in 309 genes that were up-regulated in Phureja compared with Tuberosum, whereas 555 genes were down-regulated. Major differences in expression levels were observed for genes involved in cell wall biosynthesis, including genes encoding pectin acetylesterase, xyloglucan endotransglycosylase and pectin methylesterase, which emphasizes a putative role of these genes in influencing tuber texture. Another interesting finding of this study was the detection of differentially expressed transcripts related to tuber flavour. The Phureja cultivars had higher levels of the volatile compound sesquiterpene α-copaene compared with Tuberosum types. Furthermore the authors identified a sesquiterpene synthase gene that showed higher transcript levels in Phureja tubers. Its corresponding full-length cDNA encoded α-copaene synthase. Other differentially expressed 'flavour genes' included branched-chain amino acid aminotransferases, glutamate and methionine biosynthesis genes, and a ribonuclease suggesting a mechanism for 5′-ribonucleotide formation in potato tubers during cooking.

3.4.8 Tuber Flesh Colour (Carotenoids, Anthocyanins)

Tuber flesh pigments like carotenoids, anthocyanins and their phenolic and flavonoid precursors are regarded as secondary metabolites that are beneficial for human health and nutrition. Their antioxidants potential and many other nutritional aspects are of high value in the human diet. Further enrichment of tubers with these compounds as well as refining the protein content, notably protease inhibitors, will contribute to increase the nutritional value of potato. Knowledge of the genomic positions and identity of the genes controlling tuber pigmentation facilitates the introgression of colourful alleles in novel cultivars.

Generally, tuber flesh colour varies between white, yellow and orange depending on carotenoid content. The biosynthesis and regulation during tuber life cycle of carotenoids have been described at biochemical and molecular level (Lu et al. 2001; Morris et al. 2004).

Initial studies handled tuber flesh colour as qualitative trait, distinguishing yellow and white flesh and disregarding shades of yellowness. This resulted in the identification of the *Y* locus on potato chromosome III (Bonierbale et al. 1988; Fruwirth 1912; Jacobs et al. 1995). When treating tuber flesh colour as a quantitative trait, a major QTL was identified on chromosome III at a similar position as *Y*. This QTL co-localized with the candidate gene phytoene synthase *or* β-carotene hydroxylase (Brown et al. 2006; Thorup et al. 2000). The results of Kloosterman et al. (2010) also support beta-carotene hydroxylase (bch) as being the gene that underlies the *Y* locus. The authors suggested that a dominant allele *B* enhances bch expression levels that translates in an increased enzyme activity and drives the conversion of β-carotene to zeaxanthin. However, the presence or absence of the *B* allele cannot fully explain the observed variation of yellowness. At least two additional, minor QTL for tuber flesh colour were detected on chromosomes IV and XII (Śliwka et al. 2008).

In addition to yellow, orange has been described as another flesh colour phenotype. The orange flesh phenotype was highly correlated with the zeaxanthin content in the tuber and partially associated with an exotic South American *bch* allele (Brown et al. 1993; Brown et al. 2006). A comprehensive study investigating the genetics and molecular biology of orange flesh colour in potato has recently been published (Wolters et al. 2010). Association analysis between SNP haplotypes and flesh colour phenotypes in diploid and tetraploid potato genotypes demonstrated, in agreement with other studies (Kloosterman et al. 2010), that only one dominant beta-carotene hydroxylase 2 allele had an effect on changing white into yellow flesh colour. Furthermore, the analysis of alleles of zeaxanthin epoxidase (Zep) on potato chromosome II showed that all diploid genotypes with orange tuber flesh were homozygous for one specific recessive Zep allele. This Zep allele displayed a reduced expression level, which was correlated with the presence of a non-LTR retrotransposon sequence in intron 1 of the Zep allele. The low expression of the *Zep* allele might result in reduced Zep activity and therefore low amounts of zeaxanthin. The authors

concluded that only genotypes combining the dominant bch allele with homozygosity for the recessive Zep allele accumulated large amounts of zeaxanthin and thus showed orange-fleshed tubers.

Tuber flesh and skin also shows large natural diversity of red, purple and blue pigmentation patterns, especially in Andean landraces. Red and blue pigments are caused by anthocyanins, and their localized expression suggests tissue-specific transcription of the underlying genes. The only locus thought to be implicated in pigmentation of tuber flesh, Pf, was assigned to chromosome X, due to tight linkage to I (De Jong 1987). The I or *D* locus for tuber skin colour has been mapped to chromosome X (van Eck et al. 1994) and is equivalent to the *PSC* (potato skin colour) locus mapped previously (Gebhardt et al. 1989). The *D* gene codes for a R2R3 Myb domain transcriptional regulator of the anthocyanin pathway, a homolog of the *an2* gene from Petunia (Jung et al. 2009). Pf is necessary but not sufficient for anthocyanin-pigmented tuber flesh. QTL mapping was performed in the segregating diploid progeny of a cross between a partially pigmented and an un-pigmented white potato clone (Zhang et al. 2009). The progeny showed continuous tuber flesh pigmentation phenotypes but no segregation of skin colour. QTL influencing the degree of flesh pigmentation were detected on chromosomes V, VIII and IX. CAPS markers for chalcone isomerase (chi) and *Stan1*, the potato homolog of Petunia an1, a basic helix-loop-helix (bHLH) transcriptional regulator of anthocyanin biosynthesis, co-localized with the QTL on chromosomes V and IX, respectively. A 980 bp Stan1 marker fragment was present in all tested clones with pigmented flesh but only in approximately half of the clones with white or yellow flesh (Zhang et al. 2009).

3.4.9 Glycoalkaloids

Steroidal glycoalkaloids are secondary metabolites that occur in most organs of Solanum species (Smith et al. 1996). In cultivated potato the major glycoalkaloids are α-solanine and α-chaconine. As both account for 95 % of the glycoalkaloid content in tubers, they are often referred to as total glycoalkaloids (TGA). The TGA levels increase when tubers are exposed to light, although genotype specific differences were observed (Dale et al. 1993; De Maine et al. 1988)). Due to the toxicity of glycoalkaloids, new potato varieties have to have low levels of TGA. This has to be considered when, for example, resistance factors are introgressed from wild species. Wild Solanum species often accumulate high levels or diverse forms of glycoalkaloids in comparison to those found in cultivated potato (Van Gelder et al. 1988). In order to assist the selection of cultivars with low content of tuber glycoalkaloids, it is necessary to understand the genetic basis of the natural variation of glycoalkaloid levels and composition.

Several studies have identified QTL or markers associated with the synthesis of various foliar glycoalkaloids in potato (Hutvágner et al. 2001; Medina et al. 2002; Ronning et al. 1999; Yencho et al. 1998), whereas investigations on potato

tuber glycoalkaloids remain sparse. One major QTL for both α-solanine and α-chaconine content and for TGA in tubers was mapped on potato chromosome I in diploid populations originating from reciprocal crosses between a *S. tuberosum* with a *S. sparsipilum* clone (Sorensen et al. 2008). The previous studies had located a QTL for glycoalkaloid production in foliage of different Solanum species in a similar position on chromosome I. The microsatellite marker STM5136 was found closely linked to this QTL. The diagnostic power of this marker in tetraploid breeding populations remains to be assessed. Yet, forward and reverse genetic approaches, investigating the molecular mechanisms underlying tuber glycoalkalois content have not been brought together. Only few details are known about the glycoalkaloid biosynthetic pathway, its genes and intermediates (Ginzberg et al. 2009). Candidate genes such as solanidine-UDP-glucose glucosyltransferase (Moehs et al. 1997) or sterol methyltransferases (Arnqvist et al. 2003) implied to control TGA should be tested for co-localization with QTL. Furthermore, their chromosomal position should be determined with the help of the potato genome sequence (PGSC 2011).

3.5 Conclusion and Outlook

The genetic dissection of some twenty five quantitative traits of potato by linkage analysis revealed a broad spectrum of genetic complexity, from a single locus with major QTL effects (e.g. resistance to *G. pallida*, tuber shape) to a dozen ore more QTL with small to intermediate effects (e.g. field resistance to late blight, tuber starch content). Most of these studies have been conducted in diploid, experimental mapping populations, the general agronomic quality of which is usually inferior to advanced tetraploid cultivars used for variety development. Their impact on the 'real world' of competitive, commercial potato breeding has therefore been limited. Nevertheless, these QTL maps provide invaluable knowledge on the approximate genomic positions of the unknown factors that control complex traits and they are essential for the identification of candidate genes underlying natural phenotypic variation, particularly since the first draft of an annotated potato genome sequence became available (PGSC 2011). Association genetics has the advantage that it can be performed directly in populations of tetraploid cultivars related by descent. A small number of association mapping experiments using such materials and DNA polymorphisms in candidate genes resulted in the first diagnostic markers, which can be translated into breeding applications irrespective of whether the association is direct or indirect. With one exception (Simko et al. 2004), all association mapping experiments performed so far used mostly middle European germplasm. Like in human genetics, it will be important to validate marker-trait associations in independent populations with diverse geographical provenance. Whole genome association mapping in tetraploid potato based on genome wide SNPs is now at the doorstep. SNPs are highly informative, co-dominant markers, which allow the dissection of the three heterozygous allelic configurations in tetraploid potato, thereby allowing

to include the allele dosage in the association analysis. A first SNP genotyping chip with 8300 SNPs has been constructed (Hamilton et al. 2011) and is available as Infinium platform (Illumina Inc.). Linkage as well as association mapping based on high throughput SNP genotyping will confirm known and discover novel QTL.

Association mapping based on candidate loci has revealed several known genes and their allelic variants that are very good candidates for being causal contributors to natural trait variation. Comparative analysis of the transcriptome, proteome or metabolome of single genotypes or groups of genotypes with contrasting phenotypic trait values (case-control studies), led to the identification of novel functional candidate genes. Association genetics in combination with the potato genome sequence and *omics* approaches, is expected to make further significant contributions towards the molecular identification of the genes underlying complex traits in potato. The verification of a gene's causal role by a traditional map-based cloning approach (high resolution linkage and/or association mapping followed by complementation analysis of candidate genes) is rather prohibitive for most potato complex traits due to genetic and practical constraints. Like in human genetics, association mapping combined with functional analysis of candidate gene alleles in heterologous model systems is therefore a valid alternative, as performed for *StAOS2* (Pajerowska-Mukhtar et al. 2009; Pajerowska-Mukhtar et al. 2008).

References

Albrecht T, Koch A, Lode A et al (2001) Plastidic (Pho1-type) phosphorylase isoforms in potato (*Solanum tuberosum* L.) plants: expression analysis and immunochemical characterization. Planta 213:602–613

Alcázar R, García AV, Parker JE, Reymond M (2009) Incremental steps toward incompatibility revealed by Arabidopsis epistatic interactions modulating salicylic acid pathway activation. Proc Natl Acad Sci U S A 106:334–339

Allison DB (1997) Transmission-disequilibrium tests for quantitative traits. Am J Hum Genet 60:676–690

Alyokhin A, Baker M, Mota-Sanchez D et al (2008) Colorado potato beetle resistance to insecticides. Am J Pot Res 85:395–413

Arnqvist L, Dutta PC, Jonsson L, Sitbon F (2003) Reduction of cholesterol and glycoalkaloid levels in transgenic potato plants by overexpression of a Type 1 sterol methyltransferase cDNA. Plant Physiol 131:1792–1799

Baayen R, Cochius G, Hendriks H et al (2006) History of potato wart disease in Europe—a proposal for harmonisation in defining pathotypes. Eur J Plant Pathol 116:21–31

Bachem CWB, Speckmann G-J, van der Linde PCG et al (1994) Antisense expression of polyphenol oxidase genes inhibits enzymatic browning in potato tubers. Nat Biotech 12:1101–1105

Bae J, Halterman D, Jansky S (2008) Development of a molecular marker associated with Verticillium wilt resistance in diploid interspecific potato hybrids. Mol Breed 22:61–69

Bagnaresi P, Moschella A, Beretta O et al (2008) Heterologous microarray experiments allow the identification of the early events associated with potato tuber cold sweetening. BMC Genom 9:176

Ballvora A, Ercolano MR, Weiss J et al (2002) The *R1* gene for potato resistance to late blight (*Phytophthora infestans*) belongs to the leucine zipper/NBS/LRR class of plant resistance genes. Plant J 30:361–371

Ballvora A, Flath K, Lübeck J et al (2011) Multiple alleles for resistance and susceptibility modulate the defense response in the interaction of tetraploid potato (*Solanum tuberosum*) with *Synchytrium endobioticum* pathotypes 1, 2, 6 and 18. Theor Appl Genet 123:1281–1292

Beketova M, Drobyazina P, Khavkin E (2006) The *R1* gene for late blight resistance in early and late maturing potato cultivars. Russian J Plant Physiol 53:384–389

Bhaskar PB, Wu L, Busse JS et al (2010) Suppression of the vacuolar invertase gene prevents cold-induced sweetening in potato. Plant Physiol 154:939–948

Bird AR, Brown IL, Topping DL (2000) Starches, resistant starches, the gut microflora and human health. *Curr Issues Intest Microbiol*. 1:25–37

Boccalandro HE, Ploschuk EL, Yanovsky MJ et al (2003) Increased phytochrome B alleviates density effects on tuber yield of field potato crops. Plant Physiol 133:1539–1546

Bonierbale MW, Plaisted RL, Tanksley SD (1988) RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. Genetics 120:1095–1103

Bonierbale MW, Plaisted RL, Tanksley SD (1993) A test of the maximum heterozygosity hypothesis using molecular markers in tetraploid potatoes. Theor Appl Genet 86:481–491

Bonierbale MW, Plaisted RL, Pineda O, Tanksley SD (1994) QTL analysis of trichome-mediated insect resistance in potato. Theor Appl Genet 87:973–987

Bormann CA, Rickert AM, Ruiz RA et al (2004) Tagging quantitative trait loci for maturity-corrected late blight resistance in tetraploid potato with PCR-based candidate gene markers. Mol Plant Microbe Interact 17:1126–1138

Bradshaw J, Mackay G (1994) Potato genetics. CAB International, Wallingford, UK

Bradshaw JE, Hackett CA, Meyer RC et al (1998) Identification of AFLP and SSR markers associated with quantitative resistance to *Globodera pallida* (Stone) in tetraploid potato (*Solanum tuberosum* subsp. *tuberosum*) with a view to marker-assisted selection. Theor Appl Genet 97:202–210

Bradshaw JE, Pande B, Bryan GJ et al (2004) Interval mapping of quantitative trait loci for resistance to late blight [*Phytophthora infestans* (Mont.) de Bary], height and maturity in a tetraploid population of potato (*Solanum tuberosum* subsp. *tuberosum*). Genetics 168:983–995

Bradshaw JE, Hackett CA, Lowe R et al (2006) Detection of a quantitative trait locus for both foliage and tuber resistance to late blight [*Phytophthora infestans* (Mont.) de Bary] on chromosome 4 of a dihaploid potato clone (*Solanum tuberosum* subsp. *tuberosum*). Theor Appl Genet 113:943–951

Bradshaw JE, Hackett CA, Pande B et al (2008) QTL mapping of yield, agronomic and quality traits in tetraploid potato (*Solanum tuberosum* subsp. *tuberosum*). Theor Appl Genet 116:193–211

Brown C, Edwards C, Yang C, Dean B (1993) Orange flesh trait in potato: inheritance and carotenoid content. J Amer Soc Hort Sci 118:145–150

Brown C, Kim T, Ganga Z et al (2006) Segregation of total carotenoid in high level potato germplasm and its relationship to beta-carotene hydroxylase polymorphism. Am J Potato Res 83:365–372

Brugmans B, Hutten R, Rookmaker A et al (2006) Exploitation of a marker dense linkage map of potato for positional cloning of a wart disease resistance gene. Theor Appl Genet 112:269–277

Bryan G, McLean K, Bradshaw J et al (2002) Mapping QTLs for resistance to the cyst nematode *Globodera pallida* derived from the wild potato species *Solanum vernei*. Theor Appl Genet 105:68–77

Bryan G, McLean K, Pande B et al (2004) Genetical dissection of *H3*-mediated polygenic PCN resistance in a heterozygous autotetraploid potato population. Mol Breed 14:105–116

Caromel B, Mugniéry D, Lefebvre V et al (2003) Mapping QTLs for resistance against *Globodera pallida* (Stone) Pa2/3 in a diploid potato progeny originating from *Solanum spegazzinii*. Theor Appl Genet 106:1517–1523

Caromel B, Mugniéry D, Kerlan M-C et al (2005) Resistance quantitative trait loci originating from *Solanum sparsipilum* act independently on the sex ratio of *Globodera pallida* and together for developing a necrotic reaction. Mol Plant Microbe Interact 18:1186–1194

Celis-Gamboa C, Struik PC, Jacobsen E, Visser RGF (2003) Temporal dynamics of tuber formation and related processes in a crossing population of potato (*Solanum tuberosum*). Ann Appl Biol 143:175–186

Chen X, Salamini F, Gebhardt C (2001) A potato molecular-function map for carbohydrate metabolism and transport. Theor Appl Genet 102:284–295

Clark AG, Boerwinkle E, Hixson J, Sing CF (2005) Determinants of the success of whole-genome association testing. Genome Res 15:1463–1467

Collins A, Milbourne D, Ramsay L et al (1999) QTL for field resistance to late blight in potato are strongly correlated with maturity and vigour. Mol Breed 5:387–398

Corsini D, Pavek J, Dean B (1992) Differences in free and protein-bound tyrosine among potato genotypes and the relationship to internal blackspot resistance. Am J Pot Res 69:423–435

Costanzo S, Simko I, Christ BJ, Haynes KG (2005) QTL analysis of late blight resistance in a diploid potato family of *Solanum phureja* × *S. stenotomum*. Theor Appl Genet 111:609–617

D'hoop BB, Paulo M-J, Mank R et al (2008) Association mapping of quality traits in potato (*Solanum tuberosum* L.). Euphytica 161:47–60

D'hoop B, Paulo M, Kowitwanich K et al (2010) Population structure and linkage disequilibrium unravelled in tetraploid potato. Theor Appl Genet 121:1151–1170

Dale MFB, Griffiths DW, Bain H, Todd D (1993) Glycoalkaloid increase in *Solanum tuberosum* on exposure to light. Ann Appl Biol 123:411–418

Dalianis C, Plaisted R, Peterson L (1966) Selection for freedom from after cooking darkening in a potato breeding program. Am J Pot Res 43:207–215

Danan S, Chauvin J-E, Caromel B et al (2009) Major-effect QTLs for stem and foliage resistance to late blight in the wild potato relatives *Solanum sparsipilum* and *S. spegazzinii* are mapped to chromosome X. Theor Appl Genet 119:705–719

Danan S, Veyrieras J-B, Lefebvre V (2011) Construction of a potato consensus map and QTL meta-analysis offer new insights into the genetic architecture of late blight resistance and plant maturity traits. BMC Plant Biol 11:16

De Jong H (1987) Inheritance of pigmented tuber flesh in cultivated diploid potatoes. Am Potato J 64:337–343

De Jong H, Burns V (1993) Inheritance of tuber shape in cultivated diploid potatoes. Am J Pot Res 70:267–284

De Maine MJ, Bain H, Joyce JAL (1988) Changes in the total tuber glycoalkaloid content of potato cultivars on exposure to light. J Agric Sci Camb 111:57–58

Douches DS, Freyre R (1994) Identification of genetic factors influencing chip color in diploid potato (*Solanum* spp). Am Potato J 71:581–590

Draffehn AM, Meller S, Li L, Gebhardt C (2010) Natural diversity of potato (*Solanum tuberosum*) invertases. BMC Plant Biol 10:271

Draffehn AM, Durek P, Nunes-Nesi A, Stich B, Fernie AR, Gebhardt C (2012) Tapping natural variation at functional level reveals allele specific molecular characteristics of potato invertase *Pain-1*. Plant Cell Environ 35:2143–2154

Ducreux LJ, Morris WL, Prosser IM et al (2008) Expression profiling of potato germplasm differentiated in quality traits leads to the identification of candidate flavour and texture genes. J Exp Bot 59:4219–4231

El-Shimi NM (1993) Control of enzymatic browning in apple slices by using ascorbic acid under different conditions. Plant Foods Hum Nutr 43:71–76

Ersoz ES, Yu J, Buckler ES (2007) Applications of linkage disequilibrium and association mapping in crop plants. In: Varshney RK, Tuberosa R (eds) Genomics-assisted crop improvement. Springer, Dordrecht, pp 97–119

Evans K, Trudgill DL (1992) Pest aspects of potato production. Part 1. The nematode pests of potato In: Harris P (ed) The potato crop, 2nd edn. Chapman and Hall, London, pp 438–475

Ewing EE, Struik PC (1992) Tuber formation in potato: induction, initiation and growth. Hortic Rev 14:89–197

Ewing EE, Šimko I, Smart CD et al (2000) Genetic mapping from field tests of qualitative and quantitative resistance to *Phytophthora infestans* in a population derived from *Solanum tuberosum* and *Solanum berthaultii*. Mol Breed 6:25–36

Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. Mol Ecol Notes 7:574–578

Fernie A, Tauberger E, Lytovchenko A et al (2002) Antisense repression of cytosolic phosphoglucomutase in potato (*Solanum tuberosum*) results in severe growth retardation, reduction in tuber number and altered carbon metabolism. Planta 214:510–520

Flint-Garcia SA, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrium in plants. Annu Rev Plant Biol 54:357–374

Flis B, Hermig J, Strzelczyk-Zyta D et al (2005) The *Ry-f(sto)* gene from *Solanum stoloniferum* for extreme resistant to *Potato Virus Y* maps to potato chromosome XII and is diagnosed by PCR marker GP122(718) in PVY resistant potato cultivars. Mol Breed 15:95–101

Freyre R, Douches DS (1994) Development of a model for marker-assisted selection of specific gravity in diploid potato across environments. Crop Sci 34:1361–1368

Freyre R, Warnke S, Sosinski B, Douches DS (1994) Quantitative trait locus analysis of tuber dormancy in diploid potato (*Solanum* spp.). Theor Appl Genet 89:474–480

Fridman E, Zamir D (2003) Functional divergence of a syntenic invertase gene family in tomato, potato, and *Arabidopsis*. Plant Physiol 131:603–609

Friedman M (1997) Chemistry, biochemistry, and dietary role of potato polyphenols. A review. J Agric Food Chem 45:1523–1540

Fruwirth C (1912) Zur Züchtung der Kartoffel. Deutsche Landw Presse 39:565–567

Garcia-Brugger A, Lamotte O, Vandelle E et al (2006) Early signaling events induced by elicitors of plant defenses. Mol Plant-Microbe Interact 19:711–724

Gebhardt C, Valkonen JP (2001) Organization of genes controlling disease resistance in the potato genome. Annu Rev Phytopathol 39:79–102

Gebhardt C, Ritter E, Debener T et al (1989) RFLP analysis and linkage mapping in *Solanum tuberosum*. Theor Appl Genet 78:65–75

Gebhardt C, Ballvora A, Walkemeier B et al (2004) Assessing genetic potential in germplasm collections of crop plants by marker-trait association: a case study for potatoes with quantitative variation of resistance to late blight and maturity type. Mol Breed 13:93–102

Gebhardt C, Menendez C, Chen X et al (2005) Genomic approaches for the improvement of tuber quality traits in potato. Acta Hort 684:85–92

Ghislain M, Trognitz B, del R. Herrera M et al (2001) Genetic loci associated with field resistance to late blight in offspring of *Solanum phureja* and *S. tuberosum* grown under short-day conditions. Theor Appl Genet 103:433–442

Ginzberg I, Tokuhisa J, Veilleux R (2009) Potato steroidal glycoalkaloids: biosynthesis and genetic manipulation. Potato Res 52:1–15

Gould WA (1999) Potato production, processing and technology. CTI Publications, Timonium, MD

Grant M, Lamb C (2006) Systemic immunity. Curr Opin Plant Biol 9:414–420

Hackett CA, Bradshaw JE, McNicol JW (2001) Interval mapping of quantitative trait loci in autotetraploid species. Genetics 159:1819–1832

Hackett CA, Pande B, Bryan GJ (2003) Constructing linkage maps in autotetraploid species using simulated annealing. Theor Appl Genet 106:1107–1115

Hackett CA, Milne I, Bradshaw JE, Luo Z (2007) TetraploidMap for windows: linkage map construction and QTL mapping in autotetraploid species. J Hered 98:727–729

Hamblin MT, Warburton ML, Buckler ES (2007) Empirical comparison of simple sequence repeats and single nucleotide polymorphisms in assessment of maize diversity and relatedness. PLoS ONE 2:e1367

Hamilton J, Hansey C, Whitty B et al (2011) Single nucleotide polymorphism discovery in elite North American potato germplasm. BMC Genom 12:302

Hedley PE, Machray GC, Davies HV et al (1993) cDNA cloning and expression of a potato (*Solanum tuberosum*) invertase. Plant Mol Biol 22:917–922

Hedley PE, Machray GC, Davies HV et al (1994) Potato (*Solanum tuberosum*) invertase-encoding cDNAs and their differential expression. Gene 145:211–214

Hehl R, Faurie E, Hesselbach J et al (1999) TMV resistance gene *N* homologues are linked to *Synchytrium endobioticum* resistance in potato. Theor Appl Genet 98:379–386

Henze RE (1956) Inhibition of enzymatic browning of chlorogenic acid solutions with cysteine and glutathione. Science 123:1174–1175

Hoehenwarter W, Larhlimi A, Hummel J et al (2011) MAPA distinguishes genotype-specific variability of highly similar regulatory protein isoforms in potato tuber. J Prot Res 10:2979–2991

Hougas RW, Peloquin SJ, Gabert AC (1964) Effect of seed parent and pollinator on the frequency of haploids in *Solanum tuberosum*. Crop Sci 4:593–595

Hunt MD, Eannetta NT, Yu H et al (1993) cDNA cloning and expression of potato polyphenol oxidase. Plant Mol Biol 21:59–68

Hutvágner G, Bánfalvi Z, Milánkovics I et al (2001) Molecular markers associated with leptinine production are located on chromosome 1 in *Solanum chacoense*. Theor Appl Genet 102:1065–1071

Isherwood FA (1973) Starch–sugar interconversion in *Solanum tuberosum*. Phytochemistry 12:2579–2591

Jacobs JME, Eck HJ, Arens P et al (1995) A genetic map of potato (*Solanum tuberosum*) integrating molecular markers, including transposons, and classical markers. Theor Appl Genet 91:289–300

Jobling S (2004) Improving starch for food and industrial applications. Curr Opin Plant Biol 7:210–218

Jung C, Griffiths H, De Jong D et al (2009) The potato *developer* (*D*) locus encodes an R2R3 MYB transcription factor that regulates expression of multiple anthocyanin structural genes in tuber skin. Theor Appl Genet 120:45–57

Kamoun S, Smart CD (2005) Late blight of potato and tomato in the genomics era. Plant Dis 89:692–699

Kasai K, Morikawa Y, Sorri VA et al (2000) Development of SCAR markers to the PVY resistance gene Ry_{adg} based on a common feature of plant disease resistance genes. Genome 43:1–8

Kawchuk LM, Hachey J, Lynch DR et al (2001) Tomato *Ve* disease resistance genes encode cell surface-like receptors. Proc Natl Acad Sci U S A 98:6511–6515

Kimura T, Otani M, Noda T et al (2001) Absence of amylose in sweet potato [*Ipomoea batatas* (L.) Lam.] following the introduction of granule-bound starch synthase I cDNA. Plant Cell Rep 20:663–666

Kirkman MA (2007) Global markets for processed potato products. In: Vreugdenhil D, Bradshaw J, Gebhardt C, Govers F, Mackerron DKL, Taylor MA, Ross HA (eds) Potato biology and biotechnology, advances and perspectives. Elsevier, Amsterdam, pp 27–44

Kloosterman B, Oortwijn M, uitdeWilligen J et al (2010) From QTL to candidate gene: genetical genomics of simple and complex traits in potato using a pooling strategy. BMC Genom 11:158

Kossmann J, Abel GJW, Springer F et al (1999) Cloning and functional analysis of a cDNA encoding a starch synthase from potato (*Solanum tuberosum* L.) that is predominantly expressed in leaf tissue. Planta 208:503–511

Kreike CM, de Koning JRA, Vinke JH et al (1994) Quantitatively-inherited resistance to *Globodera pallida* is dominated by one major locus in *Solanum spegazzinii*. Theor Appl Genet 88:764–769

Leister D, Ballvora A, Salamini F, Gebhardt C (1996) A PCR-based approach for isolating pathogen resistance genes from potato with potential for wide application in plants. Nat Genet 14:421–429

Leonards-Schippers C, Gieffers W, Schafer-Pregl R et al (1994) Quantitative resistance to *Phytophthora infestans* in potato: a case study for QTL mapping in an allogamous plant species. Genetics 137:67–77

Li L, Strahwald J, Hofferbert HR et al (2005a) DNA variation at the invertase locus *invGE/GF* is associated with tuber quality traits in populations of potato breeding clones. Genetics 170:813–821

Li X-Q, De Jong H, De Jong DM, De Jong WS (2005b) Inheritance and genetic mapping of tuber eye depth in cultivated diploid potatoes. Theor Appl Genet 110:1068–1073

Li L, Paulo MJ, Strahwald J et al (2008) Natural DNA variation at candidate loci is associated with potato chip color, tuber starch content, yield and starch yield. Theor Appl Genet 116:1167–1181

Li L, Paulo MJ, van Eeuwijk F, Gebhardt C (2010) Statistical epistasis between candidate gene alleles for complex tuber traits in an association mapping population of tetraploid potato. Theor Appl Genet 121:1303–1310

Lincoln SE, Lander E (1989) Mapping genes controlling quantitative traits with Mapmaker/QTL. Whitehead Institute for Biomedical Research Technical Report, Cambridge, MA

Liu X, Song B, Zhang H et al (2010) Cloning and molecular characterization of putative invertase inhibitor genes and their possible contributions to cold-induced sweetening of potato tubers. Mol Genet Genomics 284:147–159

Liu X, Zhang C, Ou Y et al (2011) Systematic analysis of potato acid invertase genes reveals that a cold-responsive member, StvacINV1, regulates cold-induced sweetening of tubers. Mol Genet Genomics 286:109–118

Lorberth R, Ritte G, Willmitzer L, Kossmann J (1998) Inhibition of a starch-granule-bound protein leads to modified starch and repression of cold sweetening. Nat Biotech 16:473–477

Lu W, Haynes K, Wiley E, Clevidence B (2001) Carotenoid content and color in diploid potatoes. J Am Soc Hort Sci 126:722–726

Luo ZW, Hackett CA, Bradshaw JE et al (2001) Construction of a genetic linkage map in tetraploid species using molecular markers. Genetics 157:1369–1385

Mackay TFC (2001) The genetic architecture of quantitative traits. Ann Rev Genet 35:303–339

Malosetti M, van der Linden CG, Vosman B, van Eeuwijk FA (2007) A mixed-model approach to association mapping using pedigree information with an illustration of resistance to *Phytophthora infestans* in potato. Genetics 175:879–889

Marczewski W, Flis B, Syller J et al (2001) A major quantitative trait locus for resistance to potato leafroll virus is located in a resistance hotspot on potato chromosome XI and is tightly linked to N-gene-like markers. Mol Plant Microbe Interact 14:1420–1425

Marczewski W, Flis B, Syller J et al (2004) Two allelic or tightly linked genetic factors at the *PLRV.4* locus on potato chromosome XI control resistance to potato leafroll virus accumulation. Theor Appl Genet 109:1604–1609

Maris B (1966) The modifiability of characters important in potato breeding. Euphytica 15:18–31

Mayton H, Griffiths H, Simko I et al (2010) Foliar and tuber late blight resistance in a *Solanum tuberosum* breeding population. Plant Breeding 129:197–201

McGregor I (2007) The fresh potato market. In: Vreudgenhil D, Bradshaw J, Gebhardt C, Govers F, MacKerron DKL, Taylor MA, Ross HA (eds) Potato biology and biotechnology advances and perspectives. Elsevier, Amsterdam, pp 3–26

Medina TM, Fogelman EF, Chani EC et al (2002) Identification of molecular markers associated with leptine in reciprocal backcross families of diploid potato. Theor Appl Genet 105:1010–1018

Menendez CM, Ritter E, Schafer-Pregl R et al (2002) Cold sweetening in diploid potato: mapping quantitative trait loci and candidate genes. Genetics 162:1423–1434

Meyer RC, Milbourne D, Hackett CA et al (1998) Linkage analysis in tetraploid potato and association of markers with quantitative resistance to late blight (*Phytophthora infestans*). Mol Gen Genet 259:150–160

Moehs CP, Allen PV, Friedman M, Belknap WR (1997) Cloning and expression of solanidine UDP-glucose glucosyltransferase from potato. Plant J 11:227–236

Moloney C, Griffin D, Jones P et al (2010) Development of diagnostic markers for use in breeding potatoes resistant to *Globodera pallida* pathotype Pa2/3 using germplasm derived from *Solanum tuberosum* ssp. *andigena* CPC 2802. Theor Appl Genet 120:679–689

Mondy NI, Munshi CB (1993) Effect of maturity and storage on ascorbic acid and tyrosine concentrations and enzymic discoloration of potatoes. J Agric Food Chem 41:1868–1871

Morris WL, Ducreux L, Griffiths DW et al (2004) Carotenogenesis during tuber development and storage in potato. J Exp Bot 55:975–982

Morrison WR, Karkalas J (1990) Starch. In Dey PM (ed) Carbohydrates. Methods in plant biochemistry, vol 2. Academic Press, London, pp 323–325

Mulder A, Turkensteen LJ (2005) Potato diseases. NIVAP Holland, Den Haag

Murphy JP, Kong F, Pinto DM, Wang-Pruski G (2010) Relative quantitative proteomic analysis reveals wound response proteins correlated with after-cooking darkening. Proteomics 10: 4258–4269

Navarro C, Abelenda JA, Cruz-Oro E et al (2011) Control of flowering and storage organ formation in potato by FLOWERING LOCUS T. Nature 478:119–122

Oberhagemann P, Chatot-Balandras C, Schafer-Pregl R et al (1999) A genetic analysis of quantitative resistance to late blight in potato: towards marker-assisted selection. Mol Breed 5:399–415

Odeny DA, Stich B, Gebhardt C (2010) Physical organization of mixed protease inhibitor gene clusters, coordinated expression and association with resistance to late blight at the *StKI* locus on potato chromosome III. Plant Cell Env 33:2149–2161

Pajerowska-Mukhtar KM, Mukhtar MS, Guex N, et al (2008) Natural variation of potato allene oxide synthase 2 causes differential levels of jasmonates and pathogen resistance in Arabidopsis. Planta 228:293–306

Pajerowska-Mukhtar K, Stich B, Achenbach U et al (2009) Single nucleotide polymorphisms in the *Allene Oxide Synthase 2* gene are associated with field resistance to late blight in populations of tetraploid potato cultivars. Genetics 181:1115–1127

Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. PLoS Genet 2:e190

Perlak FJ, Stone TB, Muskopf YM et al (1993) Genetically improved potatoes: protection from damage by Colorado potato beetles. Plant Mol Biol 22:313–321

Pflieger S, Lefebvre V, Causse M (2001) The candidate gene approach in plant genetics: a review. Mol Breed 7:275–291

PGSC (2011) Genome sequence and analysis of the tuber crop potato. Nature 475:189–195

Plaisted R, Tingey W, Steffens J (1992) The germplasm release of NYL 235–4, a clone with resistance to the Colorado potato beetle. Am Potato J 69:843–846

Portis AR, Li C, Wang D, Salvucci ME (2008) Regulation of Rubisco activase and its interaction with Rubisco. J Exp Bot 59:1597–1604

Powell W, Uhrig H (1987) Anther culture of *Solanum* genotypes. Plant Cell Tissue Organ Cult 11:13–24

Preiss J, Ball K, Smith-White B et al (1991) Starch biosynthesis and its regulation. Biochem Soc Trans 19:539–547

Price AL, Patterson NJ, Plenge RM et al (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38:904–909

Pritchard JK (2001) Deconstructing maize population structure. Nat Genet 28:203–204

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Ravi V, Aked J (1996) Review on tropical root and tuber crops. II. Physiological disorders in freshly stored roots and tubers. Crit Rev Food Sci Nutr 36:711–731

Regierer B, Fernie AR, Springer F et al (2002) Starch content and yield increase as a result of altering adenylate pools in transgenic plants. Nat Biotech 20:1256–1260

Ritter E, Gebhardt C, Salamini F (1990) Estimation of recombination frequencies and construction of RFLP linkage maps in plants from crosses between heterozygous parents. Genetics 125:645–654

Rommens CM, Ye J, Richael C, Swords K (2006) Improving potato storage and processing characteristics through all-native DNA transformation. J Agric Food Chem 54:9882–9887

Ronning CM, Stommel JR, Kowalski SP et al (1999) Identification of molecular markers associated with leptine production in a population of *Solanum chacoense* Bitter. Theor Appl Genet 98:39–46

Ross H (1986) Potato breeding—problems and perspectives. Paul Parey, Berlin

Ross HA, Wright KM, McDougall GJ et al (2011) Potato tuber pectin structure is influenced by pectin methyl esterase activity and impacts on cooked potato texture. J Exp Bot 62:371–381

Rouppe van der Voort JNAM, Lindeman W, Folkertsma R et al (1998) A QTL for broad-spectrum resistance to cyst nematode species (*Globodera* spp.) maps to a resistance gene cluster in potato. Theor Appl Genet 96:654–661

Rouppe van der Voort J, van der Vossen E, Bakker E et al (2000) Two additive QTLs conferring broad-spectrum resistance in potato to *Globodera pallida* are localized on resistance gene clusters. Theor Appl Genet 101:1122–1130

Sandbrink JM, Colon LT, Wolters P, Stiekema WJ (2000) Two related genotypes of *Solanum microdontum* carry different segregating alleles for field resistance to *Phytophthora infestans*. Mol Breed 6:215–225

Sattarzadeh A, Achenbach U, Lubeck J et al (2006) Single nucleotide polymorphism (SNP) genotyping as basis for developing a PCR-based marker highly diagnostic for potato varieties with high resistance to *Globodera pallida* pathotype *Pa2/3*. Mol Breed 18:301–312

Schäfer-Pregl R, Ritter E, Concilio L et al (1998) Analysis of quantitative trait loci (QTLs) and quantitative trait alleles (QTAs) for potato tuber yield and starch content. Theor Appl Genet 97:834–846

Scheele Cv, Svensson G, Rasmussen J (1937) Die Bestimmung des Stärkegehaltes und der Trockensubstanz der Kartoffel mit Hilfe des spezifischen Gewichts. Landw VersStn 127:67–96

Scheidig A, Fröhlich A, Schulze S et al (2002) Downregulation of a chloroplast-targeted β-amylase leads to a starch-excess phenotype in leaves. Plant J 30:581–591

Schwall GP, Safford R, Westcott RJ et al (2000) Production of very-high-amylose potato starch by inhibition of SBE A and B. Nat Biotech 18:551–554

Seo YS, Kim EY, Mang HG, Kim WT (2008) Heterologous expression, and biochemical and cellular characterization of CaPLA1 encoding a hot pepper phospholipase A1 homolog. Plant J 53:895–908

Silva GH, Chase RW, Hammerschmidt R, Cash JN (1991) After-cooking darkening of Spartan Pearl potatoes as influenced by location, phenolic acids, and citric acid. J Agric Food Chem 39:871–873

Simko I (2002) Comparative analysis of quantitative trait loci for foliage resistance to *Phytophthora infestans* in tuber-bearing *Solanum* species. Am J Pot Res 79:125–132

Simko I, McMurry S, Yang HM et al (1997) Evidence from polygene mapping for a causal relationship between potato tuber dormancy and abscisic acid content. Plant Physiol 115:1453–1459

Šimko I, Vreugdenhil D, Jung CS, May GD (1999) Similarity of QTLs detected for in vitro and greenhouse development of potato plants. Mol Breed 5:417–428

Simko I, Costanzo S, Haynes KG et al (2004) Linkage disequilibrium mapping of a *Verticillium dahliae* resistance quantitative trait locus in tetraploid potato (*Solanum tuberosum*) through a candidate gene approach. Theor Appl Genet 108:217–224

Simko I, Costanzo S, Ramanjulu V et al (2006a) Mapping polygenes for tuber resistance to late blight in a diploid *Solanum phureja* × *S. stenotomum* hybrid population. Plant Breed 125:385–389

Simko I, Haynes KG, Jones RW (2006b) Assessment of linkage disequilibrium in potato genome with single nucleotide polymorphism markers. Genetics 173:2237–2245

Śliwka J, Wasilewicz-Flis I, Jakuczun H, Gebhardt C (2008) Tagging quantitative trait loci for dormancy, tuber shape, regularity of tuber shape, eye depth and flesh colour in diploid potato originated from six *Solanum* species. Plant Breed 127:49–55

Smith DB, Roddick JG, Jones JL (1996) Potato glycoalkaloids: some unanswered questions. Trends Food Sci Technol 7:126–131

Song Y-S, Schwarzfischer A (2008) Development of STS markers for selection of extreme resistance (*Rysto*) to PVY and maternal pedigree analysis of extremely resistant cultivars. Am J Pot Res 85:159–170

Song Y-S, Hepting L, Schweizer G et al (2005) Mapping of extreme resistance to PVY (*Ry sto*) on chromosome XII using anther-culture-derived primary dihaploid potato lines. Theor Appl Genet 111:879–887

Sørensen KK, Madsen MH, Kirk HG et al (2006) Linkage and quantitative trait locus mapping of foliage late blight resistance in the wild species *Solanum vernei*. Plant Breed 125:268–276

Sorensen KK, Kirk HG, Olsson K et al (2008) A major QTL and an SSR marker associated with glycoalkaloid content in potato tubers from *Solanum tuberosum* x *S. sparsipilum* located on chromosome I. Theor Appl Genet 117:1–9

Sowokinos JR, Vigdorovich V, Abrahamsen M (2004) Molecular cloning and sequence variation of UDP-glucose pyrophosphorylase cDNAs from potatoes sensitive and resistant to cold sweetening. J Plant Physiol 161:947–955

Spielman RS, McGinnis RE, Ewens WJ (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am J Hum Genet 52: 506–516

Stam P (1993) Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. Plant J 5:739–744

Stevens LH, Davelaar E (1997) Biochemical potential of potato tubers to synthesize blackspot pigments in relation to their actual blackspot susceptibility. J Agric Food Chem 45:4221–4226

Stewart HE, Bradshaw JE, Pande B (2003) The effect of the presence of *R*-genes for resistance to late blight (*Phytophthora infestans*) of potato (*Solanum tuberosum*) on the underlying level of field resistance. Plant Pathol 52:193–198

Stich B, Gebhardt C (2011) Detection of epistatic interactions in association mapping populations: an example from tetraploid potato. Heredity 107:537–547

Stich B, Melchinger A (2009) Comparison of mixed-model approaches for association mapping in rapeseed, potato, sugar beet, maize, and Arabidopsis. BMC Genom 10:94

Stich B, Melchinger AE, Frisch M et al (2005) Linkage disequilibrium in European elite maize germplasm investigated with SSRs. Theor Appl Genet 111:723–730

Stich B, Melchinger A, Piepho H-P et al (2006) A new test for family-based association mapping with inbred lines from plant breeding programs. Theor Appl Genet 113:1121–1130

Suttle JC (2007) Dormancy and sprouting. In: Vreugdenhil D, Bradshaw J, Gebhardt C, Govers F, Mackerron DKL, Taylor MA, Ross HA (eds) Potato biology and biotechnology, advances and perspectives. Elsevier, Amsterdam, pp 287–309

Takken FLW, Joosten MHAJ (2000) Plant resistance genes: their structure, function and evolution. Eur J Plant Pathol 106:699–713

Talburt WF, Smith O (1987) Potato processing, 4th edn. Van Nostrand Reinhold, New York

Tan MYA, Park T-H, Alles R et al (2009) *GpaXI*$_{tar}$ originating from *Solanum tarijense* is a major resistance locus to *Globodera pallida* and is localised on chromosome 11 of potato. Theor Appl Genet 119:1477–1487

Taylor MA, McDougall GJ, Stewart D (2007) Potato flavor and texture. In: Vreugdenhil D, Bradshaw J, Gebhardt C, Govers F, Mackerron DKL, Taylor MA, Ross HA (eds) Potato biology and biotechnology, advances and perspectives. Elsevier, Amsterdam, pp 525–540

Thipyapong P, Stout MJ, Attajarusit J (2007) Functional analysis of polyphenol oxidases by antisense/sense technology. Molecules 12:1569–1595

Thomas P, Adam S, Diehl JF (1979) Role of citric acid in the after-cooking darkening of gamma-irradiated potato tubers. J Agric Food Chem 27:519–523

Thornsberry JM, Goodman MM, Doebley J et al (2001) *Dwarf8* polymorphisms associate with variation in flowering time. Nat Genet 28:286–289

Thorup TA, Tanyolac B, Livingstone KD et al (2000) Candidate gene analysis of organ pigmentation loci in the Solanaceae. Proc Natl Acad Sci U S A 97:11192–11197

Thygesen PW, Dry IB, Robinson SP (1995) Polyphenol oxidase in potato. A multigene family that exhibits differential expression patterns. Plant Physiol 109:525–531

Townsend LR, Hope GW (1960) Factors influencing the color of potato chips. Can J Plant Sci 40:58–64

Urbany C, Colby T, Stich B et al (2011a) Analysis of natural variation of the potato tuber proteome reveals novel candidate genes for tuber bruising. J Proteome Res 11:703–716

Urbany C, Stich B, Schmidt L et al (2011b) Association genetics in *Solanum tuberosum* provides new insights into potato tuber bruising and enzymatic tissue discoloration. BMC Genom 12:7

Valkonen JJP (2007) Viruses: economical losses and biotechnological potential. In: Vreugdenhil D, Bradshaw J, Gebhardt C, Govers F, MacKerron DKL, Taylor MA, Ross HA (eds) Potato biology and biotechnology advances and perspectives. Elsevier, Amsterdam, pp 619–633

van den Berg J, Ewing E, Plaisted R et al (1996a) QTL analysis of potato tuber dormancy. Theor Appl Genet 93:317–324

van den Berg J, Ewing E, Plaisted R et al (1996b) QTL analysis of potato tuberization. Theor Appl Genet 93:307–316

Van der Wolf JM, De Boer SH (2007) Bacterial pathogens of potato. In: Vreudgenhil D, Bradshaw J, Gebhardt C, Govers F, MacKerron DKL, Taylor MA, Ross HA (eds) Potato biology and biotechnology advances and perspectives. Elsevier, Amsterdam, pp 595–617

van Eck HJ (2007) Genetics of morphological and tuber traits. In: Vreugdenhil D, Bradshaw J, Gebhardt C, Govers F, Mackerron DKL, Taylor MA, Ross HA (eds) Potato biology and biotechnology advances and perspectives. Elsevier, Amsterdam, pp 91–115

van Eck HJ, Jacobs JME, van den Berg PMMM et al (1994) The inheritance of anthocyanin pigmentation in potato (*Solanum tuberosum* L.) and mapping of tuber skin colour loci using RFLPs. Heredity 73:410–421

Van Gelder WMJ, Vinke JH, Scheffer JJC (1988) Steroidal glycoalkaloids in tubers and leaves of Solanum species used in potato breeding. Euphytica 37S:147–158

van Loon LC, Rep M, Pieterse CMJ (2006) Significance of inducible defense-related proteins in infected plants. Ann Rev Phytopathol 44:135–162

Van Ooijen JW, Maliepaard C (1996) MAPQTL version 3.0: software for the calculation of QTL positions on genetic maps, 3 edn. CPRO-DLO, Wageningen

Villamon FG, Spooner DM, Orrillo M et al (2005) Late blight resistance linkages in a novel cross of the wild potato species *Solanum paucissectum* (series *Piurana*). Theor Appl Genet 111:1201–1214

Visker M, Keizer L, Van Eck H et al (2003) Can the QTL for late blight resistance on potato chromosome 5 be attributed to foliage maturity type? Theor Appl Genet 106:317–325

Visker M, van Raaij HMG, Keizer LCP et al (2004) Correlation between late blight resistance and foliage maturity type in potato. Euphytica 137:311–323

Visker MHPW, Heilersig HJB, Kodde LP et al (2005) Genetic linkage of QTLs for late blight resistance and foliage maturity type in six related potato progenies. Euphytica 143:189–199

Visser RGF, Somhorst I, Kuipers GJ et al (1991) Inhibition of the expression of the gene for granule-bound starch synthase in potato by antisense constructs. Mol Gen Genet 225:289–296

Wang-Pruski G, Nowak J (2004) Potato after-cooking darkening. Am J Pot Res 81:7–16

Wang-Pruski G, Zebarth B, Leclerc Y et al (2007) Effect of soil type and nutrient management on potato after-cooking darkening. Am J Pot Res 84:291–299

Wasternack C, Kombrink E (2009) Jasmonates: structural requirements for lipid-derived signals active in plant stress responses and development. ACS Chem Biol 5:63–77

Wastie RL (1991) Breeding for resistance. Adv Plant Pathol 7:193–224

Wendt UK, Wenderoth I, Tegeler A, Von Schaewen A (2000) Molecular characterization of a novel glucose-6-phosphate dehydrogenase from potato (*Solanum tuberosum* L.). Plant J 23:723–733

Werij JS, Kloosterman B, Celis-Gamboa C et al (2007) Unravelling enzymatic discoloration in potato through a combined approach of candidate genes, QTL, and expression analysis. Theor Appl Genet 115:245–252

Witek K, Strzelczyk-Żyta D, Hennig J, Marczewski W (2006) A multiplex PCR approach to simultaneously genotype potato towards the resistance alleles *Ry-fsto* and *Ns*. Mol Breed 18:273–275

Wolters AM, Uitdewilligen JG, Kloosterman BA et al (2010) Identification of alleles of carotenoid pathway genes important for zeaxanthin accumulation in potato tubers. Plant Mol Biol 73:659–671

Yang Y, Qiang X, Owsiany K et al (2011) Evaluation of different multidimensional LC-MS/MS pipelines for isobaric tags for relative and absolute quantitation (iTRAQ)-based proteomic analysis of potato tubers in response to cold storage. J Proteome Res 10:4647–4660

Yencho GC, Kowalski SP, Kobayashi RS et al (1998) QTL mapping of foliar glycoalkaloid aglycones in *Solanum tuberosum* × *S. berthaultii* potato progenies: quantitative variation and plant secondary metabolism. Theor Appl Genet 97:563–574

Yu J, Pressoir G, Briggs WH et al (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203–208

Zhang Y, Jung CS, De Jong WS (2009) Genetic analysis of pigmented tuber flesh in potato. Theor Appl Genet 119:143–150

Zheng GH, Sosulski FW (1998) Determination of water separation from cooked starch and flour pastes after refrigeration and freeze-thaw. J Food Sci 63:134–139

Zhou D, Mattoo A, Li N, Imaseki H, Solomos T (1994) Complete nucleotide sequence of potato tuber acid invertase cDNA. Plant Physiol 106:397–398

Zimnoch-Guzowska E, Marczewski W, Lebecka R et al (2000) QTL analysis of new sources of resistance to *Erwinia carotovora* ssp *atroseptica* in potato done by AFLP, RFLP, and resistance-gene-like markers. Crop Sci 40:1156–1167

Zrenner R, Schüler K, Sonnewald U (1996) Soluble acid invertase determines the hexose-to-sucrose ratio in cold-stored potato tubers. Planta 198:246–252

Chapter 4
Introgression Libraries with Wild Relatives of Crops

Silvana Grandillo

Abstract The narrow genetic base of many crops raises concerns about the prospects for continued genetic gains necessary to meeting the increasing demand for agricultural output in an age of climate changes. The development and application of the introgression line (IL) breeding approach was proposed to more efficiently harness the genetic potential stored in exotic germplasm for the improvement of agricultural performance of elite germplasm, thereby expanding the genetic base of our crops. In tomato, the IL approach has been used for almost two decades, and the studies conducted on the *Solanum pennellii* LA0716 ILs (the founding "exotic library") using cutting-edge 'omics' platforms have clearly demonstrated the effectiveness of these congenic and permanent resources in fundamental biology, and for exploring and utilizing the hidden breeding potential of wild species for practical use in agriculture. Since the pioneer studies conducted in tomato, collections of ILs representing different fractions of the exotic parent genome have been developed for a wide range of crops. The results indicate that crop wild relatives are a rich reservoir of potentially valuable alleles, many of which would not have been predicted from the mere phenotypes of the wild plants. Therefore, exotic libraries, combined with the ever-growing body of genomics tools, are expected to further improve the efficiency with which the nature of quantitative trait variation will be unveiled and wild relatives of crops will contribute to face future breeding challenges.

4.1 Introduction

Domestication and breeding of many crops have resulted in relevant improvements in yield and quality, but at the same time they have been coupled to a depletion of the genetic variation present in elite germplasm, causing the loss of valuable alleles originally present in the wild relatives of many crops (Simmonds 1976; Tanksley and McCouch 1997). This problem is particularly severe in self-pollinated crops such as tomato and rice (Miller and Tanksley 1990; Wang et al. 1992). The narrow genetic

S. Grandillo (✉)
Italian National Research Council, Institute of Biosciences and BioResources (CNR-IBBR), Research Division Portici, Via Università 133, 80055 Portici, Naples, Italy
e-mail: grandill@unina.it

R. Tuberosa et al. (eds.), *Genomics of Plant Genetic Resources,*
DOI 10.1007/978-94-007-7575-6_4,

base of modern crop varieties not only makes them more susceptible to diseases, but it also raises concerns about the prospects for continued genetic gains necessary to face the challenges of feeding 9 billion people by the year 2050, ensuring sustainable and global food security in an age of climate change (Godfray et al. 2010; Tester and Langridge 2010; Fridman and Zamir 2012). The above scenario, combined with restrictions on the commercial use of genetically modified plants, has renewed the interest in exploring and exploiting natural biodiversity as a source of novel alleles to improve the productivity, adaptation, quality and nutritional value of crops (Tanksley and McCouch 1997; Zamir 2001; McCouch 2004, 2007; Grandillo et al. 2008; Johal et al. 2008).

Genetic variability is the foundation for any crop breeding program. Nature offers a tremendous wealth of genetic variants of both basic and practical interest, which have been created and selected by nature over millions of years of evolution, as the wild ancestors of most crop plants can still be found in their natural habitats. The value of exotic germplasm, including landraces and wild relatives, as a source of new and useful alleles, that could compensate the loss caused by modern breeding, was recognized already at the beginning of the past century (Bessey 1906; McCouch 2004). Since then, considerable effort and resources have been invested worldwide in large plant collections and preservations, with a particular emphasis given to "exotics", which have resulted in more than 1,400 gene banks with about 6 million accessions representing most of the common crop species (Glaszmann et al. 2010). However, these genetic resources have been only marginally explored and exploited, leaving most of their genetic potential still untapped (Tanksley and McCouch 1997; Glaszmann et al. 2010).

A wider use of exotic germplasm in breeding programs has been hindered by several inherent problems, which are often associated with crosses involving wild and domesticated species, and in part by the lack of adequate techniques that would enable a more efficient discovery and utilization of the valuable alleles present in exotic species. Pre- and post-zygotic barriers, infertility of the segregating generations, suppressed recombination between the chromosomes of the two species, 'linkage drag', as well as the long time and effort necessary to recover the elite parent genetic background, are some of the problems often observed in wide crosses. In addition, much of the unadapted germplasm is phenotypically inferior to elite germplasm for many of the traits that breeders would like to improve. As a result, most plant breeding programs have relied, and still rely, on reshuffling the same set of genes/alleles already available in the elite lines, reducing the overall genetic variation available for future sustained crop improvements. In general, the use of exotic germplasm has mostly focused on major genes for disease and insect resistance (Plunknett et al. 1987) as shown by the high number of resistance genes derived from wild species, which can be found in elite lines (Zamir 2001; Hajjar and Hodgkin 2007). In contrast, its use as a source of valuable alleles for the improvement of other traits relevant to agriculture such as yield, stress tolerance and quality has been more limited, with differences depending on the crops (Hajjar and Hodgkin 2007). Such traits, in fact, are often quantitatively inherited, displaying continuous variation and resulting from the segregation of numerous interacting quantitative trait loci (QTL), with varying

magnitude of effect, whose expression is modified by the genetic background and the environment (Mackay 2001).

Over the past decades, improved interspecific hybridization techniques along with advances in quantitative genetics and genomic technologies have provided the necessary tools to overcome some of the difficulties associated with the use of exotic germplasm for the improvement of complex traits. High-density molecular genetic maps have allowed for the identification and characterization of single QTL contributing to complex traits while their fine-mapping allows us to distinguish pleiotropy from close linkage and, importantly, to reduce the negative effects of linkage drag (Tanksley 1993; Eshed and Zamir 1996; Frary et al. 2003). Furthermore, QTL mapping studies have also provided stronger evidence that low-performing wild and unadapted species can contribute agronomically favorable QTL alleles associated with transgressive segregation observed in several interspecific crosses that have the potential to improve yield, as well as other important traits (de Vicente and Tanksley 1993; Eshed and Zamir 1995; Tanksley et al. 1996; Xiao et al. 1996, 1998; McCouch et al. 2007; Grandillo et al. 2008). These results indicate that the phenotype of wild species is a poor predictor of their breeding value, and that the domestication process has "left behind" many favorable alleles, which could now be more efficiently "recovered" using innovative genomic-assisted breeding strategies (Tanksley and McCouch 1997; Zamir 2001; McCouch 2004; Cavanagh et al. 2008; Johal et al. 2008).

However, despite the numerous QTL-mapping studies conducted and reported for many crops, the contribution of QTL analysis to breeding new varieties has been limited. In order to bridge the gap between QTL mapping and variety development based on the use of unadapted germplam, two related molecular breeding strategies, the "Advanced Backcross QTL analysis (AB-QTL)" (Tanksley and Nelson 1996) and "exotic libraries" or introgression line (IL) libraries (Eshed and Zamir 1995; Zamir 2001) have been developed and tested in several crops (Table 4.1) (Lippman et al. 2007; McCouch et al. 2007; Grandillo et al. 2008, 2013; Swamy and Sarla 2008; Tan et al. 2008; Ali et al. 2010; Buerstmayr et al. 2011; Blair and Izquierdo 2012; Sayed et al. 2012; Varshney et al. 2013). These two approaches were proposed to more efficiently harness the genetic potential stored in seed banks and in exotic germplasm for the improvement of elite germplasm, thereby expanding the genetic base of crop species (Tanksley and McCouch 1997; Zamir 2001). Both methods have allowed the identification of favorable wild QTL alleles for numerous traits of agronomical interest and the development of introgression lines (prebred) that can be used in marker-assisted breeding programs.

Both strategies have been covered in other reviews (Lippman et al. 2007; McCouch et al. 2007; Grandillo et al. 2008; Ali et al. 2010). This paper will focus on the IL-approach, providing an overview of the results achieved over the last 20 years in tomato as well as in other crops. Considering that the principles of the IL approach were first outlined and successfully applied in tomato, a particular emphasis will be given to the efforts and accomplishments achieved within the tomato clade.

Table 4.1 Introgression libraries of crops derived from interspecific crosses

Crop	Donor parent	Recipient parent	No. of ILs and/or pre-ILs developed and/or (tested)	Estimated donor genome coverage (%)	Traits analyzed/other	References[a]
Tomato (*Solanum lycopersicum* L.)	*S. pennellii* (acc. LA0716)	cv. M82	50	100	Yield-related and fruit quality	Eshed and Zamir 1994, 1995; (Alpert et al. 1995; Alpert and Tanksley 1996; Eshed and Zamir 1996; Eshed et al. 1996; Frary et al. 2000; Fndman et al. 2000; Monforte et al. 2001; Fridman et al. 2002, 2004; Gur and Zamir 2004)
			(6)		Sugars and acid content/transcriptional profiling	Baxter et al. 2005
			(1)		Drought tolerance/fine mapping	Xu et al. 2008
			(50)		Salt tolerance, secondary metabolites	Frary et al. 2010
			(2)		Drought tolerance/ trascriptional profiling	
			(46)		Competence for adventitious organ formation	Arikita et al. 2013
			(50)		Resistance to bacteial spot	Sharlach et al. 2013
	S. pennellii (acc. LA0716)	cv. M82	76	100	NA[b]	Liu and Zamir 1999; Pan et al. 2000; httD: // sokenomics. net/
			(53)		Fusarium resistance	Sela-Buurlage et al. 2001
			(4)		Fruit volatile Compounds	Tadmor et al. 2002
			(58)		Leaf morphology	Holtan and Hake 2003
			(75)		Intensity of red color of ripe fruit/candidate gene approach	Liu et al. 2003

Table 4.1 (continued)

Crop	Donor parent	Recipient parent	No. of ILs and/or pre-ILs developed and/or (tested)	Estimated donor genome coverage (%)	Traits analyzed/other	References[a]
			(70)		Fruit size and composition/candidate gene approach	Causse et al. 2004
			(6)		Metabolite profiling	Overy et al. 2005
			(50)		Nutritional and antioxidant	Rousseaux et al. 2005
			(65) (IL & HILs)[c]		Fruit primary metabolisms, morphology and yield	Schauer et al. 2006; 2008
			(68/76) (IL & HILs)		Morphology and yield	Semel et al. 2006
			(74)		Fruit volatile Compounds and citric acid	Tieman et al. 2006; (Mageroy et al. 2006)
			(70)		Ascorbic acid	Stevens et al 2007, (2008)
			(71)		Hybrid incompatibility	Moyle and Nakazato 2008; (Bedinger et al. 2011)
			(2)		Ascorbic acid, phenolics accumulation/ transcriptional profiling	Di Matteo et al. 2010, 2013;(Sacco et al. 2013)
			(1) (IL & HILs)		Harvest index, earliness, metabolites/fine mapping	Gur et al. 2010
			BIN[d]: 1C, 2D, 41, 7H, 11C		Metabolism and yield-related /integration of physical and genetic map for 5 genomic regions (BINs 1C, 2D, 41, 7H	Kamenetzky et al. 2010
			(65)		Trichome specialized metabolites	Schilm iiier et al. 2010, (2012)
			(9)		Alfa-tocopherol content (vitE)/candidate genes	Almeida et al. 2011

Table 4.1 (continued)

Crop	Donor parent	Recipient parent	No. of ILs and/or pre-ILs developed and/or (tested)	Estimated donor genome coverage (%)	Traits analyzed/other	References[a]
			(75) (IL & HILs)		Yield-related (shoot and root)/grafting	Gur et al. 2011
			(76) (IL & HILs)		Enzyme activity in fruit pericarp	Steinhauser et al. 2011
			(76)		Fruit firmness/ fine mapping	Chapman et al. 2012
			(76)/(23)		Carotenoids/transcriptional profiling	Lee et al. 2012
			(3)		Polyphenols	Minutolo et al. 2013
			(75)		DArT markers	Van Schalkwyk et al. 2012
			(4)		Transgressive phenotypes, sRNAs	Shivaprasad et al. 2012
			(75)		SNPs	Sim et al. 2012
			(76)		Seed metabolism	Toubiana et al. 2012
			(1)		Tolerance to salt stress and blossom-end rot	Uozumi et al. 2012
			(76)		Root morpholgy and cellular development	Ron et al. 2013
	S. pennellii (acc. LA0716)	cv. Moneyberg	72 (11)		Attractivenes toB. tabaci, headspace volatiles	Bleeker et al. 2009
	S. chmielewskii (acc. LAI 840)	cv. Moneyberg	55	~83	NA	Peleman and van der Voort 2003; S. Grandillo personal communication
			(20)		Fruit quality under different fruit loads	Prudent et al. 2009, 2010, 2011
			(23)		Fruit metabolome under different fruit loads	Do et al. 2010

Table 4.1 (continued)

Crop	Donor parent	Recipient parent	No. of ILs and/or pre-ILs developed and/or (tested)	Estimated donor genome coverage (%)	Traits analyzed/other	References[a]
	S. habrockaites (acc. LA1777)	cv. E6203	99 (ILs & pre-ILs)	~85	NA	Monforte and Tanksley 2000a; (Tripodi et al. 2010; S Grandillo, personal communication)
			(2)		Yield- and fruit quality-related	(Monforte and Tanksley 2000b; Monforte et al. 2001; Yates 2004)
			(38)		Sesquiterpenes	van der Hoeven et al. 2000
			(71)		Hybrid incompatibility	Moyle and Graham 2005; (Moyle and Nakazato 2008)
			(39)		Ripening-associated ethilene emission	Dal Cin et al. 2009
			(89)		Fruit volatiles	Matbieu et al. 2009
			(93)		Cold tolerance/trascriptome analysis	Liu et al. 2012
	S. habrockaites (acc. LYC4)	cv. Moneymaker	30	~95	Gray mold resistance	Finkers et al. 2007
	S. neorickii (acc. LA2133)	cv. E6203	142	~90	Yield- and fruit quality-related	Fulton et al. 2000; D. Zamir and S. Grandillo, persoal communication
	S. pimpinellifolium (acc. LAI 589)	cv. E6203	196	NA	Yield- and fruit quality-related	Doganlar et al. 2002; D. Zamir and S. Grandillo, persoal communication
	S. pimpinellifolium (acc. TO-937)	cv. Money-maker	54	~98	Fruit quality	W Barrantes and AJ Monforte, personal communication
	S. lycopersicoides (acc LA2951)	cv. VF36	90	~96	NA	Chetelat and Meglic 2000 (Canady et al. 2005)

Table 4.1 (continued)

Crop	Donor parent	Recipient parent	No. of ILs and/or pre-ILs developed and/or (tested)	Estimated donor genome coverage (%)	Traits analyzed/other	References[a]
Barley (*Hordeum vulgare* spp. *vulgare*)	*H. vulgare* ssp. *spontaneoum* (acc. ISR42-8)	cv. Scarlett & cv. Thuringia	49 & 43 (pre-ILs)	~98.1 & ~93.0	Days until heading	von Korff et al. 2004
	H. vulgare ssp. *spontaneoum* (acc. ISR42-8)	cv. Scarlett	59	~86.6	Resistance to powdery mildew and leaf rust	Scbmalenbach et al. 2008
			73(39)		Yield-related	Sbmalenbach and Pülen 2009
			73(39)		Yield-related	Scbmalenbach et al. 2008
			73(39)		Flowering time and agronomic	Wang et al. 2010
			73	87.3	Threshability locus and SNP analysis	Sbmalenbach et al. 2011
			73(39)		Malting quality, proteomic proflies	March et al. 2012
			73(42)		Nitrogen deficiency	Hoffinanetal. 2012
	H. vulgare ssp. *spontaneoum* (H602)	cv. Haruna Nijo	99		SNP analysis	Hon et al. 2005; (Sato and Takeda 2009)
	H. vulgare ssp. *spontaneoum* (acc. Caesarea 26–24)	Harrington	NA	NA	Agronomic and malting quality	Matus et al. 2003

Table 4.1 (continued)

Crop	Donor parent	Recipient parent	No. of ILs and/or pre-ILs developed and/or (tested)	Estimated donor genome coverage (%)	Traits analyzed/other	References[a]
	H. bulbosum	NA	88	~36	NA	Johnston et al. 2009
Cotton (*Gossypium hirsutum* L.)	*G. hirsutum* TM-1 (cultivated)	G. barbadense Hai7124 (cultivated)	17	NA	Fiber quality and agronomic	Stelly et al. 2005; (Saba et al. 2006, 2010, 2011, 2013)
	G. hirsutum SG747 (cultivated)	G. barbadense Giza 75 (cultivated)	146BILs	NA	Yield and fiber quality	Yu et al. 2013
Indian mustard (*Brassica juncea* L.)	Artificially synthesized amphiploid (*B. fruticulosa* x *B. rapa* var. brown sarson)	cv. BSH	45 selected pre-ILs	NA	Resistance to mustard aphid	Atri et al. 2012
Lettuce (*Lactuca sativa* L.)	*L. saligna* (CGN 5271)	cv. Olof	28	> 90	Simple morphological	Jeuken and Lindhout 2004
Peanut (*Arachis hypogaea* L.)	*A. cardenasii*	NA	46	30	NA	Garcia et al. 1995
	Wild synthetic amphidiploid	cv. Fleur11	122 (80)	88.7	Plant morphology	Fonceka et al. 2009, 2012

Table 4.1 (continued)

Crop	Donor parent	Recipient parent	No. of ILs and/or pre-ILs developed and/or (tested)	Estimated donor genome coverage (%)	Traits analyzed/other	References[a]
Rice[e] (*Oryza sativa* L.)	*O. glumaepatula* (IRGC acc. 105668)	cv. Taichung 65	59	~ 100	NA	Sobrizal et al. 1996
	O. meridionalis	cv. Taichung 65	61	~ 100	NA	Kurakazu et al. 2001
	O. rufipogon Griff.	ssp. indica cv. Guicao 2	159	67.5	Yield-related	Tian et al. 2006
	O. rufipogon Griff. (acc. YJCWR)	ssp. indica cv. Teqing	120	100	Yield-related	Tan et al. 2007
			85	NA	Contents of mineral nutrients in grain	Gracia-Oleveria et al. 2009
			87	100	Salt tolerance	Gracia-Oleveria et al. 2009
	O. rufipogon (IRGC 105491)	ssp. indica cv. Zhenshan 97B	105	NA	NA	Ali et al. 2010
	O. glaberrima MG12 (acc. IRGC103544)	ssp. Tropical aponica *cv. Caiapo*	64	~ 100	Resistance to stripe necrosis virus, yield-related	Doi et al 1997; Gutirrez et al 2010

Table 4.1 (continued)

Crop	Donor parent	Recipient parent	No. of ILs and/or pre-ILs developed and/or (tested)	Estimated donor genome coverage (%)	Traits analyzed/other	References[a]
	O. minuta (acc. 101133)	ssp. indica cv. IR24	131	100	Resistance to bacterial bright, brown planthopper, whitebacked planthopper	Guo et al. 2012
Rye (*Secale cereale* L.)	Iranian primitive rye population Altevogt 14160	Inbred line L2053-N	2 populations of 40 pre-ILs	74 and 59	Agronomic and quality traits Pollen fertility restoration	Falke et al. 2008 (2009a; Mahone et al. 2012) Falke et al. 2009c
			39 pre-ILs	NA	Pollen fertility restoration	Falke et al. 2009c
Wheat (*Triticum aestivum* L.)	Sear's "Synthetic 6x" (derived from tetraploid emmer x wild grass *Aegilops tauschii*)	Chinese Spring	84 (52)	~ 100	Yield-related	Petsova et al. 2001, 2006
	Am3 exotic exaploid (derived from *T. carthlicum* x Aegilops taushii)	Laizhou953	97	37.7	Yield-related	Liu et al. 2006

[a] Follow-up studies are indicated in parenthesis
[b] *NA* not available
[c] *IL & HILs* the lines were tested in homozygosity (IL) and heterozygosity (HIL)
[d] *BIN* overalp between contiguous donor chromosomal segments (Pan et al. 2000)
[e] Additional IL and BIL populations are being developed using other Oryza wild species accessions (Ali et al. 2010)

4.2 IL-Based Analyses of Complex Traits

Most traits of biological and economic interest are of a quantitative nature, making the elucidation of their genetic and molecular bases a notoriously challenging task. Over the past decades numerous different types of segregating populations have been used for QTL mapping in plants (Cavanagh et al. 2008). Many QTL have been identified either using biparental populations exploiting recent recombinations, or using association analysis, which exploits historical recombination. At the beginning, early biparental segregating generations (F_2, F_3 and BC_1) or recombinant inbred lines (RILs) have been widely used. However, these populations have several limitations caused by the high proportion of donor parent alleles that still segregate, including the overshadowing effect of major QTL on the effects of independently segregating minor QTL, or the relatively high level of epistatic interactions that occur between donor QTL alleles and other donor genes. As a consequence, favorable donor QTL alleles identified in these mapping populations often lose their effects once they are introgressed into the genetic background of elite lines. In the case of interspecific crosses involving exotic germplasm, partial or complete sterility problems further complicate QTL analyses, since a few genes for sterility may hamper population development and/or the analysis of agronomical important traits (such as fruit characters).

In order to circumvent these limitations, and to gain an insight into the genetic factors underlying differences between the cultivated tomato (*Solanum lycopersicum* L.) and its wild relatives, Zamir and colleagues used RFLP (restriction fragment length polymorphism) markers to construct the first complete set of substitution lines in tomato (referred to as introgression lines—ILs), consisting of 50 near isogenic lines (NILs) carrying single marker-defined homozygous chromosomal segments of the wild green-fruited species *S. pennellii* (acc. LA0716) in an otherwise homogeneous genetic background of the processing inbred cv. M82 (Eshed and Zamir 1994, 1995). The whole donor genome is represented by the complete panel of overlapping homozygous chromosomal segments, and it is a permanent mapping population since it can be maintained by self-pollination. One of the earliest examples of this kind of genetic resources was reported by Kuspira and Unrau (1957), who analyzed quantitative traits in common wheat using whole-chromosome substitution lines (CSLs), in which the introgressions span complete chromosomes. Subsequently, to define the position of genes on substitution chromosomes, recombinant inbred chromosome substitution lines (RICSLs) have been developed (Cavanagh et al. 2008).

Since the pioneer studies conducted by Kuspira and Unrau (1957) and by Eshed and Zamir (1995, 1996) and the theoretical landmark laid by Tanksley and Nelson (1996), sets of introgression lines representing different fractions of the exotic (wild species or landrace varieties) parent genome have been developed for various crops including, barley, cotton, indian mustard, lettuce, peanut, rice, rye, and common wheat (Table 4.1). In other cases, such as cabbage (Ramsay et al. 1996), tomato (Causse et al. 2007), rice (Li et al. 2005; Ashikari and Matsuoka 2006; Mei et al. 2006; Zhao et al. 2009; Xu et al. 2010; Gu et al. 2012), melon (Eduardo et al. 2005, 2007;

Fernandez-Silva et al. 2010) and maize (Szalma et al. 2007; Pea et al. 2009; Salvi et al. 2011), ILs have been obtained using intraspecific crosses. Sets of introgression lines have also been constructed for the model species *Arabidopsis thaliana* using the three accessions Columbia, Landsberg and Niederzenz (Koumproglou et al. 2002; Torjék et al. 2008).

In the case of crosses involving cultivated and exotic germplasm, these congenic populations have been referred to as "exotic libraries" (Zamir 2001). However, since populations of ILs have been developed also using adapted germplasm as donor parents and from intraspecific crosses, in more general terms they can be referred to as "IL populations" or "IL libraries". Furthermore, while ideally an IL library should be made up of lines each containing a single chromosomal segment deriving from the donor parent, in practice, in many cases several lines in the population may still carry multiple donor introgressions (hereafter referred to as pre-ILs) and the whole set of ILs might cover variable portions of the donor genome (Table 4.1).

Although these populations are very similar in essence, different names have been used including "Introgression Lines (ILs), Backcross Recombinant Inbred Lines (BCRILs), Near Isogenic Lines (NILs) or QTL-NILs, Chromosome Segment Substitution Lines (CSSLs), Backcross Inbred Lines (BILs), Recombinant Chromosome Substitution Lines (RCSL) (see references in Table 4.1), as well as 'Stepped Aligned Inbred Recombinant Strains' (STAIRS) (Koumproglou et al. 2002), NILs (Keurentjes et al. 2007) and ILs (Torjék et al. 2008) in Arabidopsis. As mentioned before, a special case of IL populations are chromosome substitution lines such as those developed in Arabidopsis (Koumproglou et al. 2002) and cotton (Saha et al. 2006).

Similar population structures have also been produced for model animal species such as "Chromosome Substitution Strains (CSSs)" in mice (Singer et al. 2004), ILs in *Caenorhabditis elegans* (Doroszuk et al. 2009) and in Drosophila (Fang et al. 2012), and "Segmental Introgression Lines (SILs)" in parasitic wasp (Desjardins et al. 2013).

The process of IL production involves some backcrossing scheme aided by marker analysis during or after the backcross, followed by one or more generations of self-fertilization to fix the lines (Zamir 2001). The main factors influencing the efficiency of foreground and background selection are the breeding scheme, the selection strategy and the population sizes (Falke et al. 2009b; Falke and Frisch 2011). The production of such congenic and permanent resources is quite a laborious and time-consuming task which can take several years. However, the advent of high-throughput marker technologies has provided the necessary tools to make IL development a much more efficient and precise process (Severin et al. 2010; Xu et al. 2010; Schmalenbach et al. 2011).

In many instances, ILs have been used to confirm, stabilize and fine-map QTL identified in other population structures and therefore only a relatively small proportion of the donor parent genome was represented among the developed ILs (Paterson et al. 1990; Szalma et al. 2007). On the other hand, the availability of whole-genome IL populations allows screening for QTL of the entire genome (Eshed and Zamir 1995).

Several properties of these libraries of introgression lines contribute to their power in identifying and stabilizing QTL, and they have been thoroughly discussed elsewhere (Zamir 2001; Keurentjes et al. 2007; Lippman et al. 2007; Grandillo et al. 2008). In summary, in the ideal case of IL libraries made up of lines each containing a single donor parent introgression, all the phenotypic differences between an IL and the recurrent parent should be due to the allelic differences at one or more genes within the introgressed chromosomal segment. This should reduce much of the genetic background "noise", thus increasing the ability to statistically identify small phenotypic effects using a simple statistical procedure. Another important aspect of these congenic mapping populations is their "immortal nature" with a characterized genotype which eliminates the need of making crosses and of genotyping, but it also allows replicated measurements of the same line, reducing the effect of the environment and increasing the power of QTL detection. The permanent nature of these lines not only facilitates more accurate estimates of the mean phenotypic values, but replicated trials of the same line can be analyzed in different years and/or environments, which allows to estimate the extent of QTL × environment interactions (Monforte et al. 2001; Liu et al. 2003; Gur and Zamir 2004). Multiple data can be collected in different laboratories on the same lines also for multiple, even invasive and destructive traits, thereby creating a comprehensive phenotypic database for general access (Zamir 2001; Gur et al. 2004). Since the lines in the library differ from the recurrent parent by only a single chromosomal segment derived from the donor parent, their phenotypes generally resemble that of the recipient parent, which, in the case of crosses between cultivated and exotic germplasm, reduces the sterility problems that occur in other mapping population structures characterized by a higher frequency of the exotic parent genome, and also allows the lines to be evaluated for yield-associated traits. However, the advantage of single-introgressed segment ILs in resolving individual QTL is also a drawback, as epistatic interactions between unlinked loci, which are a major component of the phenotypic variation, cannot be directly estimated.

The map resolution of a population of ILs is defined by the overlap between contiguous donor introgressions (bins) to which genes or QTL can be assigned by comparing lines (Pan et al. 2000; Liu et al. 2003; Paran and Zamir 2003). The number, length and overlap of adjacent segments define bin lengths, which vary across the genome. One drawback of IL libraries is their initial relatively low level of map resolution, which in the extreme case of whole-chromosome substitution lines corresponds to the entire chromosome. However, each IL represents the starting point by which the phenotypic effects of QTL can be fine-mapped to smaller intervals (Paterson et al. 1990).

Higher resolution mapping of QTL allows us to assess whether the effect on the phenotype is due to a single QTL or to several tightly linked QTL affecting the same trait, as well as to verify whether possible undesirable effects are caused by linkage drag of other genes or by pleiotropic effects of the selected QTL (Eshed and Zamir 1996; Monforte and Tanksley 2000b; Monforte et al. 2001; Fridman et al. 2002; Frary et al. 2003; Chen and Tanksley 2004; Yates et al. 2004; Gur et al. 2010). For instance, high-resolution mapping of the *Brx9-2-5* QTL (affecting total

soluble solids of tomato fruit) in two divergent genetic backgrounds, indeterminate glasshouse tomatoes and determinate open-field varieties), enabled the mapping of a new pleiotropic QTL for the same trait that interacts with the genetic background (Fridman et al. 2002). Another example is provided by fine mapping of the major QTL *stigma exsertion 2.1* (*se2.1*), which revealed a complex locus composed of at least five closely linked genes: three controlling stamen length, one controlling style length and one conditioning anther dehiscence (Chen and Tanksley 2004). Of these five loci, the locus controlling style length (*Style 2.1*) accounted for the greatest change in stigma exertion and was subsequently cloned (Chen et al. 2007).

Besides reducing linkage drag, the development of lines with smaller introgressions (sub-ILs) allows molecular markers to be found which are more tightly linked to the QTL of interest that can be used for marker-assisted breeding. Desirable donor QTL alleles identified in IL populations can be combined in multiple-ILs by means of marker-aided QTL pyramiding approaches to improve the performance of elite lines (Gur and Zamir 2004; Ashikari and Matsuoka 2006; Zong et al. 2012; Sacco et al. 2013) (also see Sect. 4.3.1). Once introgressed, chromosome segments have been subdivided and targeted, and QTL-containing lines have been created, crosses between the lines can be used to study the phenotypic effects of QTL interactions to better understand the nature of epistasis (Tanksley 1993; Eshed and Zamir 1996; Causse et al. 2007). ILs can also be used to obtain more precise estimates of the magnitude of QTL × genetic background interaction (Eshed and Zamir 1995, 1996; Monforte et al. 2001; Gur and Zamir 2004).

Introgression lines are also a powerful tool to study the genetic basis of heterosis, since homozygous lines in a library can be crossed to different tester lines, allowing the effects of heterozygosity on the phenotype to be investigated (Semel et al. 2006), for the positional cloning of key genes underlying quantitative traits (Frary et al. 2000; Fridman et al. 2000, 2004; Salvi and Tuberosa 2005; Uauy et al. 2006; Cong et al. 2008), and for systems-based analyses aimed at identifying genes controlling complex developmental networks (Lippman et al. 2007; Lee et al. 2012; Toubiana et al. 2012) (see Sects. 4.3.1 and 4.3.2).

4.3 The IL Approach in the Tomato Clade

4.3.1 The S. pennellii LA0716 Exotic Library

Members of *Solanum* sect. *Lycopersicon*—the clade containing the cultivated tomato (*Solanum lycopersicum* L.) and its 12 wild relatives—along with the four allied species in the immediate outgroups *Solanum* sects. *Juglandifolia* and *Lycopersicoides,* are adapted to a wide variety of environmental conditions, which correspond to a wide range of variation in terms of morphological, physiological, mating system and biochemical characteristics (Peralta et al. 2008). Due to the low genetic variation of cultivated germplasm (Miller and Tanksley 1990), tomato wild species

have played an important role as sources of useful genes, and for the development of mapping populations (Rick 1982; Grandillo et al. 2011, 2013).

Solanum pennellii LA0716, is a small green-fruited desert species characterized by unique phenotypes. It is distantly related to cultivated tomato, yet the two species are sexually compatible and produce fertile hybrids. In 1969, Rick reported the development of tomato introgression lines using three chromosome segments from *S. pennellii* and recessive mutant chromosome stocks from *S. lycopersicum*. Subsequently, the development of DNA marker technology allowed the use of *S. pennellii* LA0716 as the founding donor parent of the first whole-genome exotic library in tomato (Eshed and Zamir 1994, 1995).

This population, initially consisting of 50 ILs in the genetic background of the elite inbred variety M82, allowed the identification of yield-associated QTL, and the analysis of their epistatic and environmental interactions (Eshed and Zamir 1995, 1996). These first studies also demonstrated the higher efficiency of IL populations in detecting QTL compared with conventional segregating populations such as F_2, BC_1 or RILs (Zamir and Eshed 1998). To increase the mapping resolution of the *S. pennellii* LA0716 'exotic library' additional 26 sub-ILs were added and the resulting 76 lines partition the entire genetic map into 107 bins, which are defined by singular or overlapping segments (Fig. 4.1), each with an average length of 12 cM (Liu and Zamir 1999; Pan et al. 2000; Liu et al. 2003; http://solgenomics.net/). More recently, as part of a EU project (EU-SOL, http://www.eu-sol.net/), the *S. pennellii* IL library was expanded through the addition of > 400 sub-ILs (Lippman et al. 2007; D. Zamir, personal communication). Furthermore, in order to allow the estimation of the relative contributions of epistasis to the phenotypic diversity, a new *S. pennellii* LA0716-based population of several hundreds BILs was constructed in the M82 background (D. Zamir, personal communication). Each BIL genotype carries multiple wild species introgressions permitting phenotypes to be associated with specific epistatically interacting QTL. Individual ILs and sub-ILs can then be used to reconstruct any epistasis detected in the BILs and to study the genetic and developmental components underlying the specific interactions.

Over the years, the *S. pennellii* LA0716 ILs have been evaluated for hundreds of traits allowing the identification of over ~2,800 QTL (Table 4.1) (Lippman et al. 2007; Grandillo et al. 2011, 2013). Repeated measurements have been conducted by multiple labs for numerous yield-associated, fruit morphology and biochemical traits, and the resulting raw data have been deposited in the phenotype warehouse of Phenom Networks <http://phnserver.phenome-networks.com/>.

An important aspect of IL biology, especially in the context of interspecific crosses, is the exposure of new transgressive phenotypes, not present in the parental lines. This phenomenon is caused by novel epistatic relationships between the donor parent alleles, and the independently evolved molecular networks of the recipient parent (Lippman et al. 2007; McCouch et al. 2007; L'Hôte et al. 2010).

In the *S. pennellii* ILs, transgressive phenotypes have been observed for both qualitative and quantitative traits (Lippman et al. 2007). One clear example is fruit color. In fact, while mature fruits of most cultivated tomato varieties are red and those of *S. pennellii* are green, some ILs show novel fruit color variation such as

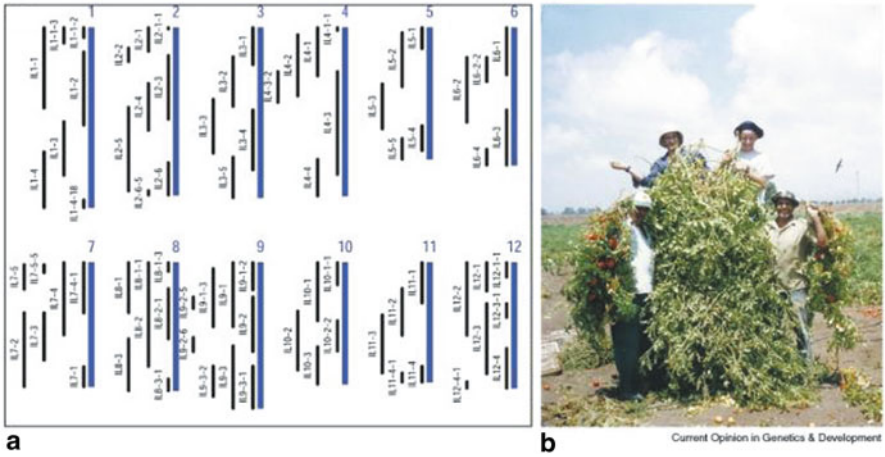

Fig. 4.1 The *Solanum pennellii* IL population. **a** Genome introgressions on the 12 tomato chromosomes of the 76 *S. pennellii* ILs, which are nearly isogenic to each other and differ only for the marked introgressed chromosome segments. **b** Heterosis for plant biomass in the F_1 hybrid of *S. pennellii* × *S. lycopersicum* (the middle plant) compared to the recurrent parent, M82 (far left and right plants). *S. pennellii*, while self-compatible in its native arid environment, does not set fruit in agricultural field conditions; however, it contributes QTL that significantly improve yield and other traits. The homozygous ILs show primarily lower yield than both parents owing to sterility, whereas certain IL hybrids show heterosis and increased yield. Interestingly, in many instances of crossing two ILs with similar QTL effects, double IL heterozygotes show lower magnitude than the sum of the effects of single heterozygotes, reflecting non-additivity of canalized phenotypes (Eshed and Zamir 1996). (Reproduced with permission from Lippman et al. (2007) Curr Opin Genet Dev 17:545, Fig. 1)

the dark orange fruits of the two lines IL6–3 and IL12–2 which are regulated by the dominant genes, *Beta* and *Delta*, respectively. The map based-cloning of both genes and their analysis indicated that the primary mechanism underlying aberrant carotenoid accumulation, and likely other transgressive phenotypes, is novel epistatic transcriptional regulation of *S. pennellii* genes (Ronen et al. 1999, 2000).

Recently Shivaprasad et al. (2012) have investigated the possibility that stable transgressive phenotypes observed in the *S. pennellii* LA0716 IL library are associated with micro or small interfering(si)RNAs. The rational for their study was based on the observation that primary sRNAs from one parent could initiate secondary siRNA on a target RNA from the other parent through an RNA-based mechanism. Such interactions would establish patterns of gene expression at either the transcriptional or posttranscriptional level that would be specific to the hybrids, and the effect would persist in lines that inherit both interacting loci. To verify their hypothesis, the authors have used high-throughput sequencing to characterize sRNAs in young seedlings of four *S. pennellii* ILs, as well as of the two parental lines, the F_1 and F_2 hybrids. They identified loci from which these sRNAs were more abundant in hybrids than in either parent and they showed that accumulation of such transgressive sRNAs correlated with suppression of the corresponding target genes. In one

case this effect was associated with hypermethylation of the corresponding genomic DNA. The results suggest that different sRNA-based mechanisms could be involved in transgressive segregation, and that the transgressive accumulation of miRNA and siRNAs is a manifestation of the hidden potential of parents that is released when hybrids are made.

The *S. pennellii* ILs have also been used to explore the underlying genetic mechanisms of heterosis, or hybrid vigor—the phenotypic superiority of a hybrid over its parents with respect to traits such as growth rate, reproductive success and yield. The genetic basis of this major genetic force that contributes to world food production is not clear yet. Possible genetic explanations include non-mutually exclusive mechanisms: dominance, true overdominance (ODO), pseudo-ODO (i.e. nearby loci at which alleles having dominant or partially dominant advantageous effects are in repulsion linkage phase) and certain types of epistasis (Lippman and Zamir 2007). For the genetic dissection of heterosis, exotic libraries have the double advantage of allowing the assessment of the contribution of ODO effects to heterosis while excluding epistasis, and to provide maximal genotypic and phenotypic diversity, which facilitates the evaluation for a broad range of phenotypes. In this respect, a phenomics study conducted on the *S. pennellii* LA0716 exotic library provided indirect support for true-ODO QTL, since ODO QTL were identified almost exclusively for the reproductive traits, while dominant and recessive QTL were detected for all analyzed traits (Semel et al. 2006). Other attempts to map ODO loci have been conducted in Arabidopsis (Meyer et al. 2010) and maize (Tang et al. 2010; Pea et al. 2009). These studies, along with the identification by Krieger et al. (2010) of a mutation in tomato with an ODO effect on yield, support the contribution of intragenic interactions to heterosis (Fridman and Zamir 2012).

Many ODO effects were confirmed over several years and environments, and a pyramiding approach was used to develop a multiple-introgression line carrying three independent *S. pennellii* yield-promoting genomic regions that had showed reproducible heterotic effect on fruit yield under irrigated and drought conditions (Gur and Zamir 2004). The pyramiding of these heterotic introgressions further increased yield beyond the individual QTL, although in a less-than-additive manner. The resulting hybrids had yields 50 % higher than leading commercial varieties when tested in multiple environments and irrigation regimes. The introduction of the *S. pennellii* introgressions into processing tomato lines resulted in the development of a leading hybrid variety, AB2 (Lippman et al. 2007).

The *S. pennellii* ILs have been a very effective tool also for the map-based cloning of the genes underlying QTL. The first QTL cloned have been *fw2.2* (fruit weight) (Frary et al. 2000; Cong et al. 2002) and *Brix9–2-5* (sugar yield, or Brix) (Fridman et al. 2000, 2004). While subtle changes in transcript quantity and in the timing of gene expression were correlated with natural variation at *fw2.2*; altered enzyme activity as a result of amino-acid substitutions in the gene was the cause for the variation between the cultivated and wild-species alleles at *Brix9-2-5*. These studies demonstrated that IL-based Mendelian segregation is a very efficient way to partition continuous variation for complex traits into discrete molecular components. Furthermore, these QTL were the first among many showing that, similarly to the variation

found for numerous genes that control quality traits, variation in QTL alleles in plants can be identified in both coding and regulatory regions of single genes (Salvi and Tuberosa 2005; Lippman et al. 2007). Besides *fw2.2* and *Brix9-2-5*, other tomato QTL have been cloned using segregating populations derived from *S. pennellii* ILs, such as *ovate*, *style2.1* and *fas* (Liu et al. 2002; Chen et al. 2007; Cong et al. 2008). An attempt was also made to clone *sun* using the ILs. However, *sun* mapped inside a paracentric inversion within the *S. pennellii* genome; this prevented map-based cloning using that resource (van der Knaap et al. 2004).

Although an extremely powerful and unbiased approach, delimiting a QTL to a single gene using genetic approaches is a time-consuming and technically demanding process (Fridman et al. 2000, 2004; Chen et al. 2007). As a consequence, while much progress has been made in mapping QTL, the elucidation of the underlying molecular mechanisms has lagged behind. Over the years, to try to accelerate the rate of QTL discovery, alternative strategies aimed at identifying candidate genes have been proposed and tested. The complexity of the approaches has evolved along with the availability of more advanced 'omics' tools. In this respect, the ILs represent a very efficient genetic resource to increase the efficiency in candidate gene identification and cloning of target QTL based on convergence of evidence deriving from QTL position, expression profiling data, functional and molecular diversity analyses of candidate genes (Li et al. 2005).

In tomato, the *S. pennellii* IL population has been used to explore the potential of the 'candidate gene approach' to identify candidate genes for QTLs influencing the intensity of tomato fruit color (Liu et al. 2003), tomato fruit size and composition (Causse et al. 2004), as well as fruit AsA content (Stevens et al. 2008), and vitamin E (Almeida et al 2011). The approach attempts to link, through mapping analysis sequences that have a known functional role in the measured phenotype with QTL that are responsible for the studied variation. While no co-locations were initially found between candidate genes and fruit color QTL (Liu et al. 2003), several apparent links were observed in the other three studies. More integrated strategies have also been tested in the *S. pennellii* ILs to find associations between trascriptomic changes and phenotypes of interest including fruit composition (Baxter et al. 2005; Di Matteo et al. 2010, 2013) and drought tolerance (Gong et al. 2010). A systems-based approach was used by Lee et al. (2012) to identify key genes regulating tomato fruit ripening (see Sect. 4.3.2).

Recently, Morgan et al. (2013) have demonstrated that individual ILs can provide useful information to guide metabolic engineering strategies. In fact, in spite of the relatively large regions of introgressed DNA from the genetically distinct donor parent contained in each IL, a detailed biochemical analysis allows pinpointing the main factor of metabolic disturbance and to identify potential candidate proteins that can subsequently be tested in a targeted manner in transgenic plants. In the specific case, one IL (IL2-5) known to have increased levels of fruit citrate and malate at the breaker stage, allowed to focus specifically on aconitase amongst a myriad of possible targets for manipulation of accumulation of carboxylic acids in tomato fruit (Morgan et al. 2013).

4.3.2 IL-Based System Analyses of Integrated Developmental Networks

Natural genetic variation stored in IL populations also facilitates the integration of multiple 'omics' techniques allowing multifaceted systems-level analysis of integrated developmental networks, and the identification of candidate genes underlying complex traits (Li et al. 2005; Schauer et al. 2006, 2008; Lippman et al. 2007; Hansen et al. 2008). These approaches can help identify uncharacterized networks or pathways, in addition to candidate regulators of such pathways (Saito and Matsuda 2010). The availability of a full-genome sequence can further facilitate filtering through genes in the QTL interval, since the examination of the annotation can often suggest a more likely candidate.

In tomato, numerous studies have already demonstrated the effectiveness of these approaches. With the aim of deciphering the genetic basis of compositional quality in tomato fruit, the high-diversity *S. pennellii* IL population was phenotyped for a wide range of plant morphology traits and for fruit pericarp 'primary' metabolites (Schauer et al. 2006). An integrated cartographical network based on correlation analysis of these diverse phenotypes allowed for the identification of morphology-dependent and morphology-independent links among a large number of QTL for fruit metabolism and yield. Moreover, the analysis revealed that harvest index (Fig. 4.2), which is a measure of the efficiency in partitioning of assimilated photosynthate to harvestable product (source-sink partitioning), was the chief pleiotropic hub in the combined network of metabolic and whole-plant phenotypic traits. These results suggest that plant structure has an important role in the final metabolite composition of the fruit. However, the strong negative association between metabolite content and yield was not found in lines heterozygous for the *S. pennellii* introgressions (ILHs) (Schauer et al. 2008). The uncoupling of the metabolic and morphological traits observed in the ILHs was explained with the reduced fertility problems and range of fruit sizes displayed by the heterozygous lines compared to the homozygous counterparts.

More recently, the *S. pennellii* IL library was used to gain insights into the genetic basis regulating natural variability in seed 'primary' metabolism and to unfold inter-organ correlations (Toubiana et al. 2012). The seed metabolite profiles were integrated with data from previous metabolic profiling studies on fruit pericarp together with plant morphological traits and yield-related parameters (Schauer et al. 2006; Lippman et al. 2007). Metabolite QTL mapping and correlation-based metabolic network analysis of the integrated heterogeneous data matrices allowed a comparison of the seed and the fruit metabolic networks. The graphic outcome and network parameters showed that the seed metabolite network displayed stronger interdependence of metabolic processes than the fruit, emphasizing the centrality of a tightly inter-regulated amino acid module in the seed metabolic network. Differently from the seed network, the fruit network was characterized by a rigid sugar module, and by the absence of a fatty acid module. In addition, the analysis allowed the identification of a number of candidate genes that may be useful to improve the nutritional values of seeds.

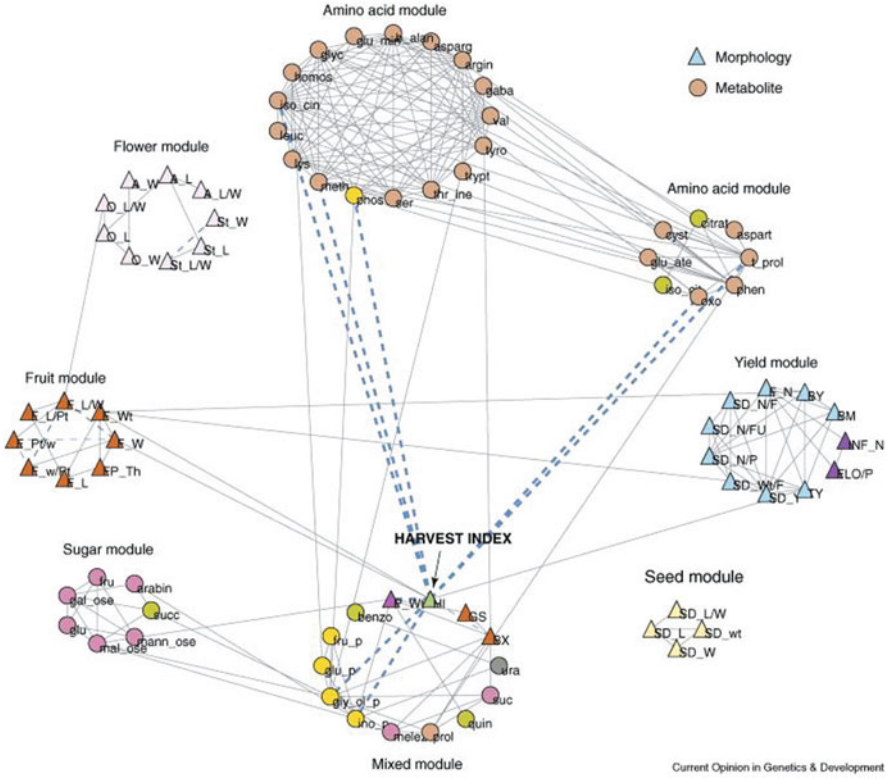

Fig. 4.2 A system view of IL-born morphology and metabolism interplay. Cartographic representation of the combined metabolic and morphological network of the tomato ILs (Schauer et al. 2006). Each trait (node) is represented by a shape (metabolites by circles and the phenotypes by triangles). The metabolites are color-coded according to type: brown, amino acids; pink, sugars; green, organic acids; yellow, phosphates; grey, miscellaneous, and module names are defined according to the most prevalent trait type. A line connecting two traits represents a significant correlation between them. Correlation of all trait pairs was calculated using IL means (total of 76 lines); gray lines represent positive correlations, blue lines represent negative correlations (significance threshold of $p < 0.0001$). Harvest index (HI), the ratio of fruit yield to total plant mass (plant weight + fruit yield), is the central pleiotropic hub of the network. (Reproduced with permission from Lippman et al. (2007) Curr Opin Genet Dev 17:545, Fig. 3)

Besides 'primary' metabolism, existing genetic variation stored in exotic libraries represents a very powerful tool also for the analysis of specialized (traditionally called 'secondary') metabolism (Schilmiller et al. 2012). For instance, glandular trichomes of cultivated tomato and wild tomato relatives produce a variety of structurally diverse volatile and non-volatile specialized metabolites, including terpenes, flavonoids and acyl sugars (Schilmiller et al. 2012). A genetic screen of leaf trichome and surface metabolite extracts of the *S. pennellii* LA0716 IL population allowed the

identification of genomic regions of the wild parent influencing mono- and sesquiterpenes or only sesquiterpenes, and the quality or quantity of acyl sugars metabolites (Schilmiller et al. 2010). In addition, the *Solanum* ILs have also been profiled for accumulation of volatile fruit compounds, allowing the identification of 25 genetic regions from *S. pennellii* LA0716 that increased emissions of at least one of the 23 volatiles measured (Tieman et al. 2006; Mathieu et al. 2009). The ability to measure the influence of many regions of the genome on multiple metabolites provided important insights into the metabolic networks. Discovery of loci that influence emissions of multiple volatile compounds led to the hypothesis that these metabolites are biosynthetically related or regulated by a common regulatory network.

Finally, Lee et al. (2012) applied ripe fruit transcriptional and metabolic profiling to the *S. pennellii* LA0716 exotic library. Candidate genes mining based on correlation analyses allowed the identification of the ethylene response factor *SlERF6*. RNAi analysis showed that *SlERF6* plays a central role in tomato ripening integrating the ethylene and carotenoid synthesis pathways.

Together, these examples illustrate that with the continued development of genetic and "omics" tools, more detailed systems-level analyses will be possible, increasing the efficiency in discovery, candidate gene identification and cloning of target QTL.

4.3.3 Other Tomato Library Resources

In order to enhance the rate of progress of introgression breeding, Zamir (2001) proposed to invest in the development of a genetic infrastructure of "exotic libraries". Along this line, for tomato, besides the *S. pennellii* LA0716 exotic library, populations of ILs and/or pre-ILs have been developed and/or further refined from other wild relatives including *S. habrochaites* (acc. LA1777) (Monforte and Tanksley 2000a; Tripodi et al. 2010; S. Grandillo, personal communication), *S. habrochaites* (acc. LA0407) (Finkers et al. 2007), *S. chmielewskii* (acc. LA1840) (Peleman and Van der Voort 2003; Prudent et al. 2009), *S. neorickii* (acc. LA2133) (Fulton et al. 2000; D. Zamir, personal communication), *S. pimpinellifolium* (acc. LA1589) (Doganlar et al. 2002; D. Zamir, personal communication), *S. pimpinellifolium* (acc. TO-937) (W. Barrantes and A.J. Monforte, personal communication) and the wild tomato-like nightshade *S. lycopersicoides* LA2951 (Chetelat and Meglic 2000; Canady et al. 2005) (Table 4.1).

The first set of *S. habrochaites* LA1777 ILs and pre-ILs was developed by Monforte and Tanksley (2000a) from the AB-QTL population (Bernacchi et al. 1998), and consisted of 99 ILs and BCRILs, in the cv. E6203 genetic background, providing an estimated coverage of approximately 85 % of the wild donor genome. The lines are highly variable for numerous traits including yield, leaf morphology and trichome density, cold tolerance, as well as fruit traits such as shape, size, color, biochemical composition and flavor volatiles. Favorable wild QTL alleles have been identified for several of the evaluated traits (Table 4.1) (Monforte and Tanksley 2000b; Van der Hoeven et al. 2000; Monforte et al. 2001; Yates et al. 2004; Dal Cin et al.

2009; Mathieu et al. 2009; Liu et al. 2012; Grandillo et al. 2011, 2013). Nevertheless, several lines of this initial population still contain multiple wild species chromosome segments. Therefore, within the framework of the EU project (EU-SOL; http://www.eu-sol.net/), an improved collection of *S. habrochaites* LA1777 ILs was developed and anchored to a shared framework of ~ 120 conserved ortholog set II (COSII) markers (Wu et al. 2006). This new population of LA1777 ILs allows a better genome coverage based on single-introgression lines (Tripodi et al. 2010; S. Grandillo, personal communication). Furthermore, leaf and fruit pericarp RNA-seq SNP data collected on this new panel of LA1777 ILs provided a better definition of the introgression boundaries and their anchoring to the tomato genome sequence (S. Grandillo and J. Giovannoni, personal communication; The Tomato Genome Consortium 2012).

From the tomato AB-QTL populations MAS has been used to develop populations of ILs and pre-ILs in the genetic background of the processing cv. E6203 also for *S. pimpinellifolium* LA1589 (196 BILs) (Grandillo and Tanksley 1996; Tanksley et al. 1996; Bernacchi et al. 1998; Doganlar et al. 2002; D. Zamir, personal communication) and *S. neorickii* LA2133 (142 BILs) (Fulton et al. 2000; Zamir, personal communication). Within the framework of the EU-SOL project the 142 *S. neorickii* BILs have been evaluated for agronomic traits, including yield, brix and fruit weight, and several favorable wild alleles were identified that could be targeted for further marker-assisted introgression into cultivated tomato (D. Zamir and S. Grandillo, personal communication).

Another population of 55 *S. chmielewskii* LA1840 ILs in the genetic background of the cv. Moneyberg was developed by KeyGene N.V. (Peleman and van der Voort 2003). A subset of these lines was used to study the effect of fruit load, and therefore of carbon availabity, on the detection of QTL underlying fruit weight and composition (Prudent et al. 2009; Do et al. 2010), and on age- and genotype- dependent gene expression (Prudent et al. 2010). A model-based approach followed by genetic analysis allowed uncoupling genetic from physiological relationships among processes, and thus provided new insights towards understanding tomato fruit sugar assimilation (Prudent et al. 2011). Furthermore, phenotypic analysis of the *S. chmielewskii* LA1840 IL population revealed three overlapping ILs on chromosome 1 with a pink fruit color, a trait known to be regulated by the *Y* locus (Ballester et al. 2010). Biochemical and molecular data, along with gene mapping, segregation analysis and virus-induced gene silencing experiments allowed the identification of *SlMYB12* as a likely candidate for the *Y* locus (Ballester et al. 2010).

In order to facilitate marker-assisted breeding based on these wild species resources, and to facilitate comparisons between function maps of tomato and potato, some of the IL libraries described above have been anchored to the potato genome using a common set of ~ 120 COSII markers and (Tripodi et al. 2010; S. Grandillo, personal communication). The multi-species IL platform include ILs and BILs derived from interspecific crosses of tomato and the five wild accessions *S. pennellii* LA0716, *S. habrochaites* LA1777, *S. neorickii* LA2133, *S. chmielewskii* LA1840, and *S. pimpinellifolium* LA1589 (Tripodi et al. 2010; Brog et al. 2011). Multi-species IL platforms are highly divergent in phenotypes

providing abundant segregation for whole genome naturally selected variation affecting yield, morphological and biochemical traits, and allow multiallelic effects to be captured. A draft sequence of *S. pimpinellifolium* LA1589 is already available (Tomato Genome Consortium 2012), and within the SOL-100 sequencing project (http://solgenomics.net/organism/sol100/view), sequences are becoming available for most of the parents of the tomato IL libraries described above, which will further enhance the value of these genetic resources.

4.4 Integrative Approaches to Genomic Introgression Mapping

Genetically well-characterized IL populations, anchored to highly saturated genetic maps, are key tools for rapid and precise localization of QTL and subsequent identification of the casual genes. Hitherto, the mapping of IL introgression boundaries has relied on a wide range of electrophoresis-based molecular tools, which have rarely ensured sufficiently high marker saturations (Severin et al. 2010). The development of new high-throughput molecular platforms that allow automated genotyping is accelerating and making more precise the process of introgressions mapping and IL library development. Dense genetic maps, in fact, allow for localizing the introgressed segments with high resolution, which is crucial for the selection of ILs carrying small marker-defined segments for genome-wide coverage of the donor parent genome.

A few studies have compared the efficiency of different genotyping platforms for genome introgression mapping. For instance, in rice, an IL population consisting of 128 ILs and pre-ILs derived from a cross between two sequenced rice cultivars, was genotyped with 254 PCR-based markers and then subjected to whole-genome re-sequencing (Xu et al. 2010). The high-quality physical map of ultrahigh-density SNPs identified 117 new segments (almost all shorter than 3 Mb) that had not been detected in the molecular marker map. The new method improved the resolution of recombination breakpoints 236-fold, and almost eliminated the likelihood of missing double-crossovers in the mapping population. Furthermore, the sequencing-based physical map allowed QTL bin mapping with higher accuracy, thus being of great potential value for gene discovery and genetic mapping.

Another study was recently conducted to compare some of the existing (Affymetrix SFP and Illumina GoldenGate) and emerging (Illumina NGS) technologies for soybean introgression mapping (Severin et al. 2010). The results show that SFP, Illumina GoldenGate, and RNA-Seq are complementary methods for identifying genetic introgressions in NILs. RNA-Seq methodologies clearly identified a much greater number of polymorphic loci within the known introgression sites, and the increased marker coverage allowed to identify the introgression boundaries at a higher resolution. Comparative NGS analyses of NILs with their respective parental lines offer the additional advantage of identifying SNP polymorphisms that are specific to the genetic material of interest. The SNPs identified de novo by RNA-Seq

can be directly used for fine-mapping on subsequent generations by means of custom SNP genotyping platforms. Furthermore, the RNA-Seq data may be mined for transcriptional differences or genetic alterations that may identify candidate genes that drive the differential phenotypes observed between the lines. In this respect, compared to the Affymetrix platform, the RNA-Seq data provide a larger sampling of transcripts and also permit the possible identification of frame-shift or nonsense mutations within introgressed loci. However, at the moment, besides the higher costs, the RNA-Seq approach has also the disadvantage of a marker depth necessarily biased for gene-rich regions and therefore, even applying bootstrapping methods to correct for gene densities, severely gene-poor regions might not be represented in the analyses (Severin et al. 2010).

High-throughput genotyping platforms have been used also on interspecific IL populations of crops including barley (Sato and Takeda 2009; Schmalenbach et al. 2011), tomato (Sim et al. 2012; Van Schalkwyk et al. 2012) and rice (Ali et al. 2010). For example, in barley, an Illumina 1536-SNP array was used for high-resolution genotyping of a set of 73 ILs (S42ILs) originating from a cross between the spring barley cv. Scarlet (*Hordeum vulgare* ssp. *vulgare*) and the wild barley accession IDSR42–8 (*H. v*. ssp. *spontaneum*) (Schmalenbach et al. 2011). The array enabled a precise localization of the wild barley introgressions in the elite barley background. In addition, to further implement this IL library into a resource for rapid identification, fine-mapping and positional cloning of QTL, segregating high-resolution mapping populations (S42IL-HRs) were developed for most ILs.

In tomato, the high-density "SolCAP" SNP array was used to genotype a large collection of tomato accessions, as well as the *S. pennellii* LA0716 ILs (Sim et al. 2012). In addition, Van Schalkwyk et al. (2012) reported the development of a diversity arrays technology (DArT) platform consisting of 6,912 clones from domesticated tomato and 12 wild tomato/Solanaceous species. The platform was validated by bin-mapping 990 polymorphic DArT markers together with 108 RFLP markers across the *S. pennellii* LA0716 IL library, resulting, on average, in a ten-fold increase of the number of markers available for each IL. A subset of DArT markers from ILs previously associated with increased levels of lycopene and carotene were sequenced, and 44 % matched protein coding genes. The conversion of the DArT markers to CAPS or SNP markers should facilitate fine mapping of QTLs in *S. pennellii* ILs.

In rice, about two dozen IL/BIL libraries have been developed representing different *O. sativa* backgrounds and wild donors, and most of the donors and recipient parents have been sequenced using second-generation sequencing technology and/or genotyped using the 44,100 SNP array (Table 4.1) (Ali et al. 2010). In addition, physical maps of 17 *Oryza* species (representing the 10 genome types) have been developed by the *Oryza* Map Alignment project (Ali et al. 2010; http://www.omap.org).

It is clear that high-throughput SNP assays and the availability of custom-designed medium- and low-density SNP arrays will greatly enhance the efficiency of whole-genome IL library development, allowing the selection of small marker-defined segments introgressed from the unadapted germplasm. Furthermore, the availability of SNP markers across the introgressed donor regions will facilitate fine-mapping and cloning of genes underlying target QTL.

4.5 Conclusions

Many crops have a very narrow genetic base that threatens future genetic gains. In contrast, wild species represent a rich, although mostly untapped, reservoir of valuable alleles that could be used to address present and future breeding challenges. For a more efficient exploitation of exotic germplasm, we need to capitalize on the acquired knowledge and on the ever-growing genetic and "omics" resources that are becoming available and that take advantage of many recently released crop genome sequences to investigate gene-function (Hamilton and Buell 2012, http://genomevolution.org/wiki/index.php/Sequenced_plant_genomes). Among all model systems, the wild and domesticated species of the tomato clade have pioneered novel population development, such as IL populations or "exotic libraries" (Zamir 2001; Lippman et al. 2007). The last 20 years of research conducted on the *S. pennellii* LA0716 ILs (the founding population) have clearly demonstrated the power of these congenic and permanent resources for the genetic and molecular analyses of QTL, for dissecting heterosis, and hence for the development of a leading hybrid variety. Over the years, the IL approach has been integrated with various state-of-the art 'omics' platforms, thus evolving beyond standard QTL identification towards a multifaceted systems-level analysis. These achievements have encouraged the research community to invest in the development of IL library resources, or related prebreds, such as BILs, representing different fractions of the exotic parent genomes, for a number of other tomato wild species, as well as for a wide range of crops. The results indicate that exotic germplsam stores a tremendous wealth of potentially valuable alleles, many of which would not have been predicted from the phenotypes of the wild plants. However, only a small fraction of the naturally occurring genetic diversity available in the world's genebanks has been explored to date, and made permanently accessible through IL population development. The advent of new cost-effective, high-throughput genotyping and sequencing technologies is expected to change this trend. Strategies based on phylogenetic approaches can be pursued to select the right parents that would maximize the probability of creating new useful transgressive segregation from which to select superior phenotypes (McCouch et al. 2012). In addition, the new high-throughput molecular platforms are accelerating and making more precise the process of introgression mapping and IL library development, and the availability of SNP markers across the introgressed donor regions will facilitate fine mapping and cloning of genes underlying target QTL.

In this context of fast-evolving technological advances, the availability of exotic libraries further increases the value of the numerous unadapted genetic resources stored worldwide in our *in situ* and *ex situ* germplasm collections.

Acknowledgments The author thanks all the colleagues who provided unpublished information and apologizes to those authors whose work could not be quoted due to space limitations. Research in the laboratory of S. Grandillo was supported in part by the EUSOL project PL 016214–2, by the Italian the Italian Ministry of University and Research (MIUR) project GenoPOM, by a dedicated grant from the Italian Ministry of Economy and Finance to the National Research Council for

the project "Innovazione e Sviluppo del Mezzogiorno—Conoscenze Integrate per Sostenibilità ed Innovazione del Made in Italy Agroalimentare—Legge n. 191/2009", and by the PON R&C 2007–2013 grant financed by the Italian MIUR in cooperation with the European Funds for the Regional Development (FESR).

References

Ali ML, Sanchez PL, Yu S-B et al (2010) Chromosome segment substitution lines: a powerful tool for the introgression of valuable genes from *Oryza* wild species into cultivated rice (*O. sativa*). Rice 3:218–234

Almeida J, Quadrana L, Asís R et al (2011) Genetic dissection of vitamin E biosynthesis in tomato. J Exp Bot 62:3781–3798

Alpert K, Grandillo S, Tanksley SD (1995) fw2.2: a major QTL controlling fruit weight is common to both red- and green-fruited tomato species. Theor Appl Genet 91:994–1000

Alpert K, Tanksley S (1996) High-resolution mapping and isolation of a yeast artificial chromosome contig containing fw2.2: a major fruit weight quantitative trait locus in tomato. Proc Natl Acad Sci U S A 93:15503–15507

Arikita FN, Azevedo MS, Scotton DC et al (2013) Novel natural genetic variation controlling the competence to form adventitious roots and shoots from the tomato wild relative *Solanum pennellii*. Plant Sci 199-200:121–130

Ashikari M, Matsuoka M (2006) Identification, isolation and pyramiding of quantitative trait loci for rice breeding. Trends Plant Sci 11:344–350

Atri C, Kumar B, Kumar H et al (2012) Development and characterization of Brassica juncea-fruticulosa introgression lines exhibiting resistance to mustard aphid (*Lipaphis erysimi* Kalt). BMC Genet 13:104

Ballester AR, Molthoff J, de Vos R et al (2010) Biochemical and molecular analysis of pink tomatoes: deregulated expression of the gene encoding transcription factor SlMYB12 leads to pink tomato fruit color. Plant Phys 152:71–84

Baxter CJ, Sabar M, Quick WP, Sweetlove LJ (2005) Comparison of changes in fruit gene expression in tomato introgression lines provides evidence of genome-wide transcriptional changes and reveals links to mapped QTLs and described traits. J Exp Bot 56:1591–1604

Bedinger PA, Chetelat RT, McClure B et al (2011) Interspecific reproductive barriers in the tomato clade: opportunities to decipher mechanisms of reproductive isolation. Sex Plant Reprod 24:171–187

Bernacchi D, Beck-Bunn T, Emmatty D et al (1998) Advanced backcross QTL analysis of tomato. II. Evaluation of near-isogenic lines carrying single-donor introgressions for desirable wild QTL-alleles derived from *Lycopersicon hirsutum* and *L. pimpinellifolium*. Theor Appl Genet 97:170-180 and 1191–1196

Bessey CE (1906) Crop improvement by utilizing wild species. Am Breed Assoc II:112–118

Blair MW, Izquierdo P (2012) Use of the advanced backcross-QTL method to transfer seed mineral accumulation nutrition traits from wild to Andean cultivated common beans. Theor Appl Genet 125:1015–1031

Bleeker PM, Diergaarde PJ, Ament K et al (2009) The role of specific tomato volatiles in tomato-whitefly interaction. Plant Physiol 151:925–935

Brog M, Tripodi P, Cammareri M et al (2011) Towards phenomics of the sequenced genomes of the cultivated tomato and its wild ancestor *Solanum pimpinellifolium*. In: Proceedings of the Joint Meeting AGI-SIBV-SIGA Assisi, Italy, 19-22 September 2011. ISBN 978-88-904570-2–9

Buerstmayr M, Lemmens M, Steiner B, Buerstmayr H (2011) Advanced backcross QTL mapping of resistance to Fusarium head blight and plant morphological traits in a *Triticum macha* × *T. aestivum* population. Theor Appl Genet 123:293–306

Canady MA, Meglic V, Chetelat RT (2005) A library of *Solanum lycopersicoides* introgression lines in cultivated tomato. Genome 48:685–697

Causse M, Chaïb J, Lecomte L et al (2007) Both additivity and epistasis control the genetic variation for fruit quality traits in tomato. Theor Appl Genet 115:429–442

Causse M, Duffe P, Gomez MC et al (2004) A genetic map of candidate genes and QTLs involved in tomato fruit size and composition. J Exp Bot 55:1671–1685

Cavanagh C, Morell M, Mackay I, Powell W (2008) From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. Curr Opin Plant Biol 11:215–221

Chapman NH, Bonnet J, Grivet L et al (2012) High-resolution mapping of a fruit firmness-related quantitative trait locus in tomato reveals epistatic interactions associated with a complex combinatorial locus. Plant Physiol 159:1644–1657

Chen KY, Cong B, Wing R et al (2007) Changes in regulation of a transcription factor lead to autogamy in cultivated tomatoes. Science 318:643–645

Chen KY, Tanksley SD (2004) High-resolution mapping and functional analysis of se2.1: a major stigma exsertion quantitative trait locus associated with the evolution from allogamy to autogamy in the genus *Lycopersicon*. Genetics 168:1563–1573

Chetelat RT, Meglic V (2000) Molecular mapping of chromosome segments introgressed from *Solanum lycopersicoides* into cultivated tomato (*Lycopersicon esculentum*). Theor Appl Genet 100:232–241

Cong B, Barrero LS, Tanksley SD (2008) Regulatory change in YABBY-like transcription factor led to evolution of extreme fruit size during tomato domestication. Nat Genet 40:800–804

Cong B, Liu J, Tanksley SD (2002) Natural alleles at a tomato fruit size quantitative trait locus differ by heterochronic regulatory mutations. Proc Natl Acad Sci U S A 99:13606–13611

Dal Cin V, Kevany B, Fei Z, Klee HJ (2009) Identification of *Solanum habrochaites* loci that quantitatively influence tomato fruit ripening-associated ethylene emissions. Theor Appl Genet 119:1183–1192

Desjardins CA, Gadau J, Lopez JA et al (2013) Fine-scale mapping of the *Nasonia* genome to chromosomes using a high-density genotyping microarray. G3 (Bethesda) 3:205–215

de Vicente MC, Tanksley SD (1993) QTL analysis of transgressive segregation in an interspecific tomato cross. Genetics 134:585–596

Di Matteo A, Ruggieri V, Sacco A et al (2013) Identification of candidate genes for phenolics accumulation in tomato fruit. Plant Sci 205-206:87–96

Di Matteo A, Sacco A, Anacleria M et al (2010) The ascorbic acid content of tomato fruits is associated with the expression of genes involved in pectin degradation. BMC Plant Biol 10:163

Do PT, Prudent M, Sulpice R et al (2010) The influence of fruit load on the tomato pericarp metabolome in a *Solanum chmielewskii* introgression line population. Plant Physiol 154:1128–1142

Doganlar S, Frary A, Ku H-M, Tanksley SD (2002) Mapping quantitative trait loci in inbred backcross lines of *Lycopersicon pimpinellifolium* (LA1589). Genome 45:1189–1202

Doi K, Iwata N, Yoshimura A (1997) The construction of chromosome substitution introgression lines of African rice (*Oryza glaberrima* Steud.) in the background of japonica (*O. sativa* L.). Rice Genet Newslett 14:39–41

Doroszuk A, Snoek LB, Fradin E et al (2009) A genome-wide library of CB4856/N2 introgression lines of *Caenorhabditis elegans*. Nucleic Acids Res 37:e110

Eduardo I, Arús P, Monforte AJ (2005) Development of a genomic library of near isogenic lines (NILs) in melon (*Cucumis melo* L.) from the exotic accession PI161375. Theor Appl Genet 112:139–148

Eduardo I, Arús P, Monforte AJ et al (2007) Estimating the genetic architecture of fruit quality traits in melon using a genomic library of near isogenic lines. J Am Soc Hortic Sci 132:80–89

Eshed Y, Gera G, Zamir D (1996) A genome-wide search for wild-species alleles that increase horticultural yield of processing tomatoes. Theor Appl Genet 93:877–886

Eshed Y, Zamir D (1994) Introgressions from *Lycopersicon pennellii* can improve the soluble solids yield of tomato hybrids. Theor Appl Genet 88:891–897

Eshed Y, Zamir D (1995) An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. Genetics 141:1147–1162

Eshed Y, Zamir D (1996) Less-than-additive epistatic interactions of quantitative trait loci in tomato. Genetics 143:1807–1817

Falke KC, Frisch M (2011) Power and false-positive rate in QTL detection with near-isogenic line libraries. Heredity (Edinb) 106:576–584

Falke KC, Miedaner T, Frisch M (2009b) Selection strategies for the development of rye introgression libraries. Theor Appl Genet 119:595–603

Falke KC, Susić Z, Hackauf B et al (2008) Establishment of introgression libraries in hybrid rye (*Secale cereale* L.) from an Iranian primitive accession as a new tool for rye breeding and genomics. Theor Appl Genet 117:641–652

Falke KC, Susić Z, Wilde P et al (2009a) Testcross performance of rye introgression lines developed by marker-assisted backcrossing using an Iranian accession as donor. Theor Appl Genet 118:1225–1238

Falke KC, Wilde P, Miedaner T (2009c) Rye introgression lines as source of alleles for pollen-fertility restoration in Pampa CMS. Plant Breeding 128:528–531

Fang S, Yukilevich R, Chen Y et al (2012) Incompatibility and competitive exclusion of genomic segments between sibling *Drosophila* species. PLoS Genet 8:e1002795

Fernandez-Silva I, Moreno E, Essafi A et al (2010) Shaping melons: agronomic and genetic characterization of QTLs that modify melon fruit morphology. Theor Appl Genet 121:931–940

Finkers R, van Heusden AW, Meijer-Dekens F et al (2007) The construction of a *Solanum habrochaites* LYC4 introgression line population and the identification of QTLs for resistance to Botrytis cinerea. Theor Appl Genet 114:1071–1080

Foncéka D, Hodo-Abalo T, Rivallan R et al (2009) Genetic mapping of wild introgressions into cultivated peanut: a way toward enlarging the genetic basis of a recent allotetraploid. BMC Plant Biol 9:103

Foncéka D, Tossim HA, Rivallan R et al (2012) Construction of chromosome segment substitution lines in peanut (*Arachis hypogaea* L.) using a wild synthetic and QTL mapping for plant morphology. PLoS One 7:e48642

Frary A, Doganlar S, Frampton A et al (2003) Fine mapping of quantitative trait loci for improved fruit characteristics from *Lycopersicon chmielewskii* chromosome 1. Genome 46:235–243

Frary A, Göl D, Keleş D et al (2010) Salt tolerance in *Solanum pennellii*: antioxidant response and related QTL. BMC Plant Biol 10:58

Frary A, Nesbitt TC, Frary A et al (2000) fw-2.2: a quantitative trait locus key to the evolution of tomato fruit size. Science 289:85–88

Fridman E, Carrari F, Liu YS et al (2004) Zooming in on a quantitative trait for tomato yield using interspecific introgressions. Science 305:1786–1789

Fridman E, Liu YS, Carmel-Goren L et al (2002) Two tightly linked QTLs modify tomato sugar content via different physiological pathways. Mol Genet Genomics 266:821–826

Fridman E, Pleban T, Zamir D (2000) A recombination hotspot delimits a wild-species quantitative trait locus for tomato sugar content to 484 bp within an invertase gene. Proc Natl Acad Sci U S A 97:4718–472

Fridman E, Zamir D (2012) Next-generation education in crop genetics. Curr Opin Plant Biol 15:218–223

Fulton TM, Grandillo S, Beck-Bunn T et al (2000) Advanced backcross QTL analysis of a *Lycopersicon esculentum* × *L. parviflorum* cross. Theor Appl Genet 100:1025–1042

Garcia GM, Stalker HT, Kochert G (1995) Introgression analysis of an interspecific hybrid population in peanuts (*Arachis hypogaea* L.) using RFLP and RAPD markers. Genome 38:166–176

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

Glaszmann JC, Kilian B, Upadhyaya HD, Varshney RK (2010) Accessing genetic diversity for crop improvement. Curr Opin Plant Biol 13:167–173
Godfray HC, Beddington JR, Crute IR et al (2010) Food security: the challenge of feeding 9 billion people. Science 327:812–818
Gong P, Zhang J, Li H et al (2010) Transcriptional profiles of drought-responsive genes in modulating transcription signal transduction, and biochemical pathways in tomato. J Exp Bot 61:3563–3575
Grandillo S, Chetelat R, Knapp S et al (2011) *Solanum* sect. *Lycopersicon*. In: Kole C (ed) Vegetables. Wild crop relatives: genomic and breeding resources, vol 5. Springer, Dordrecht, pp 129–215
Grandillo S, Tanksley SD (1996) QTL analysis of horticultural traits differentiating the cultivated tomato from the closely related species *Lycopersicon pimpinellifolium*. Theor Appl Genet 92:935–951
Grandillo S, Tanksley SD, Zamir D (2008) Exploitation of natural biodiversity through genomics. In: Varshney RK, Tuberosa R (eds) Genomics approaches and platforms. Genomics assisted crop improvement, vol 1. Springer, Dordrecht, pp 121–150
Grandillo S, Termolino P, van der Knaap E (2013) Molecular mapping of complex traits in tomato. In: Genetics, Genomics and Breeding of crop plants (Series Editor C. Kole) Volume: Genetics, Genomics and Breeding of Tomato. (Volume editors B.E. Liedl, J.A. Labate, A.J. Slade, J.R. Stommel, C. Kole). Science Publishers, Enfield, NH, USA, pp 150—227
Gu J, Yin X, Struik PC et al (2012) Using chromosome introgression lines to map quantitative trait loci for photosynthesis parameters in rice (*Oryza sativa* L.) leaves under drought and well-watered field conditions. J Exp Bot 63:455–469
Guo S, Wei Y, Li X et al (2012) Development and identification of introgression lines from cross of *Oryza sativa* and *Oryza minuta*. Rice Sci 20:95–102
Gur A, Osorio S, Fridman E et al (2010) hi2-1, a QTL which improves harvest index, earliness and alters metabolite accumulation of processing tomatoes. Theor Appl Genet 121:1587–1599
Gur A, Semel Y, Cahaner A, Zamir D (2004) Real time QTL of complex phenotypes in tomato interspecific introgression lines. Trends Plant Sci 9:107–109
Gur A, Semel Y, Osorio S et al (2011) Yield quantitative trait loci from wild tomato are predominately expressed by the shoot. Theor Appl Genet 122:405–420
Gur A, Zamir D (2004) Unused natural variation can lift yield barriers in plant breeding. PLoS Biol 2:e245
Gutiérrez AG, Carabalí SJ, Giraldo OX et al (2010) Identification of a Rice stripe necrosis virus resistance locus and yield component QTLs using *Oryza sativa* × *O. glaberrima* introgression lines. BMC Plant Biol 10:6
Hajjar R, Hodgkin T (2007) The use of wild relatives in crop improvement: a survey of developments over the last 20 years. Euphytica 156:1–13
Hamilton JP, Buell CR (2012) Advances in plant genome sequencing. Plant J 70(1):177–190
Hansen BG, Halkier BA, Kliebenstein DJ (2008) Identifying the molecular basis of QTLs: eQTLs add a new dimension. Trends Plant Sci 13:72–77
Hoffmann A, Maurer A, Pillen K (2012) Detection of nitrogen deficiency QTL in juvenile wild barley introgression lines growing in a hydroponic system. BMC Genet 13:88
Holtan HE, Hake S (2003) Quantitative trait locus analysis of leaf dissection in tomato using *Lycopersicon pennellii* segmental introgression lines. Genetics 165:1541–1550
Hori K, Sato K, Nankaku N, Takeda K (2005) QTL analysis in recombinant chromosome substitution lines and doubled haploid lines derived from a cross between *Hordeum vulgare* ssp. *vulgare* and *Hordeum vulgare* ssp. *spontaneum*. Mol Breed 16:295–311
Jeunken MJW, Lindhout P (2004) The development of lettuce backcross inbred lines (BILs) for exploitation of the *Lactuca saligna* (wild lettuce) germplasm. Theor Appl Genet 109:394–401
Johal GS, Balint-Kurti P, Weil CF (2008) Mining and harnessing natural variation: a little MAGIC. Crop science 48:2066–2072
Johnston PA, Timmerman-Vaughan GM, Farnden KJ, Pickering R (2009) Marker development and characterisation of *Hordeum bulbosum* introgression lines: a resource for barley improvement. Theor Appl Genet 118:1429–1437
Kamenetzky L, Asís R, Bassi S et al (2010) Genomic analysis of wild tomato introgressions determining metabolism- and yield-associated traits. Plant Physiol 152:1772–1786

Keurentjes JJB, Bentsink L, Alonso-Blanco C et al (2007) Development of a near-isogenic line population of *Arabidopsis thaliana* and comparison of mapping power with a recombinant inbred line population. Genetics 175:891–905

Koumproglou R, Wilkes TM, Towson P et al (2002). STAIRS: a new genetic resource for functional genomic studies of *Arabidopsis*. Plant J 31:355–364

Krieger U, Lippman ZB, Zamir D (2010) The flowering gene single flower truss drives heterosis for yield in tomato. Nat Genet 42:459–463

Kurakazu T, Sobrizal K, Ikeda K et al (2001) Oryza meridionalis chromosomal segment introgression lines in cultivated rice, *O. sativa* L. Rice Genet Newsl 18:81–82

Kuspira J, Unrau J (1957) Genetic analysis of certain characters in common wheat using all chromosome substitution lines. Can J Plant Sci 37:300–326

Lee JM, Joung JG, McQuinn R et al (2012) Combined transcriptome, genetic diversity and metabolite profiling in tomato fruit reveals that the ethylene response factor SlERF6 plays an important role in ripening and carotenoid accumulation. Plant J 70:191–204

L'Hôte D, Laissue P, Serres C et al (2010) Interspecific resources: a major tool for quantitative trait locus cloning and speciation research. Bioessays 32:132–142

Li Z-K, Fu B-Y, Gao Y-M et al (2005) Genome-wide introgression lines and their use in genetic and molecular dissection of complex phenotypes in rice (*Oryza sativa* L.). Plant Mol Biol 59:33–52

Lippman ZB, Semel Y, Zamir D (2007) An integrated view of quantitative trait variation using tomato interspecific introgression lines. Curr Opin Genet Dev 17:545–552

Lippman ZB, Zamir D (2007) Heterosis: revisiting the magic. Trends Genet 23:60–66

Liu H, Ouyang B, Zhang J et al (2012) Differential modulation of photosynthesis, signaling, and transcriptional regulation between tolerant and sensitive tomato genotypes under cold stress. PLoS One 7:e50785

Liu J, Van Eck J, Cong B, Tanksley SD (2002) A new class of regulatory genes underlying the cause of pear-shaped tomato fruit. Proc Natl Acad Sci U S A 99:13302–13306

Liu S, Zhou R, Dong Y et al (2006) Development, utilization of introgression lines using a synthetic wheat as donor. Theor Appl Genet 112:1360–1373

Liu Y-S, Gur A, Ronen G et al (2003) There is more to tomato fruit colour than candidate carotenoid genes. Plant Biotech J 1:195–207

Liu YS, Zamir D (1999) Second generation *L. pennellii* introgression lines and the concept of bin mapping. Tomato Genet Coop Rep 49:26–30

Mackay TFC (2001) The genetic architecture of quantitative traits. Ann Rev Genet 35:303–339

Mageroy MH, Tieman DM, Floystad A et al (2012) A Solanum lycopersicum catechol-O-methyltransferase involved in synthesis of the flavor molecule guaiacol. Plant J 69:1043–1051

Mahone GS, Frisch M, Miedaner T et al (2012) Identification of quantitative trait loci in rye introgression lines carrying multiple donor chromosome segments. Theor Appl Genet 126:49–58

March TJ, Richter D, Colby T et al (2012) Identification of proteins associated with malting quality in a subset of wild barley introgression lines. Proteomics 12:2843–2851

Mathieu S, Dal Cin V, Fei Z et al (2009) Flavour compounds in tomato fruits: identification of loci and potential pathways affecting volatile composition. J Exp Bot 60:325–337

Matus I, Corey A, Filchkin T et al (2003) Development and characterization of recombinant chromosome substitution lines (RCSLs) using *Hordeum vulgare* subsp. *spontaneum* as a source of donor alleles in a *Hordeum vulgare* subsp. *vulgare* background. Genome 46:1010–1023

McCouch S (2004) Diversifying selection in plant breeding. PLoS Biol 2:e347

McCouch SR, McNally KL, Wang W et al (2012) Genomics of gene banks: a case study in rice. Am J Bot 99:407–423

McCouch SR, Sweeney M, Li J et al (2007) Through the genetic bottleneck: *O. rufipogon* as a source of trait-enhancing alleles for *O. sativa*. Euphytica 154:317–339

Mei HW, Xu JL, Li ZK et al (2006) QTLs influencing panicle size detected in two reciprocal introgressive line (IL) populations in rice (*Oryza sativa* L.). Theor Appl Genet 112:648–656

Meyer RC, Kusterer B, Lisec J et al (2010) QTL analysis of early stage heterosis for biomass in *Arabidopsis*. Theor Appl Genet 120:227–237

Miller JC, Tanksley SD (1990) RFLP analysis of phylogenetic relationships and genetic variation in the genus *Lycopersicon*. Theor Appl Genet 80:437–448

Minutolo M, Amalfitano C, Evidente A et al (2013) Polyphenol distribution in plant organs of tomato introgression lines. Nat Prod Res 27:787–795

Monforte AJ, Tanksley SD (2000a) Development of a set of near isogenic and backcross recombinant inbred lines containing most of the Lycopersicon hirsutum genome in a L. esculentum genetic background: a tool for gene mapping and gene discovery. Genome 43:803–813

Monforte AJ, Tanksley SD (2000b) Fine mapping of a quantitative trait locus (QTL) from Lycopersicon hirsutum chromosome 1 affecting fruit characteristics and agronomic traits: breaking linkage among QTLs affecting different traits and dissection of heterosis for yield. Theor Appl Genet 100:471–479

Monforte AJ, Friedman E, Zamir D, Tanksley SD (2001) Comparison of a set of allelic QTL-NILs for chromosome 4 of tomato: Deductions about natural variation and implications for germplasm utilization. Theor Appl Genet 102:572–590

Morgan MJ, Osorio S, Gehl B et al (2013) Metabolic engineering of tomato fruit organic acid content guided by biochemical analysis of an introgression line. Plant Physiol 161:397–407

Moyle LC, Graham EB (2005) Genetics of hybrid incompatibility between *Lycopersicon esculentum* and *L. hirsutum*. Genetics 169:355–373

Moyle LC, Nakazato T (2008) Comparative genetics of hybrid incompatibility: sterility in two *Solanum* species crosses. Genetics 179:1437–1453

Overy, SA, Walker HJ, Malone S et al (2005) Application of metabolite profiling to the identification of traits in a population of tomato introgression lines. J Exp Bot 56:287–296

Pan Q, Liu YS, Budai-Hadrian O et al (2000) Comparative genetics of nucleotide binding site-leucine rich repeat resistance gene homologues in the genomes of two dicotyledons: tomato and Arabidopsis. Genetics 155:309–322

Paran I, Zamir D (2003) Quantitative traits in plants: beyond the QTL. Trends Genet 19:303–306

Paterson AH, DeVerna JW, Lanini B, Tanksley SD (1990) Fine mapping of quantitative trait loci using selected overlapping recombinant chromosomes, in an interspecies cross of tomato. Genetics 124:735–742

Pea G, Paulstephenraj P, Cane MA et al (2009) Recombinant near-isogenic lines: a resource for the mendelization of heterotic QTL in maize. Mol Genet Genomics 281:447–457

Peleman JD, van der Voort JR (2003) Breeding by design. Trends Plant Sci 8:330–334

Peralta IE, Spooner DM, Knapp S (2008) Taxonomy of wild tomatoes and their relatives (Solanum sections Lycopersicoides, Juglandifolia, Lycopersicon; Solanaceae). Syst Bot Monogr 84:1–186

Pestsova EG, Börner A, Röder MS (2001) Development of a set of triticum aestivum-aegilops tauschii introgression lines. Hereditas 135:139–143

Pestsova EG, Börner A, Röder MS (2006) Development and QTL assessment of triticum aestivum-aegilops tauschii introgression lines. Theor Appl Genet 112:634–647

Plunknett DL, Smith NJH, Williams JT, Murthi-Anishetti N (1987) Gene Banks and the World's Food. Princeton Univ. Press, Princeton, New Jersey

Prudent M, Bertin N, Génard M et al (2010) Genotype-dependent response to carbon availability in growing tomato fruit. Plant Cell Environ 33:1186–1120

Prudent M, Causse M, Génard M et al (2009) Genetic and physiological analysis of tomato fruit weight and composition: influence of carbon availability on QTL detection. J Exp Bot 60:923–937

Prudent M, Lecomte A, Bouchet JP et al (2011) Combining ecophysiological modelling and quantitative trait locus analysis to identify key elementary processes underlying tomato fruit sugar concentration. J Exp Bot 62:907–919

Ramsay LD, Jennings DE, Bohuon EJR et al (1996) The construction of a substitution library of recombinant backcross lines in *Brassica oleracea* for the precision mapping of quantitative trait loci. Genome 39:558–567

Rick CM (1969) Controlled introgression of chromosomes of *Solanum pennellii* into *Lycopersicum esculentum*: segregation and recombination. Genetics 62:753–768

Rick CM (1982) The potential of exotic germplasm for tomato improvement. Vasil IK, Scowcroft WR, Frey KJ (eds) Plant improvement and somatic cell genetics. Academic Press, New York, pp 1–28

Ron M, Dorrity MW, de Lucas M et al (2013) Identification of novel loci regulating inter-specific variation in root morphology and cellular development in tomato. Plant Physiol Apr 10. [Epub ahead of print]

Ronen G, Carmel-Goren L, Zamir D, Hirschberg J (2000) An alternative pathway to beta-carotene formation in plant chromoplasts discovered by map-based cloning of beta and old-gold color mutations in tomato. Proc Natl Acad Sci U S A 97:11102–11107

Ronen G, Cohen M, Zamir D, Hirschberg J (1999) Regulation of carotenoid biosynthesis during tomato fruit development: expression of the gene for lycopene epsilon-cyclase is down-regulated during ripening and is elevated in the mutant Delta. Plant J 17:341–351

Rousseaux MC, Jones CM, Adams D et al (2005) QTL analysis of fruit antioxidants in tomato using Lycopersicon pennellii introgression lines. Theor Appl Genet 111:1396–1408

Sacco A, Di Matteo A, Lombardi N et al (2013) Quantitative trait loci pyramiding for fruit quality traits in tomato. Mol Breed 31:217–222

Saha S, Jenkins JN, Wu J et al (2006) Effects of chromosome-specific introgression in upland cotton on fiber and agronomic traits. Genetics 172:1927–1938

Saha S, Wu J, Jenkins JN et al (2010) Genetic dissection of chromosome substitution lines of cotton to discover novel *Gossypium barbadense* L. alleles for improvement of agronomic traits. Theor Appl Gene 120:1193–1205

Saha S, Wu J, Jenkins JN et al (2011) Delineation of interspecific epistasis on fiber quality traits in Gossypium hirsutum by ADAA analysis of intermated *G. barbadense* chromosome substitution lines. Theor Appl Genet 122:1351–1361

Saha S, Wu J, Jenkins JN et al (2013) Interspecific chromosomal effects on agronomic traits in *Gossypium hirsutum* by AD analysis using intermated *G. barbadense* chromosome substitution lines. Theor Appl Genet 126:109–117

Saito K, Matsuda F (2010) Metabolomics for functional genomics, systems biology, and biotechnology. Annu Rev Plant Biol 61:463–489

Salvi S, Corneti S, Bellotti M et al (2011) Genetic dissection of maize phenology using an intraspecific introgression library. BMC Plant Biol 11:4

Salvi S, Tuberosa R (2005) To clone or not to clone plant QTLs: present and future challenges. Trends Plant Sci 10:297–304

Sato K, Takeda K (2009) An application of high-throughput SNP genotyping for barley genome mapping and characterization of recombinant chromosome substitution lines. Theor Appl Genet 119:613–619

Sayed MA, Schumann H, Pillen K et al (2012) AB-QTL analysis reveals new alleles associated to proline accumulation and leaf wilting under drought stress conditions in barley (*Hordeum vulgare* L.). BMC Genet 13:61

Schauer N, Semel Y, Balbo I et al (2008) Mode of inheritance of primary metabolic traits in tomato. Plant Cell 20:509–523

Schauer N, Semel Y, Roessner U et al (2006) Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. Nat Biotechnol 24:447–454

Schilmiller A, Shi F, Kim J et al (2010) Mass spectrometry screening reveals widespread diversity in trichome specialized metabolites of tomato chromosomal substitution lines. Plant J 62:391–403

Schilmiller AL, Charbonneau AL, Last RL (2012) Identification of a BAHD acetyltransferase that produces protective acyl sugars in tomato trichomes. Proc Natl Acad Sci U S A 109:16377–16382

Schmalenbach I, Körber N, Pillen K (2008) Selecting a set of wild barley introgression lines and verification of QTL effects for resistance to powdery mildew and leaf rust. Theor Appl Genet 117:1093–1106

Schmalenbach I, Léon J, Pillen K (2009) Identification and verification of QTLs for agronomic traits using wild barley introgression lines. Theor Appl Genet 118:483–497

Schmalenbach I, March TJ, Bringezu T et al (2011) High-resolution genotyping of wild barley introgression lines and fine-mapping of the threshability locus thresh-1 using the Illumina GoldenGate assay. G3 (Bethesda) 1:187–196

Schmalenbach I, Pillen K (2009) Detection and verification of malting quality QTLs using wild barley introgression lines. Theor Appl Genet 118:1411–1427

Sela-Buurlage MB, Budai-Hadrian O, Pan Q et al (2001) Genome-wide dissection of Fusarium resistance in tomato reveals multiple complex loci. Mol Genet Genomics 265:1104–1111.

Semel Y, Nissenbaum J, Menda N et al (2006) Overdominant quantitative trait loci for yield and fitness in tomato. Proc Natl Acad Sci U S A 103:12981–12986

Severin AJ, Peiffer GA, Xu WW et al (2010) An integrative approach to genomic introgression mapping. Plant Physiol 154:3–12

Shivaprasad PV, Dunn RM, Santos BA et al (2012) Extraordinary transgressive phenotypes of hybrid tomato are influenced by epigenetics and small silencing RNAs. EMBO J 31:257–266

Sim SC, Van Deynze A, Stoffel K et al (2012) High-density SNP genotyping of tomato (*Solanum lycopersicum* L.) reveals patterns of genetic variation due to breeding. PLoS One 7:e45520

Simmonds NW (1976) Evolution of crop plants. Longman, London, New York

Singer JB, Hill AE, Burrage LC et al (2004) Genetic dissection of complex traits with chromosome substitution strains of mice. Science 304:445–448

Sobrizal K, Ikeda K, Sanchez PL et al (1996) Development of *Oryza glumaepatula* introgression lines in rice, *O. sativa* L. Rice Genet Newsl 16:107

Steinhauser MC, Steinhauser D, Gibon Y et al (2011) Identification of enzyme activity quantitative trait loci in a *Solanum lycopersicum* x *Solanum pennellii* introgression line population. Plant Physiol 157:998–1014

Stelly DM, Saha S, Raska DA et al (2005) Registration of 17 Upland (*Gossypium hirsutum*) germplasm lines disomic for different G. barbadense chromosome or arm substitutions. Crop Sci 45:2663–2665

Stevens R, Buret M, Duffé P et al (2007) Candidate genes and quantitative trait loci affecting fruit ascorbic acid content in three tomato populations. Plant Physiol 143:1943–1953

Stevens R, Page D, Gouble B et al (2008) Tomato fruit ascorbic acid content is linked with monodehydroascorbate reductase activity and tolerance to chilling stress. Plant Cell Environ 31:1086–1096

Swamy BP, Sarla N (2008) Yield-enhancing quantitative trait loci (QTLs) from wild species. Biotechnol Adv 26:106–120

Szalma SJ, Hostert BM, Ledeaux JR et al (2007) QTL mapping with near-isogenic lines in maize. Theor Appl Genet 114:1211–1228

Tadmor Y, Fridman E, Gur A et al (2002) Identification of malodorous, a wild species allele affecting tomato aroma that was selected against during domestication. J Agric Food Chem 50:2005–2009

Tan L, Liu F, Xue W et al (2007) Development of *Oryza rufipogon* and *O. sativa* introgression lines and assessment for yield-related quantitative trait loci. Journal of Integrative Plant Biology 49:871–884

Tan L, Zhang P, Liu F et al (2008) Quantitative trait loci underlying domestication- and yield-related traits in an *Oryza sativa* x *Oryza rufipogon* advanced backcross population. Genome 51:692–704

Tang J, Yan J, Ma X et al (2010) Dissection of the genetic basis of heterosis in an elite maize hybrid by QTL mapping in an immortalized F2 population. Theor Appl Genet 120:333–340

Tanksley SD (1993) Mapping polygenes. Annu Rev Genet 27:205–233

Tanksley SD, Grandillo S, Fulton TM et al (1996) Advanced backcross QTL analysis in a cross between an elite processing line of tomato and its wild relative *L. pimpinellifolium*. Theor Appl Genet 92:213–224

Tanksley SD, McCouch SR (1997) Seed banks and molecular maps: unlocking genetic potential from the wild. Science 277:1063–1066

Tanksley SD, Nelson JC (1996) Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. Theor Appl Genet 92:191–203

Tester M, Langridge P (2010) Breeding technologies to increase crop production in a changing world. Science 327:818–822

Tian F, Li de J, Fu Q et al (2006) Construction of introgression lines carrying wild rice (Oryza rufipogon Griff.) segments in cultivated rice (*Oryza sativa* L.) background and characterization of introgressed segments associated with yield-related traits. Theor Appl Genet 112:570–580

Tian L, Tan L, Liu F et al (2011) Identification of quantitative trait loci associated with salt tolerance at seedling stage from Oryza rufipogon. J Genet Genomics 38:593–601

Tieman DM, Zeigler M, Schmelz EA et al (2006) Identification of loci affecting flavour volatile emissions in tomato fruits. J Exp Bot 57:887–896

Tomato Genome Consortium (2012) The tomato genome sequence provides insights into fleshy fruit evolution. Nature 485:635–641

Törjék O, Meyer RC, Zehnsdorf M et al (2008) Construction and analysis of 2 reciprocal *Arabidopsis* introgression line populations. J Hered 99:396–406

Toubiana D, Semel Y, Tohge T et al (2012) Metabolic profiling of a mapping population exposes new insights in the regulation of seed metabolism and seed, fruit, and plant relations. PLoS Genet 8:e1002612

Tripodi P, Di Dato F, Maurer S et al (2010) A genetic platform of tomato multi-species introgression lines: present and future. In: Proceedings of the 7th Solanaceae Conference, Dundee, 5-9 September 2010, pp 166

Uauy C, Distelfeld A, Fahima T et al (2006) A NAC Gene regulating senescence improves grain protein, zinc, and iron content in wheat. Science 314:1298–1301

Uozumi A, Ikeda H, Hiraga M et al (2012) Tolerance to salt stress and blossom-end rot in an introgression line, IL8-3, of tomato. Sci Hortic 138:1–6

Van der Hoeven RS, Monforte AJ, Breeden D et al (2000) Genetic control and evolution of sesquiterpene biosynthesis in Lycopersicon esculentum and L. hirsutum. Plant Cell 12:2283–2294

van der Knaap E, Sanyal A, Jackson SA, Tanksley SD (2004) High-resolution fine mapping and fluorescence in situ hybridization analysis of sun, a locus controlling tomato fruit shape, reveals a region of the tomato genome prone to DNA rearrangements. Genetics 168:2127–2140

Van Schalkwyk A, Wenzl P, Smit S et al (2012) Bin mapping of tomato diversity array (DArT) markers to genomic regions of *Solanum lycopersicum* × *Solanum pennellii* introgression lines. Theor Appl Genet 124:947–56

Varshney RK, Hoisington DA, Tyagi AK (2006) Advances in cereal genomics and applications in crop breeding. Trends Biotech 24:490–499

Varshney RK, Mohan SM, Gaur PM et al (2013) Achievements and prospects of genomics-assisted breeding in three legume crops of the semi-arid tropics. Biotechnol Adv. 2013 Jan 11. doi:pii: S0734-9750(13)00003-7. 10.1016/j.biotechadv.2013.01.001

von Korff M, Wang H, Léon J, Pillen K (2004) Development of candidate introgression lines using an exotic barley accession (*Hordeum vulgare* ssp. *spontaneum*) as donor. Theor Appl Genet. 109:1736–1745

Wang G, Schmalenbach I, von Korff M et al (2010) Association of barley photoperiod and vernalization genes with QTLs for flowering time and agronomic traits in a BC2DH population and a set of wild barley introgression lines. Theor Appl Genet 120:1559–74

Wang ZY, Second G, Tanksley SD (1992) Polymorphism and phylogenetic relationships among species in the genus Oryzae as determined by analysis of nuclear RFLPs. Theor Appl Genet 83:565–581

Wu F, Mueller LA, Crouzillat D et al (2006) Combining bioinformatics and phylogenetics to identify large sets of single-copy orthologous genes (COSII) for comparative, evolutionary and systematic studies: a test case in the euasterid plant clade. Genetics 174:1407–1420

Xiao J, Li J, Grandillo S et al (1996) Genes from wild rice improve yield. Scientific Correspondence, Nature 384:223–224

Xiao J, Li J, Grandillo S et al (1998) Identification of trait-improving quantitative trait loci alleles from a wild rice relative, *Oryza rufipogon*. Genetics 150:899–909

Xu X, Martin B, Comstock JP et al (2008) Fine mapping a QTL for carbon isotope composition in tomato. Theor Appl Genet 117:221–233

Xu J, Zhao Q, Du P et al (2010) Developing high throughput genotyped chromosome segment substitution lines based on population whole-genome re-sequencing in rice (*Oryza sativa* L.). BMC Genomics 11:656

Yates HE, Frary A, Doganlar S et al (2004) Comparative fine mapping of fruit quality QTLs on chromosome 4 introgressions derived from two wild tomato species. Euphytica 135:283–296

Yu J, Zhang K, Li S et al (2013) Mapping quantitative trait loci for lint yield and fiber quality across environments in a *Gossypium hirsutum* × *Gossypium barbadense* backcross inbred line population. Theor Appl Genet 126:275–287

Zamir D (2001) Improving plant breeding with exotic genetic libraries. Nat Rev Genet 2:983–989

Zamir D, Eshed Y (1998) Tomato genetics and breeding using nearly isogenic introgression lines derived from wild species. In: Paterson AH (ed) Molecular dissection of complex traits. CRC Press, Boca Raton, FL, pp 207–217

Zhao L, Zhou H, Lu L et al (2009) Identification of quantitative trait loci controlling rice mature seed culturability using chromosomal segment substitution lines. Plant Cell Rep 28:247–256

Zong G, Wang A, Wang L et al (2012) A pyramid breeding of eight grain-yield related quantitative trait loci based on marker-assistant and phenotype selection in rice (*Oryza sativa* L.). J Genet Genomics 39:335–350

Chapter 5
Microphenomics for Interactions of Barley with Fungal Pathogens

Dimitar Douchkov, Tobias Baum, Alexander Ihlow, Patrick Schweizer and Udo Seiffert

Abstract Current high-throughput plant phenotyping pipelines are mainly focused on quantitative assessment of macroscopic parameters. Such morphological or physiological parameters measured on entire plants or major plant parts are not well adapted to the accurate description of plant-pathogen interactions because plant pathogens are microorganisms causing only microscale changes in their hosts or non-hosts during the initial stages on infection, which often decide about susceptibility or resistance. This makes the use of microscopic phenomics techniques unavoidable. However, the high-throughput requirements of modern phenomics screens represent a considerable challenge to the available microscopic approaches and underlying instruments used to characterize plant-pathogen interactions. To meet this challenge we have developed a platform that combines high-throughput DNA cloning, single cell transformation protocols, and automated microscopy and phenotyping that we called "microphenomics". It was used to address the function of genes in nonhost- and race-nonspecific host resistance of barley interacting with the powdery mildew fungus *Blumeria graminis.* More than 1,300 genes derived from plant or fungal genomes were tested by silencing and approximately 100 of them had a significant effect on the resistance or susceptibility to the pathogen. The chapter gives an overview on the current status of this microphenomics platform for very early and early stages of plant-pathogen interactions.

D. Douchkov (✉) · P. Schweizer
Leibniz-Institute of Plant Genetics and Crop Plant Research (IPK),
Pathogen-Stress Genomics, Corrensstrasse 3, 06466 Gatersleben, Germany
e-mail: douchkov@ipk-gatersleben.de

P. Schweizer
e-mail: schweizer@ipk-gatersleben.de

T. Baum · U. Seiffert
Fraunhofer-Institute for Factory Operation and Automation (IFF),
Biosystems Engineering, Sandtorstrasse 22, 39106 Magdeburg
e-mail: mail@tobiasbaum.net

U. Seiffert
e-mail: Udo.Seiffert@iff-fraunhofer.de

A. Ihlow
Ilmenau University of Technology, Institute for Information Technology
P.O. Box 100565, 98684 Ilmenau, Germany
e-mail: alexander.ihlow@tu-ilmenau.de

R. Tuberosa et al. (eds.), *Genomics of Plant Genetic Resources,*
DOI 10.1007/978-94-007-7575-6_5,

5.1 Introduction

Recent breakthroughs in sequencing and genotyping technologies brought plant genetics to a level where other disciplines are struggling to keep pace. Sequencing of the genome of the model plant *Arabidopsis thaliana* represented a milestone in plant genomics. Subsequently the genomes of over 40 higher plant species have been partially or fully sequenced, annotated and made publicly available in the NCBI's GenBank (status 2012). Among them are several economically important crop plant species including rice, maize, tomato, potato to name just a few. New next-generation sequencing technologies will make re-sequencing of entire genomes affordable and fast to examine allelic variation. The project for sequencing of 1,001 *Arabidopsis* genomes, which is in progress, is an example for such a coordinated effort.

At the same time the existing approaches for phenotyping are usually associated with much lower performance in terms of precision and throughput, which prevents optimal use of the large amount of genotypic data by linking it with trait information. Therefore, the development of comprehensive high-throughput phenotyping systems is of crucial importance for contemporary plant science.

High-throughput phenotyping also referred to as phenomics along with the other "omics" disciplines, is one of the most recent and most challenging research directions in biology. Phenomics is asking for quantitative phenotypes in large populations of cells or organisms depending on natural or induced variability in gene expression or structure. Pioneering work in phenomics has been performed in the model organisms yeast and the worm *Caenorhabditis elegans* (Maeda et al. 2001; Warringer et al. 2003). Arabidopsis was the first plant subject to a large-scale phenotypic profiling with a pilot study of T-DNA mutant screens in *A. thaliana* being reported by (Errampalli et al. 1991). Since then the number of projects for large-scale phenotyping of plant loss-of-function and gain-of-function collections is rapidly increasing (for review Kuromori et al. (2009)). There has been a worldwide effort to meet the need to develop plant-phenomics approaches and infrastructure. Several research teams were established to set up new methods of plant phenotyping—e.g. the Australian Plant Phenomics Facility "Plant Accelerator" in Adelaide, and the "Australian High Resolution Phenomics Centre" in Canberra; the Jülich Plant Phenotyping Centre and IPK Gatersleben in Germany; the Laboratory of Plant Ecophysiological responses to Environmental Stresses, and Ecotron, both in Montpellier, France; National Plant Phenomics Centre at Aberystwyth University in UK as well as the Biotron Experimental Climate Change Research Facility and the Green Crop Network in Canada. These facilities address specific phenotyping tasks that are part of an integrated full phenotyping pipeline within the International Plant Phenotyping Network (IPPN) founded in 2009 to improve communication and to set up international quality standards in this field. European (EPPN) as well as national (DPPN in Germany) counterparts of IPPN were established too. The current high-throughput phenotyping pipelines are mainly focused on measuring macroscopic parameters such as leaf area, chlorophyll content, plant height and width, growth rate, stress pigment concentration, biomass, color etc. Large-scale studies of plant-pathogen

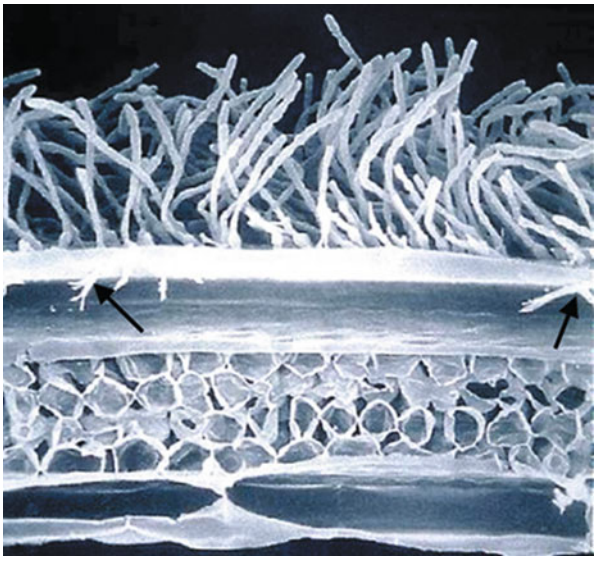

Fig. 5.1 The barley powdery mildew fungus *B. graminis* f.sp. *hordei* is an epiphytic fungal pathogen growing and heavily sporulating on the leaf surface. The only cells of the pathogen inside its host are haustoria with finger-like extrusions for nutrient uptake (*arrows*). Longitudinal section across an infected barley leaf at 120 h after inoculation

interactions such as segregation behavior in populations etc. are usually limited to disease symptoms scoring. In spite of the remarkable advances in remote-sensing technologies for plant diseases (Wijekoon et al. 2008; Bauriegel et al. 2011; Dammer et al. 2011; Peressotti et al. 2011; De Coninck et al. 2012; Mahlein et al. 2012a, 2012b) the measured macroscopic parameters are not specially designed to describe early stages of plant-pathogen interactions. In view of the fact that plant pathogens are microscopic objects during initial stages on infection, which are often decisive between susceptibility and resistance, the use of microscopic techniques is inevitable. As a consequence the development of high-throughput microscopic phenotyping systems is required. To designate specifically the combination of high-throughput phenotyping as strategy, and microscopic techniques as platform we coined the term "microphenomics".

One of the plant-pathogen systems with the status of a model interaction is the epidermis of barley attacked by the powdery mildew fungus *Blumeria graminis* f.sp. *hordei* (Bgh). An outstanding feature of this fungus is its epiphytic hyphal growth on leaf or stem surfaces with only haustorial cells (specialized for nutrient uptake) being localized inside the host where they invaginate the plasmalemma of epidermal cells (Fig. 5.1).

Therefore, the interaction is easy to follow under the microscope, and developmental stages of the fungus as well as cytological plant responses can be described in detail and also quantified. The model status of the barley/Bgh pathosystem is reflected by a large body of published information on cellular events, genes, resistance loci etc. affecting the interaction. We recently established two automated systems for quantitative microscopic analysis of the barley/powdery mildew interaction: The first system generates quantitative phenotypic data on fungal penetration efficiency

and haustorium formation, whereas the second system generates data on hyphal growth. These developmental parameters, together with sporulation efficiency, are determining the success of Bgh and other powdery mildew fungi attacking altogether hundreds of host plant species. The first of these microscopic phenotyping platforms has already been used in several mid- to high-throughput screens as summarized later in this article whereas the second system is still in its pilot testing phase.

Suitable pathosystems and automated quantification tools are two of the prerequisites for high-throughput phenomics of plant-pathogen interactions. The availability of methods to select particular gene groups or genotypes is the third requirement. Natural or induced allelic variability in genotype collections is the basis of what we called "genotype-oriented phenomics approach" (see Sect. 5.3). In contrast, the gene-oriented approach is challenging individual genes, which are manipulated via transient or stable gene silencing or gene overexpression. In the last decade RNAi became a powerful tool for gene silencing in transient whole-plant or single-cell assays such as virus-induced gene silencing (VIGS) or transient-induced gene silencing (TIGS) (see Wise et al. (2009) for an extensive review). Developing techniques for high-throughput RNAi-vector construction (Douchkov et al. 2005; Wielopolska et al. 2005), and fast methods to deliver them to the plant (Schweizer et al. 1999; Douchkov et al. 2005) are paving the way to high-throughput phenomics of interactions of barley with fungal pathogens. In this context we would like to put special emphasis on TIGS, an RNAi-based approach where the silencing constructs are delivered to epidermal plant cells via particle bombardment. Although the silencing effect is restricted to a limited number of epidermal cells the use of reporter genes allows easy localization of transformed cells, which makes TIGS a excellent method for studying cell-autonomous phenomena such as the early interaction with powdery mildew fungi.

5.1.1 Phenotype Driven Screens for Plant–Pathogen Interactions in Barley

In the past years there has been a major effort of the international community into developing barley as a model crop of the *Triticeae* tribe including wheat and rye. Barley is favorable for phenomics approaches because of the existing large *ex-situ* germplasm collections carrying desired alleles and an enormous amount of expert knowledge concerning individual traits. Valuable genomics resources were established for barley such as a large EST collection (Zhang et al. 2004), macro- and microarrays for transcriptome analysis (Close et al. 2004; Zierold et al. 2005; Zellerhoff et al. 2010), high-density genetics maps (Close et al. 2009), efficient plant transformation protocols (Tingay et al. 1997; Kumlehn et al. 2006), mutant (TILLING) platforms (http://www.gabi-till.de), model pathosystems with agronomically important fungi (Panstruga and Schulze-Lefert 2002) and, importantly, medium- to high-throughput gene silencing methods (Schweizer et al. 1999; Lacomme et al. 2003; Douchkov et al. 2005).

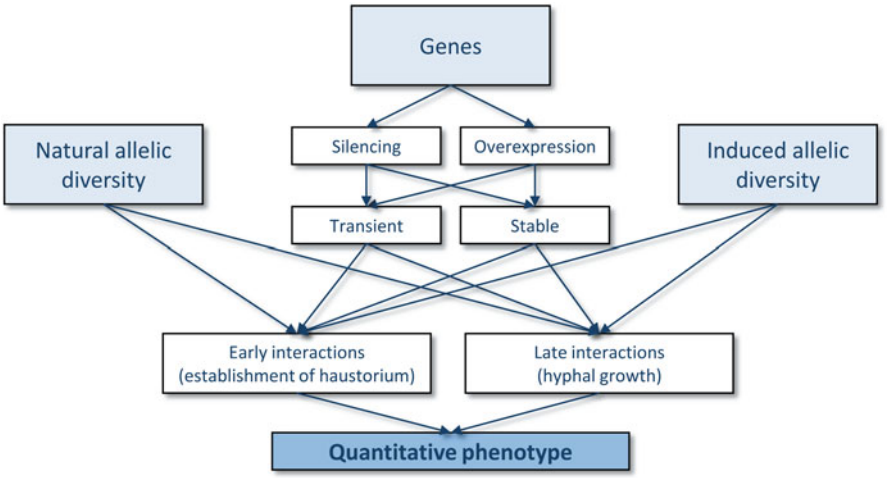

Fig. 5.2 Basic strategies for phenomics of interactions of barley with fungal pathogens. The genotype-based approach is using natural or induced allelic variability in genotype collections. The gene-oriented approach is challenging individual genes or gene groups via transient or stable gene silencing or gene overexpression

Despite the progressively increasing throughput of phenomics screens it is not yet feasible to achieve genome-wide screenings for particular phenotypes in *Triticeae* species. All current phenomics studies still begin with defining a set of candidates to be tested. There are two basic approaches for selection: The gene-oriented approach will focus on the selection of groups of genes based on predicted function, regulation or co-localization with QTLs of interest. The genotype-based approach will explore the natural or induced variability of the barley genome in gene bank material, associating-mapping collections, and biparental QTL populations at the one hand, and randomly mutagenized backcross populations or tagged-mutant collections on the other (Fig. 5.2).

5.1.2 Gene-Oriented Phenomics Strategy

The gene-based approach, as suggested by the name, is focusing on the selection of individual genes or gene groups. There are basically three ways for gene-oriented selection of candidates (Fig. 5.3): Transcript regulation, predicted function of the gene, and linkage or linkage disequilibrium with QTL or SNP-trait-association peak markers, respectively.

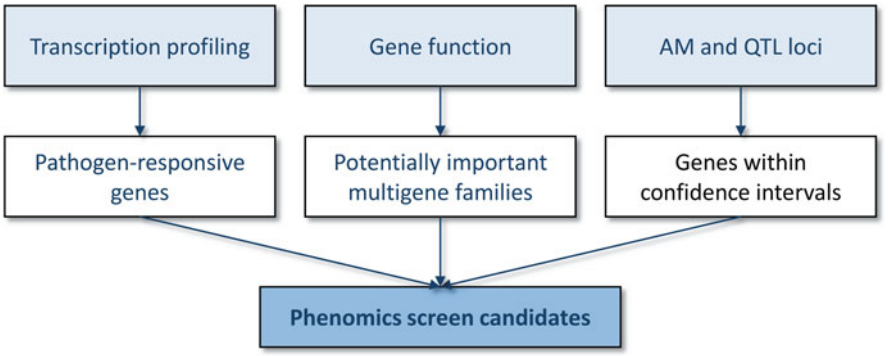

Fig. 5.3 Three main approaches to gene-oriented phenomics screen. The candidate genes can be selected based on their transcript expression profile, gene function or association mapping and co-localization with quantitative trait loci

5.1.2.1 Transcription Profiling

Micro- or macroarray-based transcript profiling, or transcriptomics, of a plant-pathogen interaction will reveal genes that are associated with transcript regulation in responses to pathogen attack. Although transcript regulation may not be the most decisive step of the regulation of protein amount or -activity, it is widely accepted that transcript regulation is a meaningful filter to narrow down the number of genes that are relevant under certain physiological (stress) conditions. In barley, a set of 206,000 ESTs has been established at the IPK Gatersleben representing a valuable resource for gene discovery based on sequence or expression (Zhang et al. 2004). Based on these and other EST libraries, arrays have been designed and used for the identification of powdery mildew-responsive host genes providing a basis for candidate-gene selection (Caldo et al. 2004; Zierold et al. 2005; Zellerhoff et al. 2010; Zhang et al. 2010). In the last years versatile tools for the meta-analysis of gene expression have been developed. PLEXdb (http://plexdb.org/) is a public resource for large-scale gene expression analysis of plants and provides a unified web interface to support the functional interpretation of microarray experiments by integrating structural-genomic and phenotypic data (Wise et al. 2006, 2007). BarleyBase (Shen et al. 2005), which is now integrated into PLEXdb, contains expression data from 54 publicly available experiments corresponding to 1,724 hybridizations using the Affymetrix 22K Barley1 GeneChip. Genevestigator (https://www.genevestigator.com) is another platform, which is defined as a high performance search engine for gene expression used to explore the spatiotemporal and response characteristics of genes of interest. Currently Genenvestigator holds 52 barley experiments with 1,650 hybridization samples. By using transcript up-regulation upon *B. graminis* infection as selection criterion we tested 468 barley genes in a TIGS screening for breakdown non-host resistance against the wheat powdery mildew and 377 genes (largely overlapping with the previous set) in a TIGS screening for modulating race-nonspecific host resistance (see Table 5.1).

A direct measure of gene expression by transcript sequencing has the advantage to be unrestricted by the availability of sequenced genome or well-characterized transcriptome. Several transcriptome sequencing methods were developed in the past including serial analysis of gene expression (SAGE) (Velculescu et al. 1995), massively parallel signature sequencing (MPSS) (Brenner et al. 2000), etc. but real break-through was made with advance of the ultra high-throughput transcript sequencing, or so called RNA-Seq (Wang et al. 2010). However, the transcript regulation of a gene may not directly reflect its function and importance. Moreover, it has been shown that protein abundance is rather weakly correlated to mRNA abundance and translational activity (Beyer et al. 2004). Therefore the transcriptomics approach is insufficient as one and only gene discovery tool and should be complemented by other criteria such as gene function or linkage to association-mapping or QTL-mapping peak markers.

5.1.2.2 Gene Function

An increasing number of genes have been functionally described with respect to plant-pathogen interactions. These often belong to multigene families, which are therefore of great potential interest for molecular plant pathologists. We used these approaches to screen for genes modulating the basal resistance in barley to Bgh and tested 410 genes belonging to eight multigene families.

ABC-Transporters The ABC-transporter superfamily is one of the largest protein families in plants. The first reports on plant ABC transporters showed that they are implicated in detoxification processes. However, recent results indicate that the function of this protein family is not restricted to detoxification processes (Martinoia et al. 2002; Crouzet et al. 2006; Yazaki et al. 2009). Several ATP-binding cassette-type (ABC) transporters were shown to be implicated in plant defense responses and secretion of plant antimicrobial compounds: Arabidopsis PEN3 (Kobae et al. 2006; Stein et al. 2006), NpABC1/PDR1 from *Nicotiana plumbaginifolia* (Jasinski et al. 2001), and wheat Lr34 (Lagudah et al. 2009).

SWEET Sugar Transporters Sugar efflux transporters are required for the production of plant nectar and for the development of plant seed and pollen. Recently a new class of sugar efflux transporters was discovered and named SWEETs. Several SWEET proteins are specifically exploited by bacterial pathogens most likely for nutrient uptake by means of direct binding of a bacterial effector to the SWEET promoter (Frommer et al. 2010). SWEETs may thus become an important target for controlling late biotrophic host-pathogen interactions, which are characterized by massive transport of nutrients across biomembranes of plant and pathogen.

Cellulose Synthase-Like Proteins Cellulose synthase-like (CLS) genes are a family of plant proteins defined by their similarity to the cellulose synthase CesA. Several members of this family were shown to be involved in the synthesis of noncellulosic cell-wall carbohydrates. Up to now CSL genes were not discussed with respect to plant defense. However, plants respond to the penetration attempts with inducible

local cell-wall appositions and cell-wall reorganization. It therefore appears possible that CLSs take part in these processes.

E3 Ubiquitin Ligases Several lines of evidences indicate that the ubiquitin/proteasome pathway plays a role in pathogen-attacked plants. Partial depletion of cellular ubiquitin levels by transient RNAi induced extreme susceptibility of transformed cells toward the appropriate host pathogen *B. graminis*f. sp. *hordei* whereas other types of resistance remained unaffected (Dong et al. 2006). RING and U-box domain E3-ligases are rapidly induced by biotic stress and may function as both positive and negative regulators of immune responses (Trujillo and Shirasu 2010). E3 ubiquitin ligases are the specificity determinants of ubiquitination. There are about 1,400 E3-ligases encoded by the *A. thaliana* genome (Mazzucotelli et al. 2006), and its U-box E3-ligases PUB22, PUB23, and PUB24 (PLANT U-BOX 22–24) were reported as negative regulators of basal resistance (Shirasu et al. 2008). Also, the rapidly Avr9/Cf9-elicited (ACRE) E3-ligase genes ACRE74 (CMPG1) and ACRE276 (PUB17) are required for the hypersensitive response of tobacco (Gonzalez-Lamothe et al. 2006; Yang et al. 2006).

Class III (Secreted) Peroxidases The secreted class III plant peroxidases (PRX) are known to be induced during host plant defense. They belong to a large multigene family and participate in a broad range of physiological processes, such as hydrogen peroxide production, lignification, cell wall cross-linking, and synthesis of phytoalexins. RRXs also appear to contribute to switching on the hypersensitive response (HR) upon attack by inappropriate or incompatible pathogens (Pedreno et al. 2009).

Receptor-Like Kinases The receptor-like protein kinases/Pelle-like kinases (RLK/Pelle) also known as IRAK in plants are a massively expanded family of proteins. The RLK/Pelle gene family is among the largest in the *A. thaliana* genome with more than 600 members, which constitutes about 60 % of all kinases of *A. thaliana* (Shiu and Bleecker 2003). Rice has nearly twice as many RLK/Pelle members as *A. thaliana*, and other grass species may have even more. RLK/Pelle proteins play roles in several cellular processes ranging from growth regulation to defense response (Shiu et al. 2009). Typical RLK/Pelle consists of ligand-binding and protein-kinase domains and are implicated in downstream signaling.

Several RLK/Pelle are reported to function as so-called pattern recognition receptors (PRRs). These include classical examples such as the Xa21 resistance gene of rice to *Xanthomonas campestris* pv. *oryzae* (Song et al. 1995; Wang et al. 1996; Zhu et al. 2011) and the pathogen-associated molecular pattern (PAMP)-receptor FLS2 of *A. thaliana* recognizing bacterial flagellin (Gomez-Gomez et al. 2001). The number of RLK/Pelle genes relevant to plant defense or susceptibility is rapidly growing and includes among others the somatic embryogenesis receptor kinase (SERK)3/brassinosteroid-associated kinase (BAK)1 (Chinchilla et al. 2007; Heese et al. 2007), EFR (Zipfel et al. 2006), the LysM receptor-like kinase 1/chitin elicitor receptor kinase 1 (LysM RLK1/CERK1) (Shibuya et al. 2007; Lipka et al. 2009, 2010; Iizasa et al. 2010), and BIK1 (Lu et al. 2010). In summary, RLK/Pelle type of proteins may play a crucial role in initial pathogen recognition and defense as PAMP/MAMP receptors.

WRKY Transcription Factors WRKY factors are central components of the innate immune system of the plant and are probably involved in all types of immune response including PAMP- and microbial-associated molecular pattern (MAMP)-triggered immunity (PTI), effector-triggered immunity (ETI), and systemic acquired resistance. WRKYs are in the focus of an active research and several review articles are available (Eulgem et al. 2000; Eulgem and Somssich 2007; Rushton et al. 2010).

5.1.2.3 Fungal Genes

An interesting set of target genes for screening consists of Bgh-encoded sequences. This approach is based on the recently discovered phenomenon called "host-induced gene silencing" (HIGS; Nowara et al. 2010). HIGS targets encode proteins of the pathogen, which are silenced by RNAi mediated via the host. The exact mechanism of HIGS is still unknown but proof of concept was obtained by the knock down of mRNA of the fungal *Avr10* effector (Nowara et al. 2010).

5.1.2.4 Random Selection of Genes

A nonbiased, random selection of genes may also lead to discovery of defense or susceptibility related genes. Such approach would offer a chance to reveal novel gene functions or even to discovery of new pathways with a role in plant-pathogen interaction. However, the expected discovery rate would be relatively low and therefore a random selection approach would be productive only if the test group consists of a large number of candidates. Smaller groups of randomly selected genes can serve as an internal control as discovery rate of nonbiased selection.

5.1.2.5 Genes Co-localizing with Resistance QTL

High-throughput phenomics can also be used for "gene landing" in genetic intervals provided these intervals can be populated with a reasonably complete list of mapped genes. Genetic intervals of interest to a large number or researchers are those either flanking association-mapping (AM) peak markers (single-nucleotide polymorphisms (SNP) or gene haplotypes) or peak markers of quantitative-trait locus (QTL) mapping. The approach to follow for gene landing would be to first define confidence intervals surrounding the QTL or AM peak markers, which is achieved by analyzing genetic linkage of nearby markers in biparental QTL-mapping- or linkage disequilibrium in AM populations. The second step then consists of populating these intervals by as many mapped genes as possible. Ideally, physical maps of sequenced and annotated genomes exist in the species of interest. Third, an RNAi collection of the identified list of co-localizing genes is generated and used for high-throughput silencing in a genetic background that should reveal a phenotype when the causative gene is knocked down. Finally the superimposition of positional and functional data together

with proposed gene functions may guide us rapidly towards most likely candidates for the respective QTL. Although a final, annotated genome of barley does not yet exist, the species is amenable to the approach because a large majority of genes have been mapped either by wet-lab approaches or by synteny-based gene-order prediction (Mayer et al. 2011). We have used this approach to screen 128 genes co-localizing inside a QTL interval for powdery mildew resistance on barley chromosome 5H.

5.1.3 Genotype-Based Phenomics Strategy

In addition to gene-driven phenomics screens plant pathologists are also interested to identify new interesting genotypes carrying potentially valuable haplotypes (alleles) that might be used for allele introgression into adapted germplasm. Although this can be basically achieved by standard, macroscopic disease rating the approach presented here has the following advantages: First, its high degree of automation allows a higher throughput. In the barley/Bgh system we estimate a throughput of 1,600 genotypes per month. Second, the result will be a precise quantitative interaction phenotype based on the average number of formed haustoria per epidermal cell, or on the surface of formed hyphae after a certain time (static) or over time (time course). This is important because QTL-mediated durable resistance is the result from multiple alleles acting together in an additive or synergistic manner, each one alone contributing sometimes as little as 5 % of the observed phenotypic variation between parents. Two sources or allelic diversity are of prime interest to barley pathologists: Natural allelic diversity in plant genetic resources such as *ex-situ* genebank collections stored at IPK Gatersleben and elsewhere, and induced allelic diversity present e.g. in mutagenized backcross populations (Fig. 5.4).

5.1.3.1 Allelic Diversity in Plant Genetic Resources

In a number of gene banks worldwide including IPK Gatersleben hundreds of thousands of barley accessions consisting of wild barley (*H. vulgare* ssp. *spontaneum* plus other species of the genus), cultivated barley landraces and old or modern cultivars are conserved. In recent years the interest of the community has shifted from conservation, pure description of plant morphology etc. and of genetic diversity to the evaluation of valuable traits including pathogen resistance (Bhullar et al. 2009). The availability of phenomics tools for the quantitative assessment of different pathogen developmental parameters will allow classifying barley genetic resources into phenotypic groups, which will add value to genome-wide or candidate gene-based AM approaches. The same also holds true for the considerable number of established biparental population for resistance-QTL mapping and for the upcoming genetic resource of nested association-mapping (NAM) populations (Pillen et al., http://www.flowercrop.uni-kiel.de/projects/pillen).

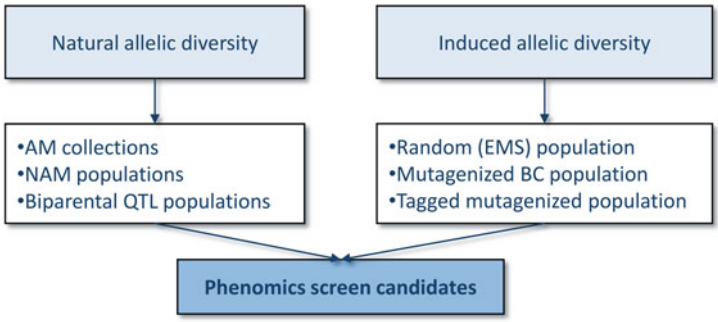

Fig. 5.4 Sources of allelic diversity for genotype-driven phenomic approaches in barley. *AM* association mapping, *NAM* nested association mapping, *QTL* quantitative trait locus, *EMS* ethyl methanesulphonate

5.1.3.2 Mutagenesis

Mutant collections are a powerful resource for the identification of genes underlying specific traits. Whereas tagged-mutant populations in barley have not yet been developed beyond proof of concept (Zhao et al. 2006), the approach of combining random mutagenesis with extensive backcrossing has yielded a valuable genetic resource to the barley community (Druka et al. 2011). The now available 881 mutagenized backcross lines in the genetic background of cv. Bowman are likely to yield very interesting data with respect to pathogen resistance and precise resistance types reflected, as discussed above, in microscopic parameters of fungal development.

5.2 Microphenomics Screen Workflow

During the last years we have established and used a high-throughput phenomics pipeline for TIGS and genotype screenings in the model system of barley and powdery mildew (Schweizer et al. 2000; Douchkov et al. 2005). Over 1,300 RNAi constructs have been used for silencing and tested for their effect on nonhost resistance, as well as for modulating the race-nonspecific (basal) host defense. The basic of this pipeline is summarized below (Fig. 5.5).

5.2.1 RNAi Library Preparation

A suitable starting material for generation of inverted-repeat constructs for RNAi in TIGS screens are existing EST clone collections, total cDNA or exon regions of genomic DNA that can be used as PCR templates. With decreasing prices synthetic DNA becomes an affordable alternative to the PCR-based techniques. The PCR

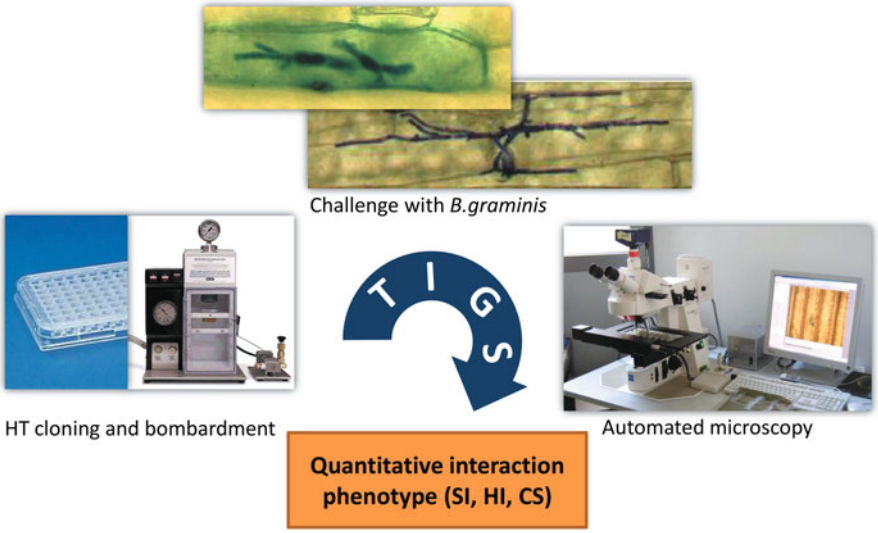

Fig. 5.5 The experimental pipeline for gene- and genotype-oriented phenomics in barley—powdery mildew system. Quantitative interaction phenotypes: susceptibility index (SI) = Σ susceptible cells/Σ GUS-cells; haustorium index (HI) = Σ haustoria/Σ GUS-cells; colony size (CS) = Σ of segmented hyphae pixels

fragments or synthetic DNA fragments are then cloned as inverted repeats into an RNAi vector in a high-throughput manner (for instance by the method described in (Douchkov et al. 2005). The obtained RNAi construct can be used in TIGS as well as for generating stable plant transformants.

5.2.2 Microscopy

Robust automated detection of pathogen structures requires high quality images with maximal contrast between fungal structures and the leaf background. Different staining methods can be applied depending on the analysis objectives.

5.2.2.1 Haustoria Staining

Expectedly the most challenging task—to render the fungal haustorium visible inside barley cells—turned out in fact fairly straight forward in practice because the X-Gluc dye for the detection of GUS reporter activity in transformed cells (Schweizer et al. 1999) produces a strong contrast for visible light microscopy. Therefore phenomics screens for early interactions (haustorium formation) typically use this type of staining.

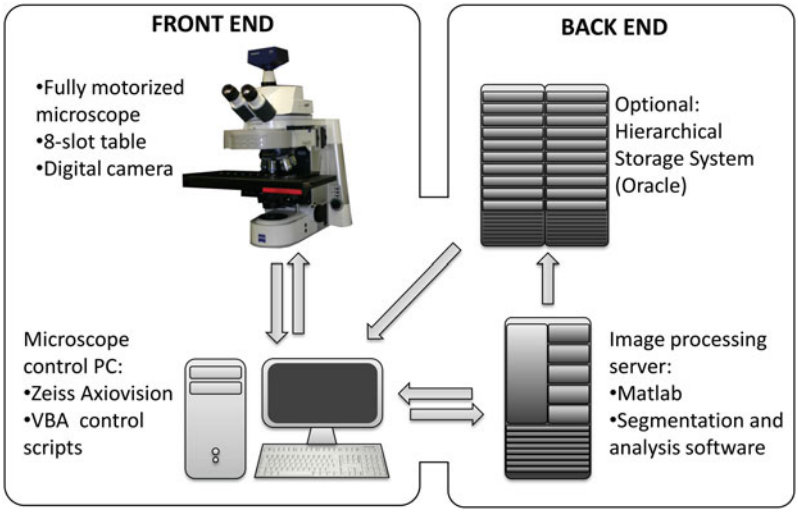

Fig. 5.6 Schematic diagram of the automated microscopic system. The front end includes a Zeiss Imager. Z1m or Axioplan2 microscope, AxioCamHRc digital camera, Märzhäuser Wetzlar SCAN 8 Präparate motorized table, and microscope control PC running Zeiss AxioVision, and VBA control scripts. The back end includes the image processing server running Matlab and segmentation analysis software, and optional hierarchical storage system for archiving and backup of the data. The *arrows* indicate the direction of data flow

5.2.2.2 Staining of Fungal Structure on the Leaf Surface

In contrast to the previous analysis, screens looking for late interaction phenotypes will need specific staining for fungal structures on the leaf surface. Modified Coomasie-blue staining protocols (Schweizer et al. 1993) represent a good compromise between technical simplicity required for high-throughput and high contrast necessary for automated analysis of the images.

5.2.3 Image Analysis and Processing

5.2.3.1 Recognition of *B. graminis* Haustoria in Barley

For the purpose of haustorium detection and its analysis a platform (Fig. 5.6) was created and described in (Ihlow et al. 2008).

As a first step of the process, an overview image of the entire slide is created at low magnification (5 × objective) and low image resolution (432 × 342 pixels), using an automated microscope multi-slide table. In this overview image, the transformed cells are being recognized as regions of interest. Subsequently, detailed images of transformed cells are taken, using the 10 × objective and a higher resolution (1,300 × 1,030 pixels).

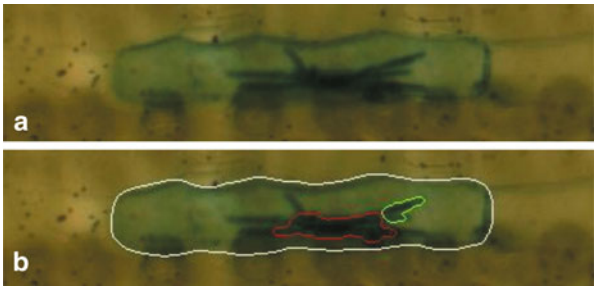

Fig. 5.7 Infected GUS–stained, transformed cell before (**a**) and after segmentation analysis (**b**). The *red* color of the boundary line indicates high probability of having a haustorium detected

The screening platform is separated into a front- and back end. The front end includes all the hardware components of the microscope and a PC running the software "AxioVision" to control the motor and electronic components of the microscope (XY stage, focus, lighting, lenses, camera, etc.). This system is connected via Ethernet network to the back end, which analyses the pre-scan and detects the GUS-transformed cells. Afterwards, it sends a position list of the regions of interest (ROIs) to the front end. The automated microscope generates a z-stacked image from each position (ROI) and sends it back to the image processing server, where it is analyzed for potential haustoria (Fig. 5.7).

5.2.3.2 HyphArea Platform

The precise scoring of fungal penetration efficiency in combination with forward or reverse genetic approaches can provide valuable information on mechanisms underlying the basal defense reaction of the plant cell, which is known to mostly affect fungal penetration. However, a system based on scoring just early interaction phenotypes will provide information only of "one side of the coin". Besides this it would be highly desirable in many cases to keep track of the plant-pathogen interaction over time. Even in the case of successful early establishment of the pathogen, its growth can be slowed down or completely detained at later stages by plant defense mechanisms known as "slow mildewing" phenotype or by environmental factors. (Ouchi et al. 1974; Stenzel et al. 1985; Slovakova 1991; Matsuda et al. 1994; Duggal et al. 2000). Studying possible gene effects later in pathogenesis requires the development of tools for high-throughput screening of altered hyphal growth phenotypes. This represents a new challenge to microscopic phenotyping platforms. To meet this challenge we decided to develop the HyphArea software tool that will allow an automated segmentation of microscopic images of growing fungal colonies acquired by different illumination techniques (bright field, fluorescence). The current version of the software is able to recognize fungal structures and to generate quantitative data with good correlation to manually acquired data (Baum et al. 2011).

Compared to the haustoria-screening platform outlined above, the analysis of hyphal structures described as follows does not need a complex hardware setup. The HyphArea tool (Baum et al. 2011) works on single microscope images that can be

taken either manually or by means of an automated screening table using the bundled control software. It works as simple as selecting the folder where the desired images are stored, followed by the selection of the host-pathogen system and the imaging protocol. Thereupon, the program segments the images, detects the colonies and calculates features like area or hull curve for each colony. The HyphArea tool is built to be extensible towards the host-pathogen system, the imaging protocol, and the morphological key values or properties.

5.2.3.3 *B. graminis* Hyphal Growth Measurement

The image data (Fig. 5.8a) is normalized with regard to magnification, scaling, and resolution. Furthermore, due to histogram operations like stretching and equalization, the intensity information is contrast-enhanced.

For bright field images possessing high color information, an additional color-space analysis improves the contrast between pathogen and background (Seiffert and Schweizer 2005; Fig. 5.8b), and a filtering reduces the variations of illumination between the single regions within the image (Fig. 5.8c). This method is called top-hat filtering and was described by (Gonzalez et al. 2004). Using a threshold-based segmentation with low threshold, the image is binarized (Fig. 5.8d). After an opening operation (Fig. 5.8e) reduces the noise, a first basic classification removes objects that are smaller than non-germinated spores (Fig. 5.8f). A morphological closing concatenates edges with low distance with each other, so that hyphal colonies are merged into one object (Fig. 5.8g). These closed objects or merged colonies are used as regions of interest (ROI). Based on their positions, the ROIs are cropped from the original image. Afterwards a rather accurate segmentation can be performed on these cropped images. Using histogram analysis and the size of the ROI, a threshold that leads to a satisfactory segmentation result can be calculated. All segmented ROIs are labeled as one unit and by using their former position superimposed into a new empty image with the size of the original one (Fig. 5.8h).

5.2.3.4 Recognition of *Rhynchosporium secalis* in Barley

HyphArea was designed as modular and adaptive concept for microscopic phenotyping of interaction of plants with filamentous fungi. The scope of the software extends beyond the barley—powdery mildew model system, and it is not fixed to particular microscopic techniques. In Baum et al. (2011) HyphArea was successfully applied for segmentation of GFP-tagged *R. secalis* in fluorescence microscopy images. Particular challenge in this case represented the two-dimensional subcuticular growth of the fungus, which was solved by using a GFP-tagged fungal strain and fluorescence microscopy.

Similar to the pre-processing of the bright-field imaging data, the contrast and brightness of fluorescence images are normalized by histogram stretching. Subsequently, due to median filtering, a reduction of the image noise can be achieved.

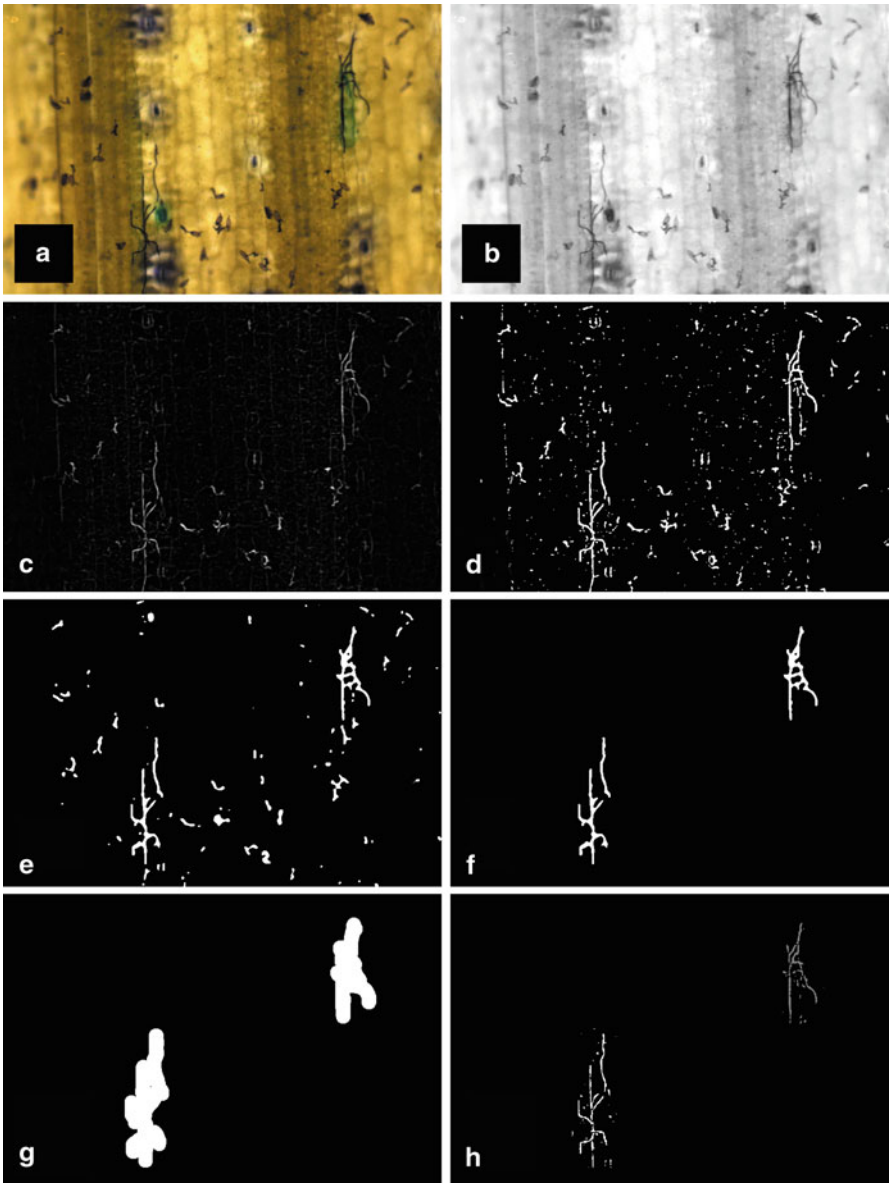

Fig. 5.8 Exemplification of each step in the segmentation chain. Showcase for an original image (**a**) containing a part of a barley leaf and two colonies of *B. graminis*. (**b**) Shows a color-transformed and gray-scaled image of (**a**). (**c**) Shows the result after tophat-transformation and (**d**) after the threshold-based segmentation. A morphological opening (**e**) combined with a preliminary classification (**e**) removes all objects smaller than ungerminated spores which results in the region of interests (**f**, after closing **g**). (**h**) Shows the result of the histogram and size-based segmentation of the ROIs

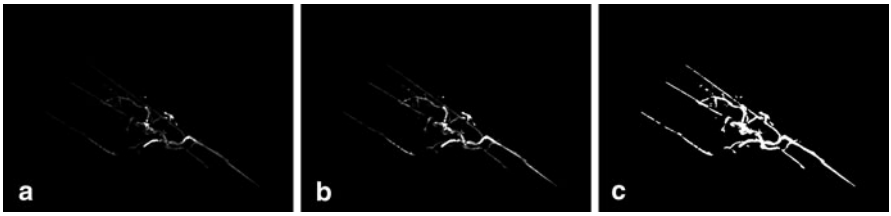

Fig. 5.9 (**a**) Shows an original partially cropped fluorescence microscope image of *R. secalis*. (**b**) Shows the result of the histogram-stretching and (**c**) of the threshold-based segmentation

Finally, a pixel-based segmentation is performed on these images. Both the histogram stretching and the result of the binarization are shown in Fig. 5.9.

Although *R. secalis*—barley interactions is not suitable for transient assay due to non-cell specific necrotrophic nature of interaction HyphArea was shown to be valuable for precise phenotyping of transgenic events and fungal mutants (Kirsten et al. 2012).

5.2.4 Detailed Analysis and Gene Validation

Both host and pathogen can exhibit a relatively high degree of variability in their response and virulence, respectively, which may cause problems in interpreting the results. Researchers usually resolve such problems by accumulating data from a high-enough number of biological replicates and by applying appropriate statistical methods. Typically we perform a first round screening to select candidates for repetitions based on pre-defined thresholds of TIGS effects. At the end of a screening all selected candidates must be tested in at least five independent experiments each, which will allow statistical calculation for the removal of outliers and for mean-value comparisons between RNAi construct and empty-vector controls.

The next step in analysis of the candidates is to confirm the specificity of the effect, which is one of the serious caveats of RNAi technology. A preliminary test can be made *in silico* by searching for potential off-targets of the used RNAi trigger sequence. Corresponding off-target prediction tools where developed recently (http://labtools.ipk-gatersleben.de/). However the trustworthiness of such tools depends on the completeness of the transcript-sequence database that is searched for finding off-targets. Unfortunately several important crop species such as barley or wheat do not yet have highly completed dataset of expressed sequences.

Another way to confirm the target-specificity of an RNAi construct is to design a number of related constructs targeting non-overlapping regions on the same mRNA sequence. The effect of the majority of constructs must be consistent.

Probably to most convincing proof of a specific RNAi effect can be achieved in RNAi-rescue experiments by overexpressing the target gene in the presence of the RNAi construct. However, an overexpression construct based on the same sequence

as the RNAi target will be silenced as well. This problem is circumvented by using synthetic overexpression constructs enriched in silent point mutations (replacing redundant codons), which renders them immune to the tested RNAi construct while being still translated to the wild type amino acid sequence (Dong et al. 2006; Nowara et al. 2010). If the RNAi effect was target-specific the RNAi-rescue construct should be able to complement for the function of the silenced gene. We are using this method routinely for the confirmation of target-specific RNAi effects of the most important candidate genes.

TILLING (Targeting Induced Local Lesions IN Genomes) was originally designed as a discovery platform for functional genomics and reverse genetic method, but it soon was developed further to a more versatile tool. Here we will mention the role of TILLING as confirmation strategy for results obtained by TIGS provided that off-target prediction clearly suggests one specific gene family member as being responsible for the observed phenotype. Existing TILLING mutant collections can be searched specifically for mutated alleles of the putative genes producing the TIGS phenotype. Co-segregation of the phenotype with mutation in the particular gene will provide strong evidence supporting the involvement of the gene in formation of the phenotype.

The Transcription activator-like (TAL) effectors (Boch et al. 2009) are a newly described class of specific DNA binding proteins that can be fused to a nuclease to achieve highly specific cleavage of DNA. The so called TALENs (TAL effector nucleases) can be used to disrupt or replace specific genes *in vivo* (Cermak et al. 2011; Li et al. 2012). TALENs may provide a straight forward way to confirm TIGS results in a similar but more specific manner than TILLING.

A high-throughput TIGS screen is able to identify among hundreds to thousands of potential candidates those genes having a high probability to be functionally relevant for the studied phenotype. However, the TIGS phenotype remains restricted to single epidermal cells and is the result of an "induced" (triggered by bombardment) but very strong silencing. Therefore independent confirmation and validation of TIGS effects in transgenic plants remains indispensable for the top candidates. During the last years we have generated and tested transgenic lines silencing and/or overexpressing about 30 genes and have so far found a good correlation of the results from the transient and stable transgenic approaches (Douchkov et al. unpublished).

5.2.5 TIGS Screening Results

Until present we have tested over 1300 RNAi constructs in TIGS screens for braking down nonhost resistance or for altering race-nonspecific host resistance, reflected by the formation of the initial fungal haustorium. The results are summarized in Table 5.1.

The test candidates were selected by several different criteria, which appeared to have an influence on the discovery rate, i.e. the percentage of RNAi targets exhibiting

Table 5.1 Summary of TIGS-screening results. Candidate barley gene groups were tested for an effect on haustorium formation of either the nonadapted pathogen wheat powdery mildew fungus *B. graminis* f.sp. *tritici* (Bgt) or of the adapted pathogen *B. graminis* f.sp. *hordei* (Bgh). A large proportion of RNAi constructs was tested in both (Bgt and Bgh) screens. The number and percent (from the initial selection) of constructs that show a statistically significant effect (one-sample t-test, 1-sided $p < 0.05$) after at least five independent experiments are listed in the columns "TIGS 1-sided significant". Selection criterion categories are described in details in the text

Screen	Selection criterion	Family	Entry No. constr	TIGS 1-sided sign.	
				No. constr	%
Bgh	Random	Multiple	88	4	4.5
Bgh	Regulated	Multiple	377	37	8.5
Bgh	Gene family	ABC transporter	94	8	9.8
Bgh		CSL	28	5	17.9
Bgh		PRR	38	8	21.1
Bgh		Peroxidase	75	6	8.0
Bgh		Proteasome lid	62	3	4.8
Bgh		RING E3-ligase	37	4	10.8
Bgh		Sweet transporter	12	4	33.3
Bgh		WKRY TF	64	7	10.9
Bgh	QTL conf. int	Multiple	128	7	5.5
Bgh	*Total*		*1003*	*93*	*9.3*
Bgt	Random	Multiple	224	3	1.3
Bgt	Regulated	Multiple	468	7	1.5
Bgt	*Total*		*692*	*10*	*1.4*
Bgh	*Fungal genes*	*Multiple*	*56*	*11*	*19.6*
Bgh and Bgt screens nonredundant			*1337*	*109*	

a significantly reproducible phenotypic effect. Two groups of randomly selected genes were used to provide a reference for the discovery rate of nonbiased selection.

5.2.6 Transcript Regulation as Selection Criterion

Selection based on transcript profiling is a relatively straight forward way to select for candidate genes relevant for the plant defense or susceptibility. In our work we have tested 377 pathogen-regulated genes in compatible interaction screens (barley/Bgh), and 468 in non-compatible interaction (barley/Bgt), ending up with 37 resp. ten genes that significantly influence the plant-pathogen interaction. Since the genes are selected "blindly" with no attention to their function or known properties, there is a high chance to end up with new and highly interesting candidates never discussed in relation to plant defense before.

5.2.7 Selection Based on Function of the Genes

The output of the knowledge-based selection of genes can vary considerably depending on the selected group. There are some extremely "hot" gene families like the only recently described SWEET sugar efflux transporters, which yielded the highest percentage of genes with an effect on haustoria formation. The observed high percentage of genes with significant TIGS effect may however still be an arbitrary result due to insufficient number of family members tested. In contrast, other gene families frequently discussed as defense- or susceptibility related such as the secreted class III peroxidases yielded rather low number of candidates, which might reflect a high degree of functional redundancy between gene paralogues. An apparently promising group of genes encodes cellulose-synthase like proteins, which has not been in the focus of plant pathologists until present.

5.2.8 Co-localization with Resistance QTL

High-throughput TIGS can be a straight forward way for discovering genes that underlie a resistance QTL as far as the corresponding QTL interval can be populated with a complete- enough list of mapped genes. In an attempt to achieve gene-landing in a confidence interval of a robust meta-QTL for resistance to Bgh on barley chromosome 5H (Schweizer and Stein 2011) we performed a TIGS screen of 128 co-localizing unigenes. This revealed four candidates with a significant Bgh-interaction phenotype upon TIGS (Spies et al., manuscript in preparation). In most cases just one gene would be causative for a given QTL. Therefore, the unexpectedly high number of candidates with a TIGS phenotype at the 5H QTL suggests that not all phenomics-based hits guide us directly to the right gene(s). Instead, false-positive results are also possible such as *per se* toxic RNAi effects of e.g. housekeeping genes that are likely to affect the delicate interaction between barley epidermal cells and the obligate biotrophic fungus Bgh requiring living cells for its development. Nevertheless, the identified shortlist helps to focus on few candidate genes in follow-up approaches such as transgenic plants or TILLING mutants.

5.2.9 Fungal Genes

Silencing of pathogen genes via HIGS (cf. 1.1.1.2) allows for the first time to functionally address genes of the obligate biotrophic fungus *B. graminis* that has been resilient to genetic transformation so far. We have tested over 50 fungal RNAi targets by HIGS and nearly 20 % of them exhibited a phenotypic effect. Such an exceptionally high discovery rate was observed otherwise only by silencing of some very specific groups of plant genes such as pathogen up-regulated RLKs or the SWEET sugar transporters. Specific silencing of these genes is not expected to produce any negative side effect on plant-encoded (off)-targets, which makes HIGS an attractive novel approach for engineering durable resistance in barley and other crop plants.

5.2.10 Random Selection of Candidates

Groups of 88 (in barley/Bgh screen), resp. 224 (barley/Bgt screen) randomly selected genes were included to the microphenomics pipeline to provide an idea for the probability of discovering genes by chance (Table 5.1). As expected the discovery rate in these two groups is lower than those of the biased selection groups, yet providing significant number of genes with effect. Authors are convinced that a systematic nonbiased screen of the entire gene space of barley will reveal exiting new functions of genes when such enormous screen becomes feasible.

5.3 Conclusions and Perspectives

Although minimal tiling-path sequencing of the barley genome is still in progress, we already now have a fairly good picture of its gene space encoding approximately 30,000 unique proteins due to large EST collections and several shotgun sequencing approaches of the whole genome and of sorted chromosome arms (Mayer et al. 2011, 2012). This leaves us with the same challenge of assigning specific biological functions to all these proposed genes as other research communities working in advanced model plants such as *A. thaliana* or rice. The only way to obtain such information is by large-scale projects of precision phenotyping, which can be conducted at whole-plant level in the field or in dedicated phenotyping facilities in the greenhouse, or by using specific (microscopic) tissue- or single cell assays. However, the development of plant-phenotyping platforms has been widely recognized as a major challenge and bottleneck because the possibilities of parallelization and high throughput become rapidly limited if one moves beyond the well-established genomic profiling approaches. Preferred plant materials for phenotyping screens may be deeply-genotyped genebank collections, mutant-backcross populations and transgenic plants. Expected results from this effort will allow for the initial discovery of gene- or allele-trait relations as well as validation and detailed description of gene function. The high-throughput microscopic phenotyping in barley leaves attacked by pathogenic fungi described here addresses a serious problem of yield security because these parasites cause most severe damage worldwide besides abiotic stresses such as drought (Reynolds and Borlaug 2006).

In order to be useful for large-scale screens by plant pathologists the phenotyping platform should fulfill several criteria: First, it should be partly or completely robotized at the level of microscopy, which was achieved for haustoria formation and hyphal growth of the powdery mildew fungus *B. graminis*. The technical set-up is based on commercially available hardware components and on proprietary software that is available under MTA conditions for non-commercial use. Still, further optimizations of the front end of the described phenotyping platform is required and planned, especially in terms of robustness to suboptimal image quality, a problem that occurred only after extensive use under daily working conditions. With the amount of parallelization established at present the haustoria phenotyping tool allows for 40

and 80 genes and genotypes, respectively, to be tested within 1 week. The calculated throughput of the HyphArea tool for hyphal growth will be approximately 40 and 400 for genes and genotypes, respectively. The reason for lower throughput to test (candidate) genes lies in the more laborious up-front work consisting in construct generation for TIGS or transient overexpression followed by bombardment of leaf segments.

So far, several screens have been performed for genes of barley or *B. graminis* affecting nonhost resistance or race-nonspecific host resistance. Two of these screens have used the automated phenotyping platform for haustoria formation. In the near future we intend to start up automated screens for hyphal growth rate, too, which will most likely expose us to similar problems of robustness of the platform as encountered for haustoria quantification. Nevertheless, the haustoria-detection tool has already contributed to the discovery of 94 barley genes with significant phenotypic effect upon silencing, and a selection of those is under further investigation in transgenic plants as well as forward-genetic approaches such as SNP-trait association in genebank collections and marker-assisted backcrossing into German elite material of spring barley. Thereby, we have started to fill an important gap between barley genome-sequence and transcript-profiling data, and meanwhile well-established approaches of association genetics and targeted introgression of potentially valuable alleles for durable resistance of barley to its major pathogens.

Acknowledgements We would like to tank to Aura Navarro-Quezada and Wolfgang Knogge from the Leibniz-Institute of Plant Biochemistry, Department of Stress and Developmental Biology, Weinberg 3, 06120 Halle (Saale), Germany for providing the *R. secalis* fluorescence microscopy images. We are especially grateful to Stefanie Lück and Gabriele Brantin from IPK for the excellent technical work. Work of own research was supported by IPK (to P.S.), the German Ministry of Education and Research BMBF (to P.S. and U.S.), and by BASF Plant Science GmbH (to P.S.).

References

Baum T, Navarro-Quezada A, Knogge W et al (2011) HyphArea-Automated analysis of spatiotemporal fungal patterns. J Plant Physiol 168:72–78

Bauriegel E, Giebel A, Herppich WB (2011) Hyperspectral and chlorophyll fluorescence imaging to analyse the impact of fusarium culmorum on the photosynthetic integrity of infected wheat ears. Sensors 11:3765–3779

Beyer A, Hollunder J, Nasheuer HP et al (2004) Post-transcriptional expression regulation in the yeast *Saccharomyces cerevisiae* on a genomic scale. Mol Cell Proteomics 3:1083–1092

Bhullar NK, Street K, Mackay M et al (2009) Unlocking wheat genetic resources for the molecular identification of previously undescribed functional alleles at the Pm3 resistance locus. P Natl Acad Sci USA 106:9519–9524

Boch J, Scholze H, Schornack S et al (2009) Breaking the code of DNA binding specificity of TAL-Type III effectors. Science 326:1509–1512

Brenner S, Johnson M, Bridgham J et al (2000) Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. Nat Biotechnol 18:630–634

Caldo RA, Nettleton D, Wise RP (2004) Interaction-dependent gene expression in Mla-specified response to barley powdery mildew. Plant Cell 16:2514–2528

Cermak T, Doyle EL, Christian M et al (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting (vol 39, pg e82, 2011). Nucleic Acids Res 39:7879–7879

Chinchilla D, Zipfel C, Robatzek S et al (2007) A flagellin-induced complex of the receptor FLS2 and BAK1 initiates plant defence. Nature 448:497–412

Close, TJ; Bhat, PR; Lonardi, S; Wu, YH; Rostoks, N; Ramsay, L; Druka, A; Stein, N; Svensson, JT; Wanamaker, S; Bozdag, S; Roose, ML; Moscou, MJ; Chao, SAM; Varshney, RK; Szucs, P; Sato, K; Hayes, PM; Matthews, DE; Kleinhofs, A; Muehlbauer, GJ; DeYoung, J; Marshall, DF; Madishetty, K; Fenton, RD; Condamine, P; Graner, A; Waugh, R (2009) Development and implementation of high-throughput SNP genotyping in barley. BMC GENOMICS 10:582

Close TJ, Wanamaker SI, Caldo RA et al (2004) A new resource for cereal genomics: 22K barley GeneChip comes of age. Plant Physiol 134:960–968

Crouzet J, Trombik T, Fraysse AS et al (2006) Organization and function of the plant pleiotropic drug resistance ABC transporter family. Febs Letters 580:1123–1130

Dammer KH, Moller B, Rodemann B et al (2011) Detection of head blight (Fusarium ssp.) in winter wheat by color and multispectral image analyses. Crop Protection 30:420–428

De Coninck BMA, Amand O, Delaure SL et al (2012) The use of digital image analysis and real-time PCR fine-tunes bioassays for quantification of Cercospora leaf spot disease in sugar beet breeding. Plant Pathol 61:76–84

Dong WB, Nowara D, Schweizer P (2006) Protein polyubiquitination plays a role in basal host resistance of barley. Plant Cell 18:3321–3331

Douchkov D, Nowara D, Zierold U et al (2005) A high-throughput gene-silencing system for the functional assessment of defense-related genes in barley epidermal cells. Mol Plant Microbe In 18:755–761

Druka A, Franckowiak J, Lundqvist U et al (2011) Genetic dissection of barley morphology and development. Plant Physiol 155:617–627

Duggal V, Jellis GJ, Hollins TW et al (2000) Resistance to powdery mildew in mutant lines of the susceptible wheat cultivar Hobbit 'sib'. Plant Pathol 49:468–476

Errampalli D, Patton D, Castle L et al (1991) Embryonic Lethals and T-DNA Insertional Mutagenesis in *Arabidopsis*. Plant Cell 3:149–157

Eulgem T, Somssich IE (2007) Networks of WRKY transcription factors in defense signaling. Curr Opin Plant Biol 10:366–371

Eulgem T, Rushton PJ, Robatzek S et al (2000) The WRKY superfamily of plant transcription factors. Trends Plant Sci 5:199–206

Frommer WB, Chen LQ, Hou BH et al (2010) Sugar transporters for intercellular exchange and nutrition of pathogens. Nature 468:527–199

Gomez-Gomez L, Bauer Z, Boller T (2001) Both the extracellular leucine-rich repeat domain and the kinase activity of FLS2 are required for flagellin binding and signaling in *Arabidopsis*. Plant Cell 13:1155–1163

Gonzalez R, Woods R, Eddins S (2004) Digital image processing using MATLAB. Prentice Hall, Upper Saddle River, NJ

Gonzalez-Lamothe R, Tsitsigiannis DI, Ludwig AA et al (2006) The U-Box protein CMPG1 is required for efficient activation of defense mechanisms triggered by multiple resistance genes in tobacco and tomato. Plant Cell 18:1067–1083

Heese A, Hann DR, Gimenez-Ibanez S et al (2007) The receptor-like kinase SERK3/BAK1 is a central regulator of innate immunity in plants. P Natl Acad Sci USA 104:12217–12222

Ihlow A, Schweizer P, Seiffert U (2008) A high-throughput screening system for barley/powdery mildew interactions based on automated analysis of light micrographs. BMC Plant Biol 8:6

Iizasa E, Mitsutomi M, Nagano Y (2010) Direct Binding of a Plant LysM Receptor-like Kinase, LysM RLK1/CERK1, to Chitin in Vitro. J Biol Chem 285:2996–3004

Jasinski M, Stukkens Y, Degand H et al (2001) A plant plasma membrane ATP binding cassette-type transporter is involved in antifungal terpenoid secretion. Plant Cell 13:1095–1107

Kirsten S, Navarro-Quezada A, Penselin D et al (2012) Necrosis-Inducing Proteins of *Rhynchosporium commune*, Effectors in Quantitative Disease Resistance. Mol Plant Microbe In 25:1314–1325

Kobae Y, Sekino T, Yoshioka H et al (2006) Loss of AtPDR8, a plasma membrane ABC transporter of *Arabidopsis thaliana*, causes hypersensitive cell death upon pathogen infection. Plant Cell Physiol 47:309–318

Kumlehn J, Serazetdinova L, Hensel G et al (2006) Genetic transformation of barley (*Hordeum vulgare* L.) via infection of androgenetic pollen cultures with *Agrobacterium tumefaciens*. Plant Biotechnol J 4:251–261

Kuromori T, Takahashi S, Kondou Y et al (2009) Phenome analysis in plant species using loss-of-function and gain-of-function mutants. Plant Cell Physiol 50:1215–1231

Lacomme C, Hrubikova K, Hein I (2003) Enhancement of virus-induced gene silencing through viral-based production of inverted-repeats. Plant J 34:543–553

Lagudah ES, Krattinger SG, Spielmeyer W et al (2009) A putative ABC transporter confers durable resistance to multiple fungal pathogens in wheat. Science 323:1360–1363

Li T, Liu B, Spalding MH et al (2012) High-efficiency TALEN-based gene editing produces disease-resistant rice. Nat Biotechnol 30:390–392

Lipka V, Gimenez-Ibanez S, Hann DR et al (2009) AvrPtoB targets the LysM receptor kinase CERK1 to promote bacterial virulence on plants. Curr Biol 19:423–429

Lipka V, Petutschnig EK, Jones AME et al (2010) The lysin motif receptor-like kinase (LysM-RLK) CERK1 is a major chitin-binding protein in *Arabidopsis thaliana* and subject to chitin-induced phosphorylation. J Biol Chem 285:28902–28911

Lu D, Wu S, He P et al (2010) Phosphorylation of receptor-like cytoplasmic kinases by bacterial flagellin. Plant Signal Behav 5:598–600

Maeda I, Kohara Y, Yamamoto M et al (2001) Large-scale analysis of gene function in *Caenorhabditis elegans* by high-throughput RNAi. Curr Biol 11:171–176

Mahlein AK, Oerke EC, Steiner U et al (2012a) Recent advances in sensing plant diseases for precision crop protection. Eur J Plant Pathol 133:197–209

Mahlein AK, Steiner U, Hillnhutter C et al (2012b) Hyperspectral imaging for small-scale analysis of symptoms caused by different sugar beet diseases. Plant Methods 8:3

Martinoia E, Klein M, Geisler M et al (2002) Multifunctionality of plant ABC transporters - more than just detoxifiers. Planta 214:345–355

Matsuda Y, Toyoda H, Morita M et al (1994) A novel method for *In Situ* hybridization in fungal cells based on pricking micro–injection of photobiotin labelled probes. J Phytopathol 141:133–142

Mayer KFX, Martis M, Hedley PE et al (2011) Unlocking the barley genome by chromosomal and comparative genomics. The Plant Cell Online 23:1249–1263

Mayer KFX, Waugh R, Langridge P et al (2012) A physical, genetic and functional sequence assembly of the barley genome. Nature 491:711–716

Mazzucotelli E, Belloni S, Marone D et al (2006) The e3 ubiquitin ligase gene family in plants: regulation by degradation. Curr Genomics 7:509–522

Nowara D, Gay A, Lacomme C et al (2010) HIGS: host-induced gene silencing in the obligate biotrophic fungal pathogen *Blumeria graminis*. Plant Cell 22:3130–3141

Ouchi S, Oku H, Hibino C et al (1974) Induction of accessibility and resistance in leaves of barley by some races of *Erysiphe graminis*. J Phytopathol 79:24–34

Panstruga R, Schulze-Lefert P (2002) Live and let live: insights into powdery mildew disease and resistance. Mol Plant Pathol 3:495–502

Pedreno MA, Almagro L, Ros LVG et al (2009) Class III peroxidases in plant defence reactions. J Exp Bot 60:377–390

Peressotti E, Duchene E, Merdinoglu D et al (2011) A semi-automatic non-destructive method to quantify grapevine downy mildew sporulation. J Microbiol Meth 84:265–271

Reynolds MP, Borlaug NE (2006) Impacts of breeding on international collaborative wheat improvement. J Agr Sci 144:3–17

Rushton PJ, Somssich IE, Ringler P et al (2010) WRKY transcription factors. Trends Plant Sci 15:247–258

Schweizer P, Stein N (2011) Large-scale data integration reveals colocalization of gene functional groups with meta-QTL for multiple disease resistance in barley. Mol Plant Microbe Interact 24:1492–1501

Schweizer P, Gees R, Mosinger E (1993) Effect of jasmonic acid on the interaction of Barley (Hordeum vulgare L.) with the powdery mildew *Erysiphe graminis* f. sp. *hordei*. Plant Physiol 102:503–511

Schweizer P, Christoffel A, Dudler R (1999) Transient expression of members of the germin-like gene family in epidermal cells of wheat confers disease resistance. Plant J 20:540–552

Schweizer P, Pokorny J, Schulze-Lefert P et al (2000) Double-stranded RNA interferes with gene function at the single-cell level in cereals. Plant J 24:895–903

Seiffert U, Schweizer P (2005) A pattern recognition tool for quantitative analysis of in planta hyphal growth of powdery mildew fungi. Mol Plant Microbe In 18:906–912

Shen LH, Gong J, Caldo RA et al (2005) BarleyBase—an expression profiling database for plant genornics. Nucleic Acids Res 33:D614–D618

Shibuya N, Miya A, Albert P et al (2007) CERK1, a LysM receptor kinase, is essential for chitin elicitor signaling in *Arabidopsis*. P Natl Acad Sci USA 104:19613–19618

Shirasu K, Trujillo M, Ichimura K et al (2008) Negative regulation of PAMP-triggered immunity by an E3 ubiquitin ligase triplet in *Arabidopsis*. Curr Biol 18:1396–1401

Shiu SH, Bleecker AB (2003) Expansion of the receptor-like kinase/Pelle gene family and receptor-like proteins in *Arabidopsis*. Plant Physiol 132:530–543

Shiu SH, Lehti-Shiu MD, Zou C et al (2009) evolutionary history and stress regulation of plant receptor-like kinase/pelle genes. Plant Physiol 150:12–26

Slovakova L (1991) Induced resistance of barley plants against powdery mildew (*Erysiphe graminis* f.sp. *hordei* Marchal).1. Influence of inducers on primary infection. Biologia 46:737–744

Song WY, Wang GL, Chen LL et al (1995) A receptor kinase-like protein encoded by the rice disease resistance gene, Xa21. Science 270:1804–1806

Stein M, Dittgen J, Sanchez-Rodriguez C et al (2006) *Arabidopsis* PEN3/PDR8, an ATP binding cassette transporter, contributes to nonhost resistance to inappropriate pathogens that enter by direct penetration. Plant Cell 18:731–746

Stenzel K, Steiner U, Schonbeck F (1985) Effect of induced resistance on the efficiency of powdery mildew haustoria in wheat and barley. Physiol Plant Pathol 27:357–367

Tingay S, McElroy D, Kalla R et al (1997) *Agrobacterium tumefaciens*-mediated barley transformation. Plant J 11:1369–1376

Trujillo M, Shirasu K (2010) Ubiquitination in plant immunity. Curr Opin Plant Biol 13:402–408

Velculescu VE, Zhang L, Vogelstein B et al (1995) Serial analysis of gene-expression. Science 270:484–487

Wang GL, Song WY, Ruan DL et al (1996) The cloned gene, Xa21, confers resistance to multiple *Xanthomonas oryzae* pv. *oryzae* isolates in transgenic plants. Mol Plant Microbe In 9:850–855

Wang L, Li PH, Brutnell TP (2010) Exploring plant transcriptomes using ultra high-throughput sequencing. Brief Funct Genomics 9:118–128

Warringer J, Ericson E, Fernandez L et al (2003) Yeast phenomics on a genome-wide scale. Yeast 20:S338–S338

Wielopolska A, Townley H, Moore I et al (2005) A high-throughput inducible RNAi vector for plants. Plant Biotechnol J 3:583–590

Wijekoon CP, Goodwin PH, Hsiang T (2008) Quantifying fungal infection of plant leaves by digital image analysis using Scion Image software. J Microbiol Meth 74:94–101

Wise R, Caldo R, Hong L et al (2006) PLEXdb: a unifited expression prolfiling database for plants and plant pathogens. Phytopathology 96:S161–S161

Wise RP, Caldo RA, Hong L et al (2007) Barleybase/PLEXdb. A unified expression profiling database for plants and plant pathogens. Methods Mol Biol 406:347–363

Wise RP, Lauter N, Szabo LJ et al (2009) Genomics of biotic interactions in the *Triticeae*. Genetics and genomics of the triticeae. C. F. G. J. muehlbauer. Springer 7:559–589

Yang CW, Gonzalez-Lamothe R, Ewan RA et al (2006) The E3 ubiquitin ligase activity of *Arabidopsis* PLANT U-BOX17 and its functional tobacco homolog ACRE276 are required for cell death and defense. Plant Cell 18:1084–1098

Yazaki K, Shitan N, Sugiyama A et al (2009) Cell and molecular biology of ATP-binding cassette proteins in plants. Int Rev Cell Mol Biol 276:263–299

Zellerhoff N, Himmelbach A, Dong WB et al (2010) Nonhost resistance of barley to different fungal pathogens is associated with largely distinct, quantitative transcriptional responses. Plant Physiol 152:2053–2066

Zhang HN, Sreenivasulu N, Weschke W et al (2004) Large-scale analysis of the barley transcriptome based on expressed sequence tags. Plant J 40:276–290

Zhang J, Li W, Xiang T et al (2010) Receptor-like cytoplasmic kinases integrate signaling from multiple plant immune receptors and are targeted by a *Pseudomonas syringae* effector. Cell Host Microbe 7:290–301

Zhao T, Palotta M, Langridge P et al (2006) Mapped Ds/T-DNA launch pads for functional genomics in barley. Plant J 47:811–826

Zhu LH, Gan QA, Bai H et al (2011) Transcriptional characteristics of xa21-mediated defense responses in rice. J Integr Plant Biol 53:300–311

Zierold U, Scholz U, Schweizer P (2005) Transcriptome analysis of mlo-mediated resistance in the epidermis of barley. Mol Plant Pathol 6:139–151

Zipfel C, Kunze G, Chinchilla D et al (2006) Perception of the bacterial PAMP EF-Tu by the receptor EFR restricts *Agrobacterium*-mediated transformation. Cell 125:749–760

Chapter 6
Genomics of Low-Temperature Tolerance for an Increased Sustainability of Wheat and Barley Production

N. Pecchioni, K. Kosová, P. Vítámvás, I.T. Prášil, J.A. Milc, E. Francia, Z. Gulyás, G. Kocsy and G. Galiba

Abstract Stability of high yields in a changing environment becomes the main aim of the future wheat and barley breeding, oriented towards development of frost-tolerant winter and facultative cultivars together with careful selection of growth cycle adaptation and drought tolerance. Since low temperature signal influences both the cold acclimation and vernalization processes the interaction between *VRN* gene expression and freezing tolerance (FT) is discussed. Recent advances in global expression changes driven by cold are reviewed in view of the immense progress in high throughput technological platforms. Different signal transduction pathways in which several transcription factors play an important role regulating the expression of whole sets of genes are presented, including CBF-regulated and CBF-independent hubs. The knowledge acquired from genomics and transcriptome analysis has been then complemented by the description of metabolomics and proteomic approaches to help unraveling the molecular changes that occur under cold stress in the cereal plants. Finally, it is surveyed the great importance of stable and well-characterized genetic resources for future breeding for FT, that could switch from marker-assisted to genomics-assisted selection.

G. Galiba (✉) · Z. Gulyás · G. Kocsy
Agricultural Institute, Centre for Agricultural Research,
Hungarian Academy of Sciences, 2462 Martonvásár, P.O.B. 19, Hungary
e-mail: galiba.gabor@agrar.mta.hu

N. Pecchioni · J.A. Milc · E. Francia
Department of Life Sciences, University of Modena and Reggio Emilia,
Via Amendola, 2—Padiglione Besta, 42122 Reggio Emilia, Italy

K. Kosová · P. Vítámvás · I.T. Prášil
Department of Genetics and Plant Breeding, Crop Research Institute, Drnovská
Street 507, Prague 6—Ruzyně, 16106 Prague, the Czech Republic

G. Galiba
Research Institute of Chemical and Process Engineering, Faculty of Information
Technology, University of Pannonia, 8200 Veszprém, P.O.B. 158, Hungary

R. Tuberosa et al. (eds.), *Genomics of Plant Genetic Resources,*
DOI 10.1007/978-94-007-7575-6_6,

6.1 Introduction

Global climate change, causing more dramatic and erratic temperature shifts and rainfall events, has increased the importance of an old problem. Due to variations in environmental stresses such as drought, flooding, heat, salinity, low temperature, etc., the increasing variation in crop yield from year to year throughout the world posses a serious problem for agricultural production and the market of agricultural commodities. Although farmers in the northern hemisphere face the possible risk of winter damage, they still grow winter cultivars. The explanation is very simple: winter cereals are sown in the fall, and provided they have adequate tolerance to survive freezing temperatures in winter, because of their longer growing period they usually have higher yield potential than spring varieties, planted later in the spring. In addition, they are usually earlier in flowering and maturity than their spring counterparts, thus they escape the summer heat and drought, the other significant abiotic stress factors which may reduce yield significantly. Maintaining yield stability while enhancing productivity in the driest and coldest regions of cereal growing areas is becoming increasingly important for securing a level of food production sufficient to meet the needs of mankind. Stability of high yields in a rapidly changing environment becomes the main aim of stress biology, from a theoretical point of view, to be applied to future crop breeding. In this context, it seems unavoidable the trend towards an increasing adoption of frost-tolerant winter and facultative cultivars in temperate cereals, coupled with careful selection of growth cycle length and drought tolerance. Thus there are reasons for further efforts to improve cereal freezing tolerance, particularly among winter materials.

The ability of plants to survive freezing is based on the effectiveness of cold acclimation process (Thomashow 1999; Skinner 2009). Cold acclimation is a relatively slow, adaptive response during fall, when the temperature, day length and light intensity decrease gradually. All these factors are important for attaining genetically determined freezing tolerance (Sandre et al. 2011; Schoot and Rinne 2011). Apart from the cold acclimation, decreasing day length and an extended cold period influence the plant development also through the vernalization process mediated by *VRN* genes (Distelfeld et al. 2009; Greenup et al. 2009). Plants are able to cold acclimate only in the vegetative developmental phase, during which in most cereal growing areas wheat and barley experience frost. From the freezing tolerance point of view, the vegetative/reproductive transition phase is the most critical period. Following the vernalization saturation, when the shoot apex reaches the double ridge phase (floral initiation), the leaf initiation is terminated. This developmental stage is irreversible and reduces the ability of the plant to maintain enhanced freezing tolerance upon cold acclimation conditions (Fowler et al. 1996; Vítámvás and Prášil 2008; Galiba et al. 2009). Although cold stress in the reproductive phase is quite infrequent, it can occur due to cold winds (Reinheimer et al. 2004). In Australia, the predominant frost damage occurs from radiation-frost events in spring during the reproductive stages of barley and wheat. Reproductive tissue is very sensitive and can be damaged by small sub-zero temperatures drops ($\leq -2\,^{\circ}\mathrm{C}$) that can lead to spikelet sterility. Genetic studies performed in barley concluded that chromosomes

2H and 5H were responsible for reproductive freezing tolerance, and showed that winter habit alleles at the *VRN-H1* vernalization gene on 5H were linked in coupling with reproductive FT (Chen et al. 2009). Since low temperature influences the cold acclimation and vernalization processes, we will discuss the interaction between *VRN* genes and freezing tolerance.

Low temperature-specific gene expression is mediated by different parallel signal transduction pathways. Two main signalling pathways have been described in *Arabidopsis*; one is dependent on the involvement of the phytohormone abscisic acid (ABA), while the other is not (Zhang et al. 2004). Both pathways trigger the expression of a range of transcription factors that bind, among others, to the C-repeats or Dehydration-Responsive Element (CRT/DRE), ABA-Responsive Element (ABRE), and MYC/MYB Recognition Sequence (MYCR/MYBR) binding sites of their target genes, thus regulating the transcription of these genes. The altered expression of these downstream genes leads to cold acclimation and to an increased level of freezing tolerance (Chinnusamy et al. 2004). During cold acclimation many molecular processes are conserved in the dicot and monocot lineages (Nakashima et al. 2009). One of the best examples of this phenomenon is the central role of the C-repeat Binding Factor or Dehydration-Responsive Element Binding Factor (CBF/DREB1) regulon in the cold acclimation process both in model plants and crops. Accordingly, this review will discuss the role of the CBF regulon in the cold acclimation process in cereals. Since *CBF* genes are important but not the only regulators involved in the cold-response regulatory pathway, our current knowledge on other transcription factors and their target genes will also be described. In addition to transcriptomic analysis, metabolomics and proteomic approaches helped to highlight the molecular changes under cold stress in plants as well (Fig. 6.1).

FT is a polygenic trait affected by several genes located on different chromosomes. Since bread wheat (*Triticum aestivum* L.) is hexaploid, its vital genes are replicated. This allowed Sears (1953) to develop a series of nullisomic lines (i.e. lines lacking one of the normal chromosomal pairs) from Chinese Spring (a freezing-susceptible spring wheat). These lines served as the basis for developing intervarietal chromosome substitution lines, thus providing one of the best means of studying the genetic control of freezing tolerance.

In the last two decades, following the omics breakthrough, from genomics (Tuberosa et al. 2002) to the introduction of high-throughput technological platforms, there has been a shift from genetic, gene-by-gene, studies to genomic ones, in which sequence as well as expression variation are studied on a global scale. Since the pioneering works to develop suitable genetic materials used for genetic studies of FT (Galiba et al. 2009), a large amount of genomic data have been generated in wheat and barley, and two international sequencing initiatives led to the recent release of the first genomic frameworks of the wheat and barley genomes. In hexaploid wheat, Brenchley et al. (2012) used a whole-genome shotgun sequencing strategy to obtain a low-pass coverage of the genome. In barley, Mayer et al. (2012) reported an integrated and ordered physical, genetic and functional sequence resource of the cultivar Morex gene-space in a structured whole-genome context.

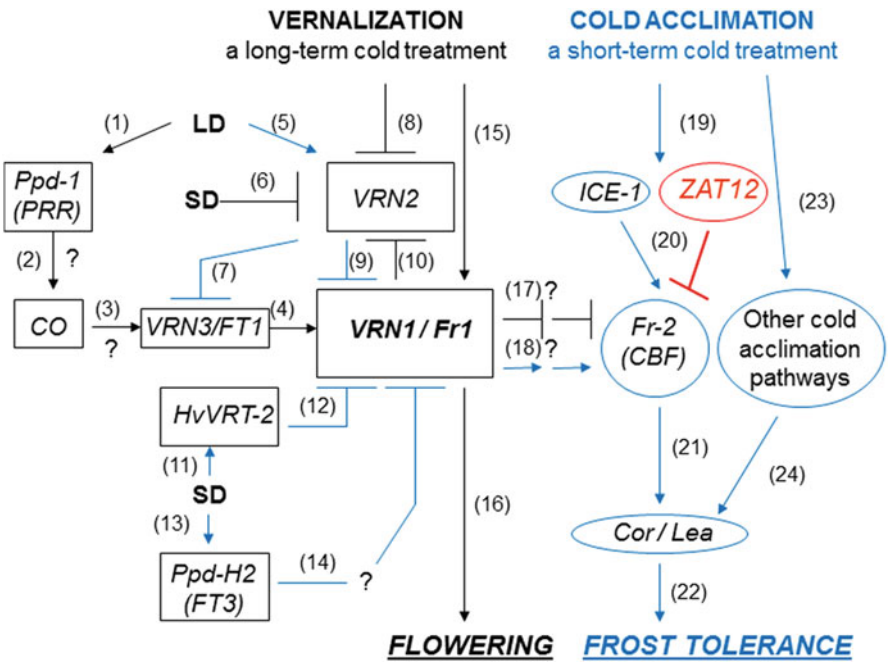

Fig. 6.1 Relationships between vernalization and FT in *Triticeae*

Descriptive scheme of the relationships between vernalization- and photoperiodically-regulated loci (their candidate genes, respectively) which control the transition to flowering and loci which regulate the development of FT upon cold acclimation (CA). (1) Laurie et al. (1994), (2) Turner et al. (2005), Beales et al. (2007), (3) Griffiths et al. (2003), Turner et al. (2005), (4) Yan et al. (2006), (5) Yan et al. (2004), Loukoianov et al. (2005), LD induces expression of *VRN2* gene, (6) Yan et al. (2004), (7) Yan et al. (2006), (8) Dubcovsky et al. (2006), Trevaskis et al. (2006), (9) when *vrn-1* is recessive and *Vrn-2* is dominant, von Zitzewitz et al. (2005), (10) possibly when *Vrn-1* is dominant or when *vrn-2* is recessive under prolonged LD treatment, Loukoianov et al. (2005), Trevaskis et al. (2006), (11) Kane et al. (2005), (12) Kane et al. (2007), (13) Laurie et al. (1995), (14) Faure et al. (2007), (15) vernalization induces expression of *VRN1* gene, Oliver et al. (2009), (16) expression of *VRN1* gene is necessary for flowering, Shitsukawa et al. (2007), (17) *VRN1* gene downregulates expression of *CBF* genes, but only under LD conditions, Danyluk et al. (2003), Kane et al. (2005), Kobayashi et al. (2005), Stockinger et al. (2007), Dhillon et al. (2010), (18) Vágújfalvi et al. (2000), Kobayashi et al. (2005), Stockinger et al. (2007), (19) cold induces expression of *ICE-1* and *ZAT12* genes, Zarka et al. (2003), Skinner et al. (2006), (20) Skinner et al. (2006), (21) Vágújfalvi et al. (2000), Choi et al. (2002), Kume et al. (2005), Skinner et al. (2005), Miller et al. (2006), (22) Houde et al. (1992), Crosatti et al. (1995), Vágújfalvi et al. (2000), Vítámvás et al. (2007), Kosová et al. (2008a),

(23) Thomashow (1999), (24) Thomashow (1999), Choi et al. (2002). Question marks indicate uncertain or unknown components of the regulatory pathways. Blue lines indicate signalling pathways active before vernalization, black lines indicate pathways active after vernalization fulfillment. Modified after Kosová et al. (2008b).

Moreover, knowledge acquired from *Triticeae* genomics is complemented by a deeper understanding of the proteome, metabolome, and more recently, epigenome. In a research environment where throughput and amount of data generated are no longer limiting, the availability of appropriate and well-characterized plant genetic resources will become a crucial issue for further knowledge advancements in stress biology. This article reviews how the best allelic variants for freezing tolerance can be combined in a single wheat or barley genotype.

6.2 "Omic" Approaches to Study Cold Acclimation

6.2.1 Transcriptomics

The effect of low temperature on the transcriptome was first investigated mainly in the model plant *Arabidopsis* (Seki et al. 2001; Fowler and Thomashow 2002; Nakashima and Yamaguchi-Shinozaki 2006) but later on, these studies were also extended to an increasing number of crop species including cereals (for review see Vij and Tyagi 2007, Sreenivasulu et al. 2007). Thus, cold-induced transcriptome changes were investigated in rice (Rabbani et al. 2003), wheat (Gulick et al. 2005; Herman et al. 2006) and barley (Svensson et al. 2006; Greenup et al. 2011). Among the cold-responsive genes, transcription factors play a pivotal role since they regulate the expression of whole sets of genes called "regulons" and therefore, they can switch on arrays of metabolic activities necessary for cold acclimation and freezing tolerance. Special attention was given to the CBF regulon controlled by the C-repeat binding factors (CBFs), also called dehydration-responsive element binding (DREB1) factors (Chinnusamy et al. 2006; Nakashima and Yamaguchi-Shinozaki 2006; Van Buskirk and Thomashow 2006). From a comparison of the expression of several *CBF* genes in *Triticeae* (rye, wheat, barley) it turned out that sample timing, induction temperature and light-related factors have significant effects on transcript levels, and must be considered in future transcriptomic studies involving functional characterization of low temperature-induced genes in cereals (Campoli et al. 2009). Although the expression of 10 *CBF* genes was compared in three rye genotypes with different freezing tolerance, it is unclear which of them have the greatest influence on FT and whether they have an additive effect. The regulation of cold acclimation by CBFs is described in more detail in Sect. 6.3.

CBF genes are arranged in clusters on the homeologous group 5 chromosomes of *Triticeae*, and coincide with *FR2* QTL for FT (Vágújfalvi et al. 2003, 2005; Miller et al. 2006; Baga et al. 2007; Francia et al. 2007; Galiba et al. 2009). Besides *Fr-A2*, *Fr-A1,* a vernalization gene, *Vrn-A1* and several additional genes involved in the response to low temperature and other stresses were localised on the long arm of

wheat chromosome 5A (Dubcovsky et al. 1995; Galiba et al. 1995; Cattivelli et al. 2002; Galiba 2002; Danyluk et al. 2003; Yan et al. 2004; Ramalingam et al. 2006), which has a major effect on freezing tolerance (Sutka 1994). The arrangement of loci and respective functions is very similar in the barley H genome (Francia et al. 2004, 2007; von Zitzewitz et al. 2005). *VRN* genes regulate the transition from the vegetative to the reproductive phase, and act as a master switch controlling the duration of the expression of low temperature-induced structural genes (Danyluk et al. 2003). Both changes in temperature and day length are involved in the control of cold-regulated (*COR*) genes by *VRN1*, since mutations in the *VRN-1* promoter, resulting in high transcript levels under both long and short days, lead to down-regulation of the *COR14b* gene under long days (LD) but not under short days (SD) (Dhillon et al. 2010). Several cold-responsive and chromosome 5A-regulated genes, including those ones of LEA proteins and antioxidants were determined by comparison of the transcript profile of a freezing-tolerant and a freezing-sensitive chromosome 5A substitution line during a 21-day-long hardening period (Kocsy et al. 2010).

Besides the *CBF* downstream cascade, other CBF-independent pathways involving ZAT12 or the homeodomain transcription factor HOS9 have been shown to control cold-regulated genes (Vogel et al. 2005; Chinnusamy et al. 2006, Nakashima and Yamaguchi-Shinozaki 2006) (Fig. 6.1). Additonal CBF-independent cold acclimation pathways were discovered by characterization of several mutants. Thus, mutations in ESKIMO1 protein having unknown function and in the transcriptional adaptor protein ADA2 resulted in constitutively increased freezing tolerance with simultaneous induction of genes distinct of CBF regulons and without influencing genes of this regulon (Vlachonasios et al. 2003; Xin et al. 2007). A detailed description of CBF-independent transcriptional regulation of cold acclimation can be also found in Sect. 6.3.

Comparison of transcript profile changes in genotypes with different freezing tolerance and vernalization requirement is a powerful tool for the identification of genes having a role both in the effective cold acclimation and in the vegetative to reproductive transition. While in most studies an abrupt decrease in temperature was used for cold acclimation (Monroy et al. 2007; Kocsy et al. 2010), in a recent experiment a gradual reduction of temperature was applied which is more similar to the natural acclimation process during autumn. In this case, the underlying rationale was that those genes having different cold induction threshold temperatures could have been discovered in this way (Winfield et al. 2009). The earlier induction of cold-responsive genes at higher temperatures in the freezing-tolerant wheat genotypes compared to the freezing-sensitive ones, which was shown in the case of *COR14b* gene (Vágújfalvi et al. 2003), ensures their more efficient cold acclimation and greater level of freezing tolerance. Both gradual (Winfield et al. 2010) and abrupt decrease in temperature (Monroy et al. 2007; Kocsy et al. 2010) induced the expression changes of more genes during long-term cold acclimation in the freezing-tolerant wheat genotypes than in the sensitive ones. During a short-term exposure to cold (2 d, 4 °C), Winfield et al. (2010) did not find a such a difference by separate investigation

of crowns and leaves, but other authors observed alterations in the level of more transcripts in the tolerant genotype whole shoots than in the sensitive one (1 d at 2 °C and 1 d at 4 °C) (Monroy et al. 2007; Kocsy et al. 2010). This contradiction may be due to the different plant organs used for the analysis in the two experimental systems.

The importance of separate investigation of leaves and crown was first supported by the results of Winfield et al. (2009), who interestingly found many genes which were affected by cold either only in the crown or only in the leaves. They investigated cold-induced changes in the transcriptome of wheat leading to the vegetative to generative transition identifying several MADS-box genes, as others related to the gibberellin pathway, which may play an important role in the onset of flowering. Another numerous set of cold-responsive genes was found common to both organs, including genes encoding DEAD-box RNA helicase, choline-phosphate cytidyltransferase and delta-1-pyrroline carboxylate synthetase (Ganeshan et al. 2011). However, the same authors, in agreement with the previous report, found more genes being affected only in one of the two organs than common ones; and emphasized how cold acclimation mechanisms were likely differently regulated in crowns and leaves.

Additional genes involved in the acclimation process were identified by transcriptome analysis in crowns of wheat subjected to subzero acclimation (−3 °C for 12–18 h) after the cultivation at non-freezing low temperature (Herman et al. 2006). These plants exhibited additional 3–5 °C increase in freezing tolerance and the expression of genes related to carbohydrate metabolism, photosynthesis and defence processes (dehydrins, ice recrystallization inhibition protein) significantly changed compared to the plants acclimated at non-freezing temperatures. The induction of dehydrins was also confirmed at proteome level (Kosová et al. 2008a, 2010). In another experimental system where the temperature was slowly decreased to −10 or −12 °C after the cold acclimation, genes involved in transcription and defence processes were affected by the subzero acclimation in wheat crowns (Skinner 2009). In summary, the results of transcriptome analysis carried out with plants acclimated at various temperatures indicate that the gradual decrease in temperature during autumn triggers continuous alterations in the transcriptome pattern in order to ensure an effective adaptation to the environment.

6.2.2 Proteomics

A wide range of proteomic studies of temperature stress in plants are currently underway, using numerous methodologies, species, and stress conditions. Despite results obtained so far, information on the systemic response to temperature stress is still quite limited because plant perception and response is often based on factors common to the response to other stresses. What is at least clear is that high and low temperature stresses cause distinct proteome responses in plant tissues (Neilson et al. 2010; Kosová et al. 2011).

Similarly to transcriptome analysis, plant cold acclimation processes at the proteome level were first studied in model plants *Arabidopsis* and rice, where large sequence data sets were available (Bae et al. 2003; Kawamura and Uemura 2003; Imin et al. 2004; Cui et al. 2005; Amme et al. 2006; Yan et al. 2006; Hashimoto and Komatsu 2007). Subsequently, proteome responses to cold were studied in a much broader range of plants including *Arabidopsis* cold- and salt-tolerant relative *Thellungiella salsuginea* (Gao et al. 2009), chicory (Degand et al. 2009), *Festuca pratensis* (Kosmala et al. 2009), soybean (Cheng et al. 2010), pea (Dumont et al. 2011), *Lolium perenne* (Bocian et al. 2011), woody plants such as peach (Renaut et al. 2008), and also wheat (Vítámvás et al. 2007; Sarhadi et al. 2010; Rinalducci et al. 2011a, b) (reviewed in Kosová et al. 2011). Unfortunately, for barley limited information is available on proteome variation following exposure to low-temperature.

Based on studies in cyanobacterium *Synechocystis* strain PCC6803 carried out several years ago, the plasma membrane has been proposed as the primary site of cold-signal sensing (Murata and Los 1997; Suzuki et al. 2000). A temperature decline leads to a decrease in plasma membrane fluidity with changes in composition of membrane phospholipids, which also affects conformation of several transmembrane protein complexes. At the protein level, changes in the composition of plasma membrane proteins in response to cold were first studied in *A. thaliana* by Kawamura and Uemura (2003). The increase in ERD10, ERD14 and COR47 dehydrins indicated protection of membrane-associated proteins against dehydration and denaturation. The increase in outer membrane, lipoprotein-like proteins belonging to the lipocalin family was associated with their important role in membrane biogenesis and repair, as in cellular adaptation to high osmotic stress (Bishop 2000). A similar behaviour was observed for two VfENOD18-like proteins, known to confer tolerance to various stresses including oxidative stress and osmotic shock. From signalling studies, it is known that cold signal is transduced from plasma membrane to nucleus via Ca^{2+} signalling pathway (reviewed in Yamaguchi-Shinozaki and Shinozaki 2006). Accordingly, changes in the abundance of several proteins involved in Ca^{2+} transduction pathway such as a homologue of tobacco DREPP-like protein, synaptotagmin-like protein and phospholipase D δ (PLD δ) have been detected.

In nucleus, cold signal induces profound changes in gene expression underlying the plant cold acclimation process. At the proteome level, Bae et al. (2003) observed in *A. thaliana* a significant increase during the first 4 h of cold followed by a decline in several transcriptional regulators involved in RNA processing and export, such as U2 snRNP-A× and DEAD box RNA helicase; the latter also involved in the regulation of *CBF* gene expression (Gong et al. 2002). A longer lasting upregulation (up to 24 h) by cold has been reported for several heat-shock proteins such as dnaK-type HSP or HSC70-1. In addition, changes in several proteins involved in cold-induced Ca^{2+} signalling, such as calmodulin isoforms, and protein metabolism, such as 20S proteasome alpha subunit G, have also been observed.

At the whole-cell level, cold induces profound changes in hydration status and increases the risk of osmotic and oxidative stress. Cold acclimation process also

displays increased demands on energy metabolism and is associated with a profound redirectioning of the whole cellular metabolism from an active growth and development to an active cold acclimation. The changes in cellular metabolism are also reflected at proteome level. Cold, similarly to other stress factors, leads to disbalances in cellular metabolism increasing the risk of oxidative stress (ROS—Reactive Oxygen Species—formation). Enhancement of antioxidative enzymes involved in ROS scavenging, namely enzymes of ascorbate-glutathione cycle, has been reported in all proteomic studies. Several classes of superoxide dismutase (SOD), ascorbate peroxidase (APX), monodehydroascorbate reductase (MDAR), dehydroascorbate reductase (DHAR), glutathione reductase (GR), as well as several classes of glutathione-S-transferases (GST) and thioredoxin *h* were increased by cold (e.g. Amme et al. 2006; Yan et al. 2006; Degand et al. 2009; Gao et al. 2009; Dumont et al. 2011; Vítámvás et al. 2012).

Chilling similarly to several abiotic stresses such as freezing, drought, salinity or osmotic stress, is also a dehydrative stress, i.e. it leads to a disbalance between water uptake and water release resulting in cellular dehydration. As a response, a *de novo* synthesis of several osmoprotective compounds including both low-molecular hydrophilic compounds, especially carbohydrates (mono- and oligosaccharides), and high-molecular hydrophilic proteins from COR/LEA group occur (Amme et al. 2006; Vítámvás et al. 2007; Degand et al. 2009; Gao et al. 2009; Vítámvás et al. 2012). Apart from COR/LEA proteins revealing hydrophilic and chaperone properties (reviewed in Kosová et al. 2007; Tunnacliffe and Wise 2007; Battaglia et al. 2008; Kosová et al. 2010), changes in several other proteins with chaperone functions, namely HSP proteins, have been observed. An increased abundance of HSP70 proteins and a decreased abundance of HSP90 proteins was found in cold-treated winter wheat (Vítámvás et al. 2012). HSP90 is known to conserve allele variation due to its role in protein folding. Proteins revealing differential primary structure (sequence) encoded by different alleles adopt similar tertiary structure with the aid of HSP90. A decrease in HSP90 abundance could lead to an increased variation in protein conformation which may be advantageous upon stress (Queitsch 2002). Furthermore, an increased abundance of several classes of pathogenesis-related (PR) proteins has been observed upon cold expsosure (Sarhadi et al. 2010).

An increased risk of ROS formation and an enhanced demand of osmoprotectants leads to profound changes in energy metabolism driven by cold. Degradation of polysaccharides, namely starch, followed by glycolysis and Krebs cycle respiratory pathway are upregulated (Amme et al. 2006; Gao et al. 2009). In contrast, an abundance of enzymes incorporating activated glucose into polysaccharide molecules (UDP-glucose pyrophosphorylase) is usually decreased during long-term cold treatments (Gao et al. 2009; Vítámvás et al. 2012); however, during a short-term cold response in young seedlings, it can be increased with respect to control (Cui et al. 2005).

The other major cold-affected process in plants is photosynthesis. An increased risk of ROS formation as a consequence of chilling-enhanced photoinhibition processes results in a down-regulation of electron transport in photosynthetic transport

chain. This is accompanied by a decrease in several protein components of the photosynthetic electron transport chain, as plastocyanin, cytochrome b_6-f complex, Fe-S protein (Rieske protein), and parts of OEC complex (OEE1 protein) (Amme et al. 2006; Gao et al. 2009; Kosmala et al. 2009; Dumont et al. 2011). In contrast, changes in abundance of proteins involved in CO_2 assimilation (Calvin cycle), especially Rubisco large and small subunits and Rubisco activase proteins, are more complex and less clear. Increased abundance of several Calvin cycle components is in fact often observed (Gao et al. 2009; Bocian et al. 2011; Dumont et al. 2011) since rate of enzymatically-catalyzed reactions generally decreases at low temperatures. However, cold acclimation processes reveal increased energy demands, therefore in cold conditions, the CO_2 assimilation needs to be efficiently regulated. As indicated by physiological studies comparing acquired freezing tolerance (FT) levels in long-day (LD) versus short-day (SD) grown barley plants, LD-grown plants reveal higher acquired FT levels (lower LT50 values) in comparison to SD-grown plants when all plants were in vegetative stage (Limin et al. 2007). This result is probably caused by longer photosynthetic periods and higher accumulation of soluble sugars in LD-grown plants in comparison with the SD-grown ones. Thus, a sufficient rate of CO_2 assimilation during cold conditions would be crucial for plant ability to efficiently acclimate to cold.

Another widely reported proteome change due to cold is related to metabolism of methionine and S-adenosylmethionine (SAM) (Cui et al. 2005; Amme et al. 2006; Yan et al. 2006; Vítámvás et al. 2012). SAM is well-known not only as a universal methyl donor in plant methylation reactions, but also as a precursor of ethylene and polyamine biosynthesis. Proteomic studies reporting increased abundance of methionine synthase and SAM synthase under cold thus indicate profound metabolic changes dealing with stress signalling and osmoprotection.

In *Triticeae*, proteome changes upon cold have been studied in bread wheat with respect to both cold acclimation and vernalization processes (Vítámvás et al. 2007; Sarhadi et al. 2010; Rinalducci et al. 2011a, b; Vítámvás et al. 2012). Proteome changes with respect to cold acclimation process have been studied both in spring and winter backgrounds. From physiological studies, it is well-known that upon cold spring genotypes can acquire only a limited level of FT in comparison to winter ones (Fowler et al. 1996b; Fowler 2008). The differences in acquired FT between spring and winter genotypes, but even between winter genotypes with different levels of acquired FT, are mirrored at relative abundance levels of several COR/LEA proteins, namely wheat LEA-II WCS120 proteins (Houde et al. 1992; Vítámvás et al. 2007), their barley homologue DHN5 (Kosová et al. 2008), and stromal LEA-III COR14b protein (Crosatti et al. 1995; Vágújfalvi et al. 2003).

A complex comparison of dynamics of FT acquisition in a winter and a spring wheat genotype with respect to phytohormone levels has revealed that in the winter genotype cold acclimation process is associated with a significant upregulation of several "stress-responsive" phytohormones such as ABA, jasmonic acid (JA), salicylic acid (SA), together with a significant downregulation of growth regulators such as active gibberellins (GAs), cytokinins (CKs) and auxin (IAA). On the other hand, in the spring genotype an initial upregulation of stress-responsive phytohormones,

with a downregulation of positive growth regulators is reversed in later phases of cold treatment (Kosová et al. 2012). These data thus could elucidate why spring genotypes can acquire only a limited level of FT in comparison to the winter ones. At the proteome level, a decrease of several photosynthetic proteins (OEE1, OEE2, ferredoxin NADPH oxidoreductase, fragmentation of Rubisco large subunits, several Calvin cycle enzymes) indicates an impairment of photosynthetic processes.

From studies on transgenic *Arabidopsis*, overexpression of *CBF1* is known to result in plant dwarfism and to positively regulate accumulation of DELLA proteins which act as GA inhibitors (Achard et al. 2008). Conversely, upregulation of transcripts induced by GA seems to represent a logical consequence of a downregulation of *CBFs*. At protein level, Rinalducci et al. (2011b) reported vernalization-induced upregulation of two important proteins, a glycine-rich RNA binding protein (GR-RBP), whose *A. thaliana* homologues are known to act as repressors of FLC (Quesada et al. 2005), and a lectin protein VER2, which is involved in β-N-acetylglucosamine signaling during vernalization (Xing et al. 2009). However, when accumulation of cold-inducible protective proteins (COR/LEA, chaperones) has been evaluated, no dramatic changes caused by vernalization have been reported with respect to the control (Sarhadi et al. 2010; Rinalducci et al. 2011b; Vítámvás et al. 2012). These findings are caused by the fact that the vernalization process has been studied on cold-treated plants and that both unvernalized and vernalized plants have been exposed to conditions inducing cold acclimation. Proteins, unlike transcripts, are direct effectors of plant cold response; a downregulation of cold-inducible proteins in vernalized, but still cold-treated, winter wheat plants would have fatal effects on their survival. Therefore, it can be proposed that there is a time lag between a downregulation of cold-inducible transcripts and cold-inducible proteins in vernalized winter wheat plants, especially when COR/LEA proteins are investigated (Sarhadi et al. 2010; Vítámvás et al. 2012). However, when both unvernalized and vernalized winter wheat plants were exposed to warm temperatures (cold deacclimation) and then reacclimated to cold, significant differences in WCS120 protein accumulation between unvernalized and vernalized winter wheat plants were reported (Vítámvás and Prášil 2008).

From a proteomic point of view, it can therefore be concluded that vernalization fulfillment in cold-treated winter wheat plants is associated with profound changes at the signalling level, but not with rapid changes in cold-inducible protective proteins. A rapid decline of cold-inducible proteins can be achieved only by cold deacclimation, since accumulation of cold-inducible proteins is necessary for plant survival under low temperatures.

6.2.3 Metabolomics

Low temperature-induced changes in transcriptome and proteome finally result in the reconfiguration of metabolomes. Alterations in the level of osmolytes (carbohydrates and free aminoacids), antioxidants, polyamines and other metabolites are

important in the cold acclimation process, and result in an increased freezing tolerance. A coordinated increase in the concentration of amino acids derived from pyruvate and oxalacetate, of polyamine precursors and compatible solutes was observed during cold shock experiments in *Arabidopsis* (Kaplan et al. 2004). The role of CBF regulon was confirmed also at metabolite level in the model plant, since the cold acclimation-induced extensive changes in metabolome could be mimicked by constitutive overexpression of *CBF3*. At the same time, the low temperature metabolome was depleted in metabolites controlled by the CBF regulon in a more sensitive *Arabidopsis* ecotype (Cook et al. 2004). In contrast to *Arabidopsis*, complex metabolome studies were not done with cold-treated wheat and barley. However, several metabolites being important in cold acclimation were studied in these cereals.

A central role of carbohydrate metabolism in the reprogramming of metabolome during temperature stress was suggested by Guy et al. (2008). The positive role of carbohydrates in the cold acclimation process was demonstrated several years ago in wheat, since the activity of sucrose phosphate synthase, sucrose fructosyl transferase and acid invertase increased only in the winter-tolerant genotype but not in the sensitive one, and the fructan accumulation was greater in the former one (Savitch et al. 2000). Using other wheat genotypes with different freezing tolerance, a significant correlation was found between fructose and sucrose content, and freezing tolerance. The comparison of chromosome 5A substitution lines with different freezing tolerance indicated the effect of this chromosome on cold-induced fructose accumulation (Vágújfalvi et al. 1999). The cold-induced up-regulation of sucrose synthase, with the down-regulation of beta-fructofuranosidase genes may result in sucrose accumulation (Kocsy et al. 2010).

Cold acclimation altered both the composition and amount of free amino acids in wheat (Kovács et al. 2011). The ratio of amino acids belonging to the glutamate family increased and the ratio of those ones of aspartate family decreased. Considering the individual amino acids, Asp, Glu, Gln and Pro levels were greatly induced by cold, and these changes were also observed at gene expression level in the case of Pro and Glu. Cold-induced increase in total amino acid content derived mainly from the accumulation of Pro, Glu and Gln in bluegrass (Dionne et al. 2001). The increase in Arg content would drive the greater rate of polyamine synthesis which was described in cold-hardened wheat (Kovács et al. 2010). On the other hand, the cold-induced accumulation of Glu should influence not only polyamine, but also glutathione (GSH) synthesis, the activation of which at low temperature was described in wheat (Kocsy et al. 2000). Cold treatment resulted in increased free amino acid levels in barley as in wheat, and it was suggested that Glu decarboxylation together with GABA metabolism would play a role in freezing tolerance (Mazzucotelli et al. 2006). As an osmolyte, Pro has a special role during cold acclimation, since its accumulation may prevent the water loss from the cells occurring at subzero temperatures, due to extracellular ice formation. In winter wheat, cold acclimation increased Pro content and the change was greater in the tolerant genotype than in the sensitive one (Dörffling et al. 1990; Macháčková et al. 2006).

The cold-induced increase in the amount of the amino acids, which are precursors of polyamines is thus associated with an increased polyamine synthesis, as described

in wheat (Kovács et al. 2010). In the case of putrescine (Put), the increase was not only observed at the metabolite level, but also at the transcript level of the corresponding gene. The importance of Put in the response to low temperature stress was demonstrated in tomato leaves, in which exogenous Put decreased the cold-induced electrolyte leakage, while the inhibition of Put synthesis increased membrane damage (Kim et al. 2002). The accumulation of Put was also observed in alfalfa and wheat during cold hardening, with a decrease in control plants in the activity of arginine decarboxylase, a key enzyme of Put synthesis (Nadeau et al. 1987). In addition, the Put, spermidine and cadaverine contents increased after cold hardening in a winter wheat genotype, while only the concentrations of the polyamines spermidine and spermine increased in a spring wheat variety, indicating the involvement of polyamines in the response to low temperature stress.

Combinatorial approaches that integrate metabolome and transcriptome data have recently elucidated regulatory networks acting in response to environmental stresses. Clear examples are from the model species *Arabidopsis*, in which the metabolome was analyzed using various types of mass spectrometry after cold and dehydration exposure, and metabolic profiles have been then combined into regulatory networks together with transcriptome data. Maruyama et al. (2009) coupled microarray analysis of transgenic plants overexpressing of genes encoding DREB1A/CBF3 and DREB2A transcription factors with the metabolic pathways that act in response to cold and dehydration.

6.3 Molecular Networks: Key Steps in Cold Acclimation (Genomic Role of the CBF Regulon)

The fundamental acquisition of FT in fall-sown temperate cereals is associated with several interconnected physiological and molecular changes, aimed at protecting critical cell structures and vital processes during freezing. The adaptive phenomenon of cold acclimation typically occurs when plants are exposed to low non-freezing temperatures (usually below 9 °C) for several weeks (4–6) and contributes significantly to strengthen the capacity to cope with the stress (Fowler et al. 1996a; Kosova et al. 2008a; Rizza et al. 2011). During the past two decades, molecular biology research of model and crop plants has taught us that cold acclimation is an extremely complex process, whose final outcome consists of increased levels of sugars, soluble proteins, proline and organic acids, the appearance of new enzyme isoforms, and the modifications in lipid membrane composition (Heidarvand and Maali Amiri 2010, and as already described in previous chapters). On the whole, the molecular mechanisms leading to such physiological changes can be ascribed to three main steps: (1) perception of external/physical changes related to temperature drop; (2) transduction of the signal to the nucleus, including the activation of transcription factors; and (3) activation of the low-temperature regulated gene batteries acting as the real effectors of the response (for in-depth reviews see Penfield 2008 and Nakashima et al. 2009). As one can imagine, such view of the phenomenon through

linear steps may be too reductive, especially in light of the results accumulated in the eudicot model *Arabidopsis*. This information provided plant scientists with efficient and valuable resources for co-expression and comparative studies aimed to reconstruct gene networks and pathways that are in turn used to infer functional interactions among genes, proteins and metabolites (Sasaki et al. 2010; Barrett et al. 2011; Mochida et al. 2011). Drawing on Thomashow's commentary on the recent insights gained in the molecular basis of plant cold acclimation, it seems promising that questions on: (i) how low temperature signal is sensed/processed to activate the first wave of cold-regulated genes; (ii) what regulatory logic underlies the cascading pattern of the low-temperature gene network; and (iii) what biological functions can be ascribed to the genes that constitute the various circuits of the network, should find a complete answer soon (Thomashow 2010). The aim of the present paragraph is to make a comprehensive summary of the fundamental players that act in the regulatory networks leading to cold acclimation.

6.3.1 Function of the CBF Regulatory Hub

After exposure to low temperatures, the transcriptome of the *Triticeae* undergoes an extensive progressive reorganization with thousands of genes involved being up- or down-regulated (Svensson et al. 2006; Greenup et al. 2011; Laudencia-Chingcuanco et al. 2011). As estimated in *Arabidopsis*, among these waves of cold-regulated genes, the impressive number of more than 200 transcription factors act in concert for transcriptome reconfiguration and serve as major regulatory for "hubs" acclimation (Thomashow 2010). Although a precise topology of the regulatory network activated in response to cold exposure is still far from being fully understood, the most clearly understood node from both the *cis*-regulatory and transcription factor side is represented by the CBF circuit. The CBFs are characterized by a plant specific APETALA2/ethylene-responsive element binding factor (AP2/ERF) domain flanked by a typical sequence motif, the so called "CBF signature" (Stockinger et al. 1997; Skinner et al. 2005). It is the AP2/ERF domains that interacts with the CRT elements present in the promoter region of their target genes, referred to as the CBF regulon, to increase their expression level. Starting from the early 1990s, the CRT elements, defined by the 5'-CCGAC-3' core motif, and the CBF proteins were isolated and characterized in several plant species from *Arabidopsis* to *Triticeae* crops (Baker et al. 1994; Stockinger et al. 1997; Ouellet et al. 1998; Choi et al. 1999; Xue 2003; Dubouzet et al. 2003). Much of our understanding of this circuit was built on a bottom-up approach, by linking the results of promoter binding assays and mutant analysis (for a recent review see Chinnusamy et al. 2010), whereas the top-down studies used DNA microarrays to establish components and wiring of the CBF regulon, and to add new putative components (Vogel et al. 2005; Greenup et al. 2011; Laudencia-Chingcuanco et al. 2011).

As summarized earlier, a series of stress-inducible genes that play a major role in acclimation and acquisition of FT are downstream *CBFs*. Examples of targets

within the CBF regulon are genes encoding enzymes involved in the biosynthesis of osmo- and cryo-protectant molecules (such as sucrose, raffinose and proline) but also other transcription factors, late-embryogenesis abundant (LEA) proteins, COR and cold-inducible (KIN) proteins, phospholipase C enzymes, and sugar transport proteins (Heidarvand and Maali Amiri 2010). The primary functional involvement of the target stress-inducible genes is thus the protection of membranes and proteins against the severe water deprivation stress that occurs with freezing (Penfield 2008). However, beside production/accumulation of cryoprotectants, other metabolic functions are affected by the CBF regulatory hub in contributing to the ability of plants to survive frost. A well-known phenomenon in *Arabidopsis* plants that constitutively overexpress *CBF* genes under normal growth conditions is severe growth and development retardation (Jaglo-Ottosen et al. 1998). Results accumulated in transgenic tomato plants overexpressing *AtCBF1* suggested a link between the CBFs and giberellic acid (GA) through nuclear-localized DELLA proteins (Hsieh et al. 2002). Afterwards, Achard et al. (2008) demonstrated that the CBFs enhance the expression of GA-inactivating GA2-oxidases allowing the accumulation of DELLA protein repressors, which in turn lead to dwarfism and late flowering. Although the role of DELLA-mediated change in plant architecture has still not been completely clarified, one could speculate on how this phenotype would potentially contribute to winter survival of fall sown cereals. For example, a more prostrate growth habit should result in a greater chance of the plants being covered by snow and insulated against harsh air temperatures (Roberts 1990).

It is worthy to note that the majority of the studies done on wheat and barley CBFs focused only on transcript levels, and not on CBF protein abundance, post-transcriptional and post-translational modifications. However, a further interesting perspective related to the role played by CBFs in the grasses concerns their *cis*- and *trans*-acting regulation. The rationale is that once the transcription factors involved in expression of the first wave of stress-responsive genes are identified, further studies can be followed to work back upstream into the stress-sensing system and thus understanding to which extent the activities of the transcription factors are affected. In this view, the *Arabidopsis* research community provided indications about two cold sensing pathways upstream of the CBF regulatory hub: post-translational modification and calcium signaling (Thomashow 2010).

In the first pathway, *Inducer of CBF expression1* (*ICE1*) gene is involved, and it binds to multiple MYC DNA regulatory elements in the promoters of certain *AtCBF*s to stimulate their transcription (Chinnusamy et al. 2003). Such a major positive regulator mediates cold induction via low temperature-induced activation rather than by transcription modulation. On one side, ICE1 is activated by small ubiquitin-related modifier (SUMO) so that sumoylation (i.e. SUMO conjugation) of ICE1 is directly responsible of the positive induction of the CBF regulon and is mediated by the SIZ1 protein, a SUMO E3 ligase (Miura et al. 2007). On the other side, this activation process is finely modulated by *High expression of osmotically responsive gene 1* (HOS1), a RING finger E3 ligase that mediates ubiquitination and targeted proteolysis of ICE1 (Dong et al. 2006). The SIZ1-HOS1 system thus appears to contribute to the "transient" nature of ICE1 as a master regulator controlling the

CBF regulon and many other cold-responsive genes. In wheat, the ICE1 homologs *TaICE141* and *TaICE187* are constitutively expressed and activate the wheat *CBF* genes of group IV, which are associated with freezing tolerance. Overexpression of *TaICE141* and *TaICE187*in *Arabidopsis* positively stimulated *CBF* and *COR* gene expression and enhanced freezing tolerance only after cold acclimation. This finding suggests that, similar to AtICE1, also wheat ICE1 proteins need to be activated by cold acclimation (Badawi et al. 2008).

In addition to the ICE1 mediated pathway, the CBF regulatory hub is upstream-regulated by a second cold-sensing pathway in which cytosolic signatures of the second messenger Ca^{2+} are major players. Calcium signatures are sensed by calcium sensor family proteins, namely calcium-dependent protein kinases (CDPKs), calmodulins (CaMs), and salt overly sensitive 3-like (SOS3-like) or calcineurin B-like (CBL) proteins (for a review see Reddy et al. 2011). Several studies linked the low-temperature induced intracellular Ca^{2+} spikes to *COR* gene expression in *Arabidopsis* (Knight et al. 1996; Knight 2000), and it is demonstrated that cold induction of certain CBF regulon genes is weaker either in mutants impaired in Ca^{2+} channels and Ca^{2+}/H^{+} antiporters or using chemical agents that chelate/block calcium influx into the cytoplasm (Monroy et al. 1997; Catala et al. 2003). Recently, Doherty et al. (2009) analyzed the promoter region of the three *AtCBFs* and found seven conserved DNA motifs (CM1-7) responsive to cold induction. In particular, CM2 matched the calmodulin-binding transcription activator (CAMTA) binding sequence 5'-CGCG-3' (which also overlaps the ICE1 regulatory element) and brought to cold-induced expression of *AtCBF*s. Additional information about interconnection between calcium cellular concentration and gene regulation in response to stress is beyond the scope of this paragraph and can be found reviewed in detail by Chinnusamy et al. (2010) and Reddy et al. (2011). As far as we know, no or very few studies on isolation and characterization of CAMTA genes involved in low-temperature tolerance induction in wheat and barley have been published. However, the evidence accumulated in the model plant *Arabidopsis* would suggest that the *Triticeae* research community deepen the study about the existence and diffusion of such a conserved regulatory pathway.

6.3.2 Early Events in the Cascade

An interactive reactive oxygen species (ROS)—nitric oxide (NO) concentration-dependent signaling mechanism has been recently proposed by Cantrel et al. (2011). The authors showed how, after membrane rigidification at low temperature exposure, NO act as a key player at the very beginning of the regulatory network activation in modulating the synthesis of sphingolipids signals. They also proposed a model in which the cold stress is perceived as a result of plasma membrane rigidification, and this in turn results in the production of phosphatidic acid through the activation of diacylglicerol kinase (DAGK) and/or phospholipase D (PLD). Phosphatidic acid activates a short spike of ROS and NO that is quickly modulated to a subtoxic

concentration by nonsymbiotic haemoglobins (nHb). Similarly to the first report in which either sphingolipid phosphorilation or its regulation by NO were implicated in plant cold-signal transduction, the findings by Cantrel et al. (2011) open the way for future studies aimed to further increase our understanding of these complex and inter-related processes in model and crop plants.

6.3.3 Integration of the Circadian Control

Besides the described molecular mechanisms leading to the development of freezing tolerance, it is important to briefly mention here also the influence of the circadian clock on the CBF regulatory hub. The circadian clock has been shown to control the expression of *CBF* genes in *Arabidopsis* and tomato (Fowler et al. 2005; Pennycooke et al. 2008), suggesting that this could be a highly conserved way of regulation. Expression analyses in both barley and hexaploid wheat also indicate that certain *CBF* genes are circadian-regulated, while others are not (Badawi et al. 2007; Stockinger et al. 2007), with the maximal cold-induced increase in CBF transcription that occur when cold stress is imposed 4 h after dawn. Determining how the low temperature regulatory network and the circadian clock are integrated will not be trivial. However, some insights were recently obtained in *Arabidopsis* where a complex molecular circuit seems to directly affect the CBF regulatory hub both in a negative and in a positive way of action. On one hand, it was demonstrated as the *phytochrome-interacting factor 7* (PIF7), a bHLH transcriptional repressor, interact with *timing of cab 1* (TOC1), a component of the central circadian oscillator, and a red light photoreceptor PHYB (Kidokoro et al. 2009). On the other hand, a direct positive action of two transcription factors that are core components of the clock, i.e. *Circadian Clock-Associated 1* (CCA1) and *Late Elongated Hypocotyl* (LHY) has been found (Dong et al. 2011). In this latter report, the authors proposed a first comprehensive model for circadian regulation and gated cold induction of *AtCBF -1*, *-2*, and *-3* in response to day/night oscillation coupled to warm/cold temperature exposure, and also include the mechanism of action of the above mentioned factors ICE1 and CAMTA (Dong et al. 2011). Due to the expected conservation of the circadian clock regulation system between *Arabidopsis* and the *Triticeae*, one could predict that in the near future the orthologous components of the regulatory network should be identified in wheat and barley as well.

6.3.4 CBF-Independent Hubs

Genes induced by abiotic stresses (mainly dehydration and cold) can be classified according to two main signal transduction pathways that able to convert the external stimulus into cellular responses: the ABA-independent and the ABA-dependent. The CBF regulatory hub discussed so far functions in the ABA-independent pathway. However, transcriptomic analyses involving *Arabidopsis* mutants identified through

genetic screening for freezing tolerance, revealed that several classes of transcription factors are affected by ABA and are related to multiple stresses (for recent reviews, see Nakashima et al. 2009; Chinnusamy et al. 2010). For example, after a reporter-gene genetic screening (based on P_{RD29A}*:LUC* construct) two constitutively expressed transcription factors, HOS9 (a homeodomain protein) and HOS10 (a R2R3-type MYB), which are necessary for developing low-temperature tolerance were identified (Zhu et al. 2004, 2005). Through transcriptome analysis it was possible to identify the HOS9 regulon as distinct from the CBF regulon (Zhu et al. 2004) and, since HOS10 positively regulates NCED3 (9-cisepoxycarotenoid dioxygenase), this transcription factor has been assigned to ABA-dependent cold acclimation pathways (Zhu et al. 2005). Among the CBF-independent regulons those under the control of ZAT12 and HOS9 show remarkable characteristics (Vogel et al. 2005; Chinnusamy et al. 2010) (Fig. 6.1). The ZAT12 regulon contains 9 cold-induced and 15 cold-repressed genes and the overexpression of *ZAT12* gene resulted in increased freezing-tolerance. In *hos9-1* mutant seedlings, the expression of 140 genes was higher and that of 35 genes was lower than in wild type, and from the highly expressed genes 41 appeared to be cold-inducible (Zhu et al. 2004). Both *ZAT12* and *HOS9* genes are negative regulators of some genes of the CBF regulon (Zhu et al. 2004; Vogel et al. 2005), which indicates a possible interaction between the various cold-responsive regulons. *Eskimo1* (*esk1*) is a constitutively freezing-tolerant *Arabidopsis* mutant characterised by a 30-fold higher level of proline than wild-type plants, supporting an important role of the osmolytes in stress resistance (Xin et al. 2007). ESK1 is constitutively expressed and encodes the protein domain of unknown function which mechanism of action has yet to be revealed. However, transcriptome comparison of CBF2-overexpressing plants and *esk1* mutants showed that different sets of genes are regulated by CBF2 and ESK1 (Fowler et al. 2005).

Arabidopsis showed its potential as a well-established model system also through the transgenic analysis of cold-inducible transcription factors identified in crop species and helped in validation of functions of some transcription factors in cold tolerance. As intriguing examples, overexpression in *Arabidopsis* of the cold-regulated rice transcription factors OsMYB4 (an R2R3-type MYB) and OsMYB3R-2 (an R1R2R3 MYB) enhanced freezing tolerance (Vannini et al. 2004; Dai et al. 2007). Similarly, transgenic *Arabidopsis* overexpressing *TaERF1*, a wheat gene induced by cold, drought salinity, ABA, ethylene, salicylic acid, and infection by *Blumeria graminis* f. sp. *tritici*, exhibited enhanced tolerance to cold, salt, and drought stresses, as well as pathogens (Xu et al. 2007). Finally, although still not deeply characterized, the two transcription factors *wheat low-temperature-induced protein 19* (WLIP19) and *T. aestivum ocs-element bindingfactor 1* (TaOBF1), two bZIP-type proteins, were found to form heterodimers that were able to activate the expression of *COR/LEA* genes in the development of abiotic stress tolerance. In conclusion, these results suggest that a number of transcriptional networks operate during cold acclimation and cold stress tolerance in addition to the major role played by the CBF hub. Comparative analysis of gene expression patterns between *Arabidopsis* and the Triticeae are important future tasks, and further research efforts are expected to better clarify the interconnection of the whole gene regulatory network involved in cold acclimation.

6.4 Exploiting Genetic Resources and Genomic Selection for FT

6.4.1 Genetic Resources

The progenitors of current wheat and barley varieties come from the Fertile Crescent and most wild *Triticeae* species possess a winter growth habit (Kosová et al. 2008b) as they are adapted to climatic conditions where the annual rainfall is concentrated in the autumn and spring, followed by hot and dry summers. Plants develop vegetative organs during rainy autumns, and use vernalization to delay flowering until winter is over, while the photoperiod sensitivity allows them to flower and complete grain filling before the hot summer begins. The first domesticated cereals shared the growth habit of their wild relatives. Farmers selected for spring forms which could be sown and harvested in a shorter season and allow growing two crops in succession each year (Cockram et al. 2007b). Mutations in spring barley resulted in reduced photoperiod responsiveness alleles that removed the promotion of flowering in response to long days (Turner et al. 2005). Those factors are likely to have favoured expansion of barley production to higher latitudes allowing it to avoid injury during cold winters (Cockram et al. 2009), while taking advantage of the long cool and wet summers of northern Europe (Pourkheirandish and Komatsuda 2007; Comadran et al. 2012). Distinct clustering of types classified as winter or spring has been maintained and is reported in all surveys of cultivated crops diversity (Cockram et al. 2007b; Comadran et al. 2012). Some, rather infrequent, materials are classified as "alternative" or "facultative". No unambiguous description for this type exists; von Zitzewitz et al. (2005) classified them as cold-tolerant and vernalization unresponsive, while according to the International Union for the Protection of New Varieties of Plants (UPOV) they display an intermediate flowering time (respect to that of winter and spring types) when grown under inductive photoperiods without vernalization (Cockram et al. 2009). Many winter, most facultative, and few spring barleys are sensitive to short days, whereas winter and facultative barleys are more cold-tolerant than spring types. Since those traits are controlled by different genes, cold-tolerant, facultative varieties could be the best choice (www.barleyworld.org). Wheat is considered to have the broadest adaptation of all cereal crops and is cultivated in a wide range of environments due, largely, to its tolerance to cold. Winter and facultative wheats are grown on one third of the 220 million ha devoted to wheat worldwide. The most winter-hardy wheat cultivars are required for areas in the northern Great Plains of North America, the Russian Federation and Ukraine and, to less extent, ineastern, central and northern Europe, eastern Turkey, northwest Iran and China (Braun and Sãulescu 2002).

Breeding for uniformity for the development of high-yielding cultivars as requested by modern agriculture, rapid population growth and economical changes led to a drastic intraspecific narrowing of the genetic base of barley and wheat, leaving behind many potential useful genes, and enhancing the risk of losing adaptation to abiotic stress such as frost. The conservation and availability of genetic diversity of crops and their wild relatives provide farmers and breeders with materials for

improving and adapting the crops to face future environmental, climatic and economic changes in a sustainable way (FAO 2010). Harlan and de Wet (1971) were the first authors to classify plants using the "three gene pool" concept. The primary gene pool consists of species which can be easily intercrossed; the secondary gene pool includes related species; their crossing with the target produces at least some fertile hybrids. Species in the tertiary gene pool can be crossed only by applying techniques such as embryo rescue, bridge crossing or protoplasm fusion (Acquaah 2006). In the case of barley, wild barley (*H. vulgare* spp *spontaneum)* belongs to the primary genepool; *H. bulbosum*is the only member of the secondary genepool, while the tertiary genepool of barley comprises about 30 *Hordeum* species (Pickering and Johnston 2005). The primary pool of wheat comprises all *Triticum* species, the secondary pool—all *Aegilops* species, the tertiary wild relatives pool includes some remote members of the tribe *Triticeae*.

In order to preserve the existent crop diversity, major cereal collections have significantly increased in recent years. In total 466,531 barley accessions are held worldwide, and major holders are Plant Gene Resources of Canada (PGRC) with 9 % of them, and National Small Grains Collection (NSGC) USA with 6 % (FAO 2010). If we consider the division of resources according to the growth habit, the European Barley Database (EBDB) of the Institute of Plant Genetics and Crop Plant Research (IPK) reports for advanced/improved cultivars 1,431 winter; 2,875 spring and 10 facultative accessions. For wild barley there are 135 winter, 320 spring and 2 facultative, while among landraces there are 1,485 winter, 14,180 spring and 90 facultative accessions (http://barley.ipk-gatersleben.de/ebdb.php3).

According to a FAO's recent report there are 856,168 total world accessions of wheat, with major holders being CIMMYT (13%) and NSGC (7 %) (FAO 2010). If we consider the division according to the growth habit, there are 19,179 accessions of winter, 11,419 of spring and 119 intermediate advanced/improved cultivars in the European Wheat database (EWDB) (http://genbank.vurv.cz/ewdb/) that contains data of wheat collections stored in European countries. Among landraces we find 8,058 winter, 9,921 spring and 180 intermediate accessions. As far as wild *Triticum aestivum* is considered, there are 126 winter accessions and 142 characterized as spring. For durum wheat, there are 614 winter, 3,145 spring and 11 intermediate improved/advanced accessions, 644 winter, 4,044 spring and 109 intermediate traditional varieties/landraces.

Variation in physiological traits associated with salt, cold and drought tolerance and N-starvation has been reported in the highly variable wild progenitor *H. spontaneum* (Hussain 2006). Perennial *H. bulbosum* that includes both diploid and tetraploid forms is an obligate outbreeder (self-incompatible) and the only member of the secondary genepool of *Hordeum* that represents a valuable source of genetic diversity for barley crop improvement. A set of diploid introgression lines (ILs) containing chromatin introgressed from *H. bulbosum* into cultivated barley (*H. vulgare*) were generated that represent a significant germplasm resource likely to contain genetic diversity that can be mined for improvement of traits of interest for barley breeders and researchers (Johnston et al. 2009).

A large number of accessions of *A. tauschii* have been screened for FT and found to contain cold hardy accessions, although none of these were as cold hardy as winter wheat cultivars (Limin and Fowler 1991). Moreover, the cold hardiness levels of synthetic hexaploid wheat, produced by combining tetraploid wheat with *A. tauschii*, to introduce new cold hardiness genes into the common hexaploid wheat gene pool, did not identify transgressive segregates for improved cold hardiness, even if the cold hardiness levels of hybrids ranged from similar to equal to the hardy parent. These observations led Limin and Fowler (1993) to conclude that the close wheat relatives, sharing common genomes with *T. aestivum*, cannot be considered promising sources of new genes for cold hardiness improvement. On the other hand, recently wild emmer *T. dicoccoides* has been reported to be a rich, mostly untapped genetic resource for improvement of cultivated wheat and genetic traits of economic significant such as earliness, yield and cold tolerance (Nevo 2011).

Related cultivated species could provide another potential source of genetic variability. The much greater winter hardiness of rye than that known in wheat and barley proves that its potential to improve wheat for the trait has been largely unused. Hexaploid or octaploid *Triticales* are synthesized by crossing rye either with tetraploid or hexaploid wheats, respectively. Secondary hexaploid triticales, that nowadays are the most common commercial *Triticales* worldwide, derive from intercrosses between different primary hybrids, after backcrosses with wheat. At present, winter *Triticale* varieties are as winter hardy as the best winter wheat varieties, but less than winter rye (Alberta Agriculture, Food and Development 2001). Rizza et al. (1997), comparing the behaviour of the most frost resistant *Triticale* (cv. Aubrac) with a winter wheat, demonstrated that this genotype could perform much better than the winter wheat and, under specific test conditions, even as good as a typical frost-resistant rye.

6.4.2 Genomics-Assisted Breeding

Until present, prerequisite for any application of DNA-based technologies to plant breeding is sufficient knowledge of the genetic bases for agronomically relevant traits: inheritance, number of loci and weight on the traits of single loci. Secondly, sufficient knowledge of markers associated (in linkage disequilibrium) with loci supporting the trait: nature and information content, position in the genome, distance from the gene responsible for the trait. Gene and QTL cloning allow in this case moving from associated, linked, markers, to "perfect" candidate gene-derived markers.

As underlined in the previous paragraphs, genomic research has allowed in the last decade to make significant advances in knowledge of the genetic bases for freezing tolerance in plants, as well as in *Triticeae*. Although considered a polygenic trait, since the early 90's QTL mapping has led to the identification in *Triticeae* of a relatively small number of quantitative trait loci having major effects on the ability of the plant to survive freezing (Galiba et al. 2009; Pecchioni et al. 2012). After the first reports that identified a highly significant genomic region of chromosome

group 5, associated with *VRN-1* vernalization requirement locus (Galiba et al. 1995), Francia et al. (2004) demonstrated that FT was mainly controlled by two linked QTLs, *FR-H1* and *FR-H2*, thanks to the Nure x Tremois (winter x spring), unique mapping population where both FR- QTLs were segregating in the *Triticeae*; in wheat *FR-A2* (Vágújfalvi et al. 2003) and *FR-A1* were mapped in different genetic systems. A cluster of at least 12 C-repeat binding factor (*CBF*) genes are the best candidate genes underlying *FR-2*, whereas *FR-1* was later identified with the *VRN-1* vernalization response locus on chromosome 5A, and then cloned by Yan et al. (2004). The two QTLs together explain a large part of variation for the trait (e.g. 65.9 % of tolerance in a controlled freeze test in barley, Francia et al. 2004), and this fact limited the success of identification of other, minor effect, QTLs for freezing tolerance through such linkage mapping approach. A probably hortologous locus, *FR-B1*, with a smaller effect than *FR-1*, was mapped on the homoeologous chromosome 5B (Tóth et al. 2003). In barley, Reinheimer et al. (2004) indicated two chromosomes, 2H and 5H, as implicated in the genetic control of reproductive FT. A pioneering work by Tuberosa et al. (1997) found a slightly more complex situation in a winter x winter type barley cross Arda x Opale. Nine freezing tolerance QTLs were mapped on chromosomes 2H, 3H, 6H and 5H, after a screening conducted in controlled environment, with the one of 5H roughly coinciding with *FR-1* (Tuberosa et al. 1997). For other identified loci, due to the nature of markers used, it was not possible to identify precisely colinearity regions with more dense genetic maps (Cattivelli et al. 2002). Recently and still in barley, 285 spring type cultivars were evaluated for FT and genotyped with 1,536 gene-based SNPs for an incremental association mapping approach and significant new marker/trait associations have been detected on chromosome 4H and 5H, that do not co-localize with *FR-H1* and *FR-H2* loci (Tondelli et al. 2009).

Varshney et al. (2005) discussed how for some traits it could be necessary to use crop wild relatives to introgress some of the diversity that was lost during domestication. That might be the case of freezing tolerance too. In the cross Arta x *Hordeum spontaneum 41-1,* it was shown that whereas cultivated barley contributed higher scoring alleles for the major effect QTLs for cold tolerance, for some QTLs with minor effects (on 2H, 4H and 6H), the allele from the *H. spontaneum* conferred tolerance (Baum et al. 2003), suggesting a deeper investigation of the wild resources. The use of association mapping to position loci controlling economically important traits in a wild crop progenitor was tested on a collection of 318 accessions from the Fertile Crescent, Central Asia, and North Africa assembled to form the Wild Barley Diversity Collection (WBDC) genotyped with 50 SSR and 1090 DArT markers, with 3,000 additional SNP markers to follow and evaluated for resistance to six barley diseases and 25 agronomic and morphological traits (Steffenson et al. 2007).

Although it is possible that in the future new introgressions, new crossing schemes, especially among either winter or spring types, or very large populations like NAM (Nested Association Mapping; Buckler et al. 2008), will allow to identify other QTLs, it would be less likely that they have a large influence on tolerance to frost. For this reason, the routine application of DNA-based technologies to breeding for this trait could be focused more on marker-assisted selection (MAS) than on genomic selection (GS) approaches. Owing to Kumar et al. (2012), the first approach could be

named LD-MAS, i.e. MAS using markers in LD with a QTL. Moreover, this breeding strategy could be based on candidate genes (CGs) and "functional" markers rather than on anonymous markers.

6.4.2.1 LD-MAS for Freezing Tolerance

Availability of functionally characterized genes, ESTs collections and ongoing genome sequencing projects have facilitated in important crops like wheat and barley the development of molecular markers directly from the transcribed regions that are commonly referred to as "genic" or "functional" markers (FMs) as their putative functions can be derived from homology searches. Such markers can target directly the functional polymorphism within the gene that is causing variation in the influenced trait, enabling selection in different genetic backgrounds without the need to validate the marker. That is why they are also commonly named "perfect markers" (Varshney et al. 2005).

The development of a reliable molecular test for GH in wheat and barley could find numerous applications. Crosses within GH classes, predominant in modern breeding contribute to corresponding genotypic division of spring and winter types, diagnostic markers for *VRN* genes could be thus deployed in breeding programs in order to increase genetic diversity by utilizing winter × spring crosses (Cockram et al. 2008, 2009). For example, the identification of *VRN-H1* and *VRN-H2* genes in barley allows screening of germplasm collections and classification of the alleles and allele combinations present in modern cultivars, as well as discovery and characterization of novel alleles and allele combinations controlling vernalization requirement (Cockram et al. 2007a). Since a highly conserved VRN-1 peptide is likely to be essential for the plant (for both GH), the functional marker is based in barley on an In/Del polymorphism in putative *cis*-regulatory regions of VRN-H1 intron 1 (von Zitzewitz et al. 2005). Similar intron 1 deletions of the orthologous genes of the hexaploid and tetraploid wheats are associated with spring alleles suggesting that those deletions are responsible for the differences in GH (Fu et al. 2005). Because coincident with *FR-1*, the same GH molecular marker for *VRN-H1* (*HvBM5*) was also reported as the best predictor for marker-assisted selection within highly frost-tolerant accessions from Turkey and other winter, facultative and spring barley germplasm. The CG marker can thus be used not only for selection of GH types in winter x spring crosses, but also for fast routine selection of frost tolerant genotypes (Akar et al. 2009). Similar indications came from the work of Rapacz et al. (2010) that found variation in the promoter region of *Vrn-H1* (Hv*BM5*) directly connected with freezing tolerance of plants partially de-acclimated in the field.

In a study by Akar et al. (2009), only one out of the three markers designed on *CBF* genes was moderately associated with FT. Polymorphisms used and derived from linkage mapping were in fact not associated with any functional diversity in the transcripts. In barley, it is still not clarified if a distinct CBF element could be causative of the *FR-2* effect on the trait, as tentatively hypothesized by Fricano et al. (2009), or if, more likely, the composition of the gene cluster in terms of copy

number variation (CNV) of at least two elements could constitute the functional difference between freezing-tolerant and susceptible genotypes (Knox et al. 2010). Yet, further studies would be necessary to validate the feasibility of a MAS approach based on single *CBF* elements rather than on the introgression of the whole *CBF* cluster region; especially in winter x winter genotype crosses, where recombinants between *CBF* sub-clusters could be not easy to obtain. The simplest approach to select with the LD-MAS is the introduction of (SNP) functional marker panels into normal pedigree breeding schemes. The CG markers can be introduced as support or substitution of phenotypic selection also in case of normal backcrosses, or doubled-haploid (DH) line evaluation. An interesting application of LD-MAS is the parent building by means of candidate gene pyramiding (Sabatini et al. 2013). Mainly used to cumulate pathogen resistances into a single genotype (Mago et al. 2011), it could also be used to cumulate CGs for freezing and other abiotic stress tolerances.

6.4.2.2 Genomic Selection (GS)

Genomic selection (Meuwissen et al. 2001) could be defined as a method of prediction of the breeding value of lines by analyzing phenotype together with high-density marker scores. Basically, GS simultaneously estimates all locus, haplotype, or marker effects across the entire genome to calculate genomic estimated breeding values (GEBVs). It incorporates all marker information in a prediction model, thereby avoiding biased marker effect estimates and capturing more of the variation due to small-effect QTL. Genomic selection (GS) has been proposed to overcome the limits of application of LD-based MAS to polygenic traits selection. While GS should substantially accelerate the breeding cycle, it would also dramatically change the role of phenotyping, that could be used more to update the prediction models driving GS than to select lines (Heffner et al. 2009).

By using high-density SNP panels, the genotype that would better fit with the genomic prediction model should be selected in order to combine and cumulate the positive effects from all beneficial alleles at contributing genes and minor QTLs. Heffner et al. (2010) demonstrated in a population of 374 winter wheat characterised for 13 agronomic traits that the average prediction accuracies for GS would be 28 % higher than MAS, while 955 as accurate as phenotypic selection. Since a high-density SNP panel could be excessively costly per single analysis, Habier et al. (2009) proposed to use a panel of low-density evenly spaced SNPs. Diffusion of GS in practical breeding thus strongly depends on the increasing availability of cheap high throughput markers systems.

Moreover, GS can be better proposed for species where genomic constitution is known in terms of sequences and their physical position, and for which cultivar resequencing projects are in progress, as in apple (Kumar et al. 2012). In this view, the success in sequencing all gene-containing regions of barley and wheat is a necessary requirement to allow GS-based schemes. Recently, Paux et al. (2012) report that GS methods are under evaluation for crops such as maize and wheat and in some cases are being applied in wheat commercial breeding programs, although details have yet to be published.

Genomic selection could be particularly useful to cumulate durable resistance QTLs in wheat as in other plants. Rutkoski et al. (2011) propose a GS-based wheat breeding scheme for quantitative (durable) resistance to stem rust, where the multigenic nature of adult plant resistance hampers the efficiency of MAS-based pyramiding. Due to the lack of mapped minor QTLs in wheat as in barley, all affecting the final level of freezing tolerance, GS for FT could be an option, together with other agronomically relevant traits. Once high density SNP panels can be made available and at reasonable assay costs, there should not in fact be the need to know all the (minor) QTL positions to select associated markers. Rather, and as proposed by Meuwissen et al. (2001), GS analyzes jointly all markers on a population attempting to explain the total genetic variance with dense, genomewide marker coverage through summing the marker effects to predict the final breeding value of individuals.

6.5 Conclusions and Perspectives

Global climate change poses increasing constraints on the ability of crops to acclimate to abiotic stresses. Cereal breeding for an enhanced FT and winter hardiness will remain an important part of winter cereal breeding programmes in temperate climate zone. New approaches of structural as well as functional genomics will facilitate the identification of candidate genes underlying FT and winter-hardiness QTLs as well as the functional characterization of their protein products. Recently, fine mapping of the cluster of *CBF* transcriptional activators at *Fr-2* locus in Triticeae and their further characterization has revealed genetic differences in *CBF* gene sequence and gene-copy number between frost-tolerant winter and frost-sensitive spring genotypes of *Triticeae* (Knox et al. 2008, 2010). Differences in *TmCBF12* sequence at AP2 CRT/DRE promoter-binding domain between winter line G3116 and spring line DV92 of einkorn wheat and differences in the number of *HvCBF2* and *HvCBF4* paralogues (*HvCBF* gene copy number) between winter barley Nure and spring barley Tremois can be utilised as genetic markers to distinguish frost-tolerant winter and frost-sensitive spring genotypes of *Triticeae* (reviewed in Tondelli et al. 2011). Combination of "omics" approaches with transgenic techniques will help us to identify key factors at all levels (genes, transcripts, proteins and metabolites) involved in the processes of FT acquisition and vernalization. Transgenic techniques will help us to functionally characterize key proteins involved in cold acclimation and vernalization. An example is provided by the progress in understanding the role of the major cereal vernalization gene *VRN1* using *mvp* (*maintained vegetative phase*) mutants of einkorn wheat coding for a non-functional *VRN1* gene (Dhillon et al. 2010). A better understanding of the physiological mechanisms involved in cold acclimation and FT acquisition will lead to breeding progress using modern genetic approaches (e.g. MAS, gene pyramiding, etc.). Modern high-throughput screening methods will enable us to carry out genome-based selection of genotypes with desired sets of alleles of multiple genes regulating stress tolerance as well as grain quality and other desired agronomic traits. The new approaches described herein

will eventually lead to the development of new cultivars with tailored characteristics and with an improved versatility underlying the ability to cope with environmental challenges.

Acknowledgements This work was supported by the European Union (AGRISAFE 203288—EU-FP7-REGPOT 2007-1), by a joint grant of EU and the European Social Fund, by the FROSTMAP project of the Fondazione Cassa di Risparmio di Modena, by the Hungarian Scientific Research Fund and the National Office for Research and Technology (OTKA K 83642, NKTH-OTKA CNK 80781), by the National Development Agency grant TÁMOP-4.2.2/B-10/1-2010-0025 Doctoral School of Molecular- and Nanotechnologies, Faculty of Information Technology, University of Pannonia, by the Czech Ministry of Agriculture (MZe ČR), project no. 0002700604 and by the Grant Agency of the Czech Republic (GA ČR) postdoctoral project P501/11/P637.

References

Achard P, Gong F, Cheminant S et al (2008) The cold-inducible CBF1 factor-dependent signaling pathway modulates the accumulation of the growth-repressing DELLA proteins via its effect on gibberellin metabolism. Plant Cell 20:2117–2129

Acquaah G (2006) Principles of plant genetics and breeding. Blackwell Publishing Ltd., Oxford

Akar T, Francia E, Tondelli A et al (2009) Marker-assisted characterization of frost tolerance in barley (*Hordeum vulgare* L.). Plant Breed 128:381–386

Alberta Agriculture, Food and Development (2001) The Growth Potential of Triticale in Western Canada: Section B—Genetic Basis, Breeding and Varietal Performance of Triticale Genetic Sources and Potential of 21st Century Triticale Germplasm: Past, Current and Future Breeding Goals, Achievements and Limitations. Government of Alberta, Agricultural and Rural Development. http://www1.agric.gov.ab.ca/$department/deptdocs.nsf/all/fcd4232. Accessed 25 February 2013

Amme S, Matros A, Schlesier B, Mock HP (2006) Proteome analysis of cold stress response in *Arabidopsis thaliana* using DIGE-technology. J Exp Bot 57:1537–1546

Badawi M, Danyluk J, Boucho B et al (2007) The CBF gene family in hexaploid wheat and its relationship to the phylogenetic complexity of cereal CBFs. Mol Genet Genomics 277:533–554

Badawi M, Reddy YV, Agharbaoui Z et al (2008) Structure and functional analysis of wheat ICE (Inducer of CBF Expression) genes. Plant Cell Physiol 49:1237–1249

Bae MS, Cho EJ, Choi EY, Park OK (2003) Analysis of the *Arabidopsis* nuclear proteome and its response to cold stress. Plant J 36:652–663

Baga M, Chodaparambil SV, Limin ASE et al (2007) Identification of quantitative trait loci and associated candidate genes for low-temperature tolerance in cold-hardy winter wheat. Funct Integr Genom 7:53–68

Baker SS, Wilhelm KS, Thomashow MF (1994) The 5'-region of *Arabidopsis thaliana* cor15a has cis-acting elements that confer cold-, drought- and ABA-regulated gene expression. Plant Mol Biol 24:701–713

Barrett T, Troup DB, Wilhite SE et al (2011) NCBI GEO: archive for functional genomics data sets—10 years on. Nucleic Acids Res 39:D1005–D1010

Battaglia M, Olvera-Carillo Y, Garciarrubio A, Covarrubias AA (2008) The enigmatic LEA proteins and other hydrophilins. Plant Physiol 48:6–24

Baum M, Grando S, Backes G et al (2003) QTLs for agronomic traits in the Mediterranean environment identified in recombinant inbred lines of the cross 'Arta' x H. spontaneum 41–1. Theor Appl Genet 107:1215–1225

Beales J, Turner A, Griffiths S et al (2007) A *Pseudo-Response Regulator* is misexpressed in the photoperiod insensitive *Ppd-D1a* mutant of wheat (*Triticum aestivum* L.). Theor Appl Genet 115:721–733

Bishop RE (2000) The bacterial lipocalins. Biochim Biophys Acta 1482:73–83

Bocian A, Kosmala A, Rapacz M et al (2011) Differences in leaf proteome response to cold acclimation between Lolium perenne plants with distinct levels of frost tolerance. J Plant Physiol 168:1271–1279

Braun HJ, Săulescu NN (2002) Breeding winter and facultative wheat. In: Curtis BC, Rajaram S, Gomez Macpherson H (eds) FAO Plant production and protection series. Bread wheat. Improvement and production. Food and Agriculture and Food Organization of the United Nations, Rome

Brenchley R, Spannagl M, Pfeifer M et al (2012) Analysis of the bread wheat genome using whole-genome shotgun sequencing. Nature 491:705–710

Buckler DR, Park A, Viswanathan M et al (2008) Screening isolates from antibody phage-display libraries. Drug Discov Today 13:318–324

Campoli C, Matus-Cádiz MA, Pozniak CJ et al (2009) Comparative expression of *Cbf* genes in the Triticeae under different acclimation induction temperatures. Mol Genet Genomics 282:141–152

Cantrel C, Vazquez T, Puyaubert J et al (2011) Nitric oxide participates in cold-responsive phosphosphingolipid formation and gene expression in *Arabidopsis thaliana*. New Phytol 189:415–427

Catala R, Santos E, Alonso JM et al (2003) Mutations in the Ca2+/H+ transporter CAX1 increase CBF/DREB1 expression and the cold-acclimation response in *Arabidopsis*. Plant Cell 15:2940–2951

Cattivelli L, Baldi P, Crosatti C et al (2002) Chromosome regions and stress-related sequences involved in resistance to abiotic stress in Triticeae. Plant Mol Biol 48:649–655

Chen A, Reinheimer J, Brulé-Babel A et al (2009) Genes and traits associated with chromosome 2H and 5H regions controlling sensitivity of reproductive tissues to frost in barley. Theor Appl Genet 8:1465–1476

Cheng L, Gao X, Li S et al (2010) Proteomic analysis of soybean [*Glycine max* (L.) Meer.] seeds during imbibition at chilling temperature. Mol Breed 26:1–17

Chinnusamy V, Ohta M, Kanrar S et al (2003) ICE1: a regulator of cold-induced transcriptome and freezing tolerance in *Arabidopsis*. Genes Dev 17:1043–1054

Chinnusamy V, Schumaker K, Zhu JK (2004) Molecular genetic perspectives on cross-talk and specificity in abiotic stress signalling in plants. J Exp Bot 55:225–236

Chinnusamy V, Zhu J, Zhu JK (2006) Gene regulation during cold acclimation in plants. Physiol Plant 126:52–61

Chinnusamy V, Zhu JK, Sunkar R (2010) Gene regulation during cold stress acclimation in plants. Methods Mol Biol 639:39–55

Choi DW, Zhu B, Close TJ (1999) The barley (*Hordeum vulgare* L.) dehydrin multigene family: sequences, allele types, chromosome assignments, and expression characteristics of 11 Dhn genes of cv. Dicktoo. Theor Appl Genet 98:1234–1247

Choi DW, Rodriguez EM, Close TJ (2002) Barley *Cbf3* gene identification, expression pattern, and map location. Plant Physiol 129:1781–1787

Cockram J, Chiapparino E, Taylor SA et al (2007a) Haplotype analysis of vernalization loci in European barley germplasm reveals novel *VRN-H1* alleles and predominant winter *VRN-H1/VRN-H2* multi-locus haplotype. Theor Appl Genet 115:993–1001

Cockram J, Jones H, Leigh FJ et al (2007b) Control of flowering time in temperate cereals: genes, domestication, and sustainable productivity. J Exp Bot 58:1231–1244

Cockram J, White J, Leigh FJ et al (2008) Association mapping of partitioning loci in barley. BMC Genet 9:16

Cockram J, Norris C, O'Sullivan DM (2009) PCR-based markers diagnostic for spring and winter seasonal growth habit in barley. Crop Sci 49:404–410

Comadran J, Kilian B, Russell J et al (2012) Natural variation in a homolog of *Antirrhinum CENTRORADIALIS* contributed to spring growth habit and environmental adaptation in cultivated barley. Nat Genet 44:1388–1392

Cook D, Fowler S, Fiehn O, Thomashow MF (2004) A prominent role for the CBF cold response pathway in configuring the low-temperature metabolome of *Arabidopsis*. Proc Natl Acad Sci USA 42:15243–15248

Crosatti C, Soncini C, Stanca AM, Cattivelli L (1995) The accumulation of a cold-regulated chloroplastic protein is light-dependent. Planta 196:458–463

Cui S, Huang F, Wang J et al (2005) A proteomic analysis of cold stress responses in rice seedlings. Proteomics 5:3162–3172

Dai X, Xu Y, Ma Q et al (2007) Overexpression of a *R1R2R3 MYB* gene, *OsMYB3R-2*, increases tolerance to freezing, drought, and salt stress in transgenic *Arabidopsis*. Plant Physiol 143:1739–1751

Danyluk J, Kane NA, Breton G et al (2003) TaVRT-1, a putative transcription factor associated with vegetative to reproductive transition in cereals. Plant Physiol 132:1849–1860

Degand H, Faber AM, Dauchot N et al (2009) Proteomic analysis of chicory root identifies proteins typically involved in cold acclimation. Proteomics 9:2903–2907

Dhillon T, Pearce SP, Stockinger EJ et al (2010) Regulation of freezing tolerance and flowering in cereals: the *VRN-1* connection. Plant Physiol 153:1846–1858

Dionne J, Castonguay Y, Nadeau P, Desjardnis Y (2001) Amino acid and protein changes during cold acclimation of green-type annual bluegrass (*Poa annua* L.) ecotypes. Crop Sci 41:1862–1870

Distelfeld A, Li C, Dubcovsky J (2009) Regulation of flowering in temperate cereals. Curr Opin Plant Biol 12:1–7

Doherty CJ, van Buskirk HA, Myers SJ, Thomashow MF (2009) Roles for *Arabidopsis* CAMTA transcription factors in cold-regulated gene expression and freezing tolerance. Plant Cell 21:972–984

Dong CH, Agarwal M, Zhang Y et al (2006) The negative regulator of plant cold responses, HOS1, is a RING E3 ligase that mediates the ubiquiti-nation and degradation of ICE1. Proc Natl Acad Sci USA 103:8281–8286

Dong MA, Farré EM, Thomashow MF (2011) Circadian clock-associated 1 and late elongated hypocotyl regulate expression of the C-repeat binding factor (CBF) pathway in *Arabidopsis*. Proc Natl Acad Sci USA 108:7241–7246

Dörffling K, Schulenburg S, Lesselich G, Dörffling H (1990) Abscisic acid and proline levels in cold-hardened winter wheat leaves in relation to variety-specific differences in freezing resistance. J Agr Crop Sci 165:230–239

Dubcovsky J, Luo MC, Dvorak J (1995) Linkage relationships among stress-induced genes in wheat. Theor Appl Genet 91:795–801

Dubcovsky J, Loukoianov A, Fu D et al (2006) Effect of photoperiod on the regulation of wheat vernalization genes *VRN1* and *VRN2*. Plant Mol Biol 60:469–480

Dubouzet JG, Sakuma Y, Ito Y et al (2003) *OsDREB* genes in rice, *Oryza sativa* L, encode transcription activators that function in drought-, high-salt- and cold-responsive gene expression. Plant J 33:751–763

Dumont E, Bahrman N, Goulas E et al (2011) A proteomic approach to decipher chilling response from cold acclimation in pea (*Pisum sativum* L.). Plant Sci 180:86–98

FAO (2010) The second report on the state of the world's plant genetic resources for food and agriculture. Food and Agriculture and Food Organization of the United Nations, Rome

Faure S, Higgins J, Turner A, Laurie DA (2007) The flowering locus *T-like* gene family in barley (*Hordeum vulgare*). Genet 176:599–609

Fowler DB (2008) Cold acclimation threshold induction temperatures in cereals. Crop Sci 48:1147–1154

Fowler DB, Chauvin LP, Limin AE, Sarhan F (1996a) The regulatory role of vernalization in the expression of low-temperature-induced genes in wheat and rye. Theor Appl Genet 93:554–559

Fowler DB, Limin AE, Wang SY, Ward RW (1996b) Relationship between low-temperature tolerance and vernalization response in wheat and rye. Can J Plant Sci 76:37–42
Fowler S, Thomashow MF (2002) *Arabidopsis* transcriptome profiling indicates that multiple regulatory pathways are activated during cold acclimation in addition to the CBF cold response pathway. Plant Cell 14:1675–1690
Fowler SG, Cook D, Thomashow MF (2005) Low temperature induction of *Arabidopsis* CBF1, 2 and 3 is gated by the circadian clock. Plant Physiol 137:961–968
Francia E, Rizza F, Cattivelli L et al (2004) Two loci on chromosome 5H determine low temperature tolerance in a 'Nure' (winter) x 'Tremois' (spring) barley map. Theor Appl Genet 108:670–680
Francia E, Barabaschi D, Tondelli A et al (2007) Fine mapping of a HvCBF gene cluster at the frost resistance locus *Fr-H2* in barley. Theor Appl Genet 115:1083–1091
Fricano A, Rizza F, Faccioli P et al (2009) Genetic variants of *HvCbf14* are statistically associated with frost tolerance in a European germplasm collection of *Hordeum vulgare*. Theor Appl Genet 119:1335–1348
Fu D, Szücs P, Yan L et al (2005) Large deletions within the first intron in *VRN-1* are associated with spring growth habit in barley and wheat. Mol Genet Genomics 273:54–65
Galiba G (2002) Mapping of genes regulating abiotic stress tolerance in cereals. Acta Agron Hung 50:235–247
Galiba G, Quarrie SA, Sutka J et al (1995) RFLP mapping of the vernalization (*Vrn1*) and frost resistance (*Fr1*) genes on chromosome 5A of wheat. Theor Appl Genet 90:1174–1179
Galiba G, Vágújfalvi A, Li CX et al (2009) Regulatory genes involved in the determination of frost tolerance in temperate cereals. Plant Sci 176:12–19
Ganeshan S, Sharma P, Young L et al (2011) Contrasting cDNA-AFLP profiles between crown and leaf tissues of cold-acclimated wheat plants indicate differing regulatory circuitries for low temperature tolerance. Plant Mol. Biol 75:379–398
Gao F, Zhou Y, Zhu W et al (2009) Proteomic analysis of cold stress-responsive proteins in *Thellungiella* rosette leaves. Planta 230:1033–1046
Gong Z, Lee H, Xiong L et al (2002) RNA helicase-like protein as an early regulator of transcription factors for plant chilling and freezing tolerance. Proc Natl Acad Sci USA 99:11507–11512
Greenup AG, Peacock WJ, Dennis ES, Trevaskis B (2009) The molecular biology of seasonal flowering-responses in *Arabidopsis* and the cereals. Ann Bot (Lond) 103:1165–1172
Greenup AG, Sasani S, Oliver SN et al (2011) Transcriptome analysis of the vernalization response in Barley (*Hordeum vulgare*) Seedlings. PLOS ONE 6:e17900
Griffiths S, Dunford RP, Coupland G, Laurie DA (2003) The evolution of *CONSTANS*-like gene families in barley, rice, and *Arabidopsis*. Plant Physiol 131:1855–1867
Gulick P, Drouin S, Yu Z et al (2005) Transcriptome comparison of winter and spring wheat responding to low temperature. Genome 48:913–923
Guy C, Kaplan F, Kopka J et al (2008) Metabolomics of temperature stress. Physiol Plant 132:220–235
Habier D, Fernando RL, Dekkers JCM (2009) Genomic selection using low-density marker panels. Genetics 182:343–353
Harlan J, de Wet J (1971) Towards a rational classification of cultivated plants. Taxon 20:509–517
Hashimoto M, Komatsu S (2007) Proteomic analysis of rice seedlings during cold stress. Proteomics 7:1293–1302
Heffner EL, Sorrells ME, Jannink JL (2009) Genomic selection for crop improvement. Crop Sci 49:1–12
Heffner EL, Jannink JL, Sorrells ME (2010) Genomic selection accuracy using multifamily prediction models in a wheat breeding program. Plant Genome 4:65–75
Heidarvand L, Maali Amiri R (2010) What happens in plant molecular responses to cold stress? Acta Physiole Plant 32:419–431
Herman EM, Rotter K, Premakumar R et al (2006) Additional freeze hardiness in wheat acquired by exposure to –3°C is associated with extensive physiological, morphological, and molecular changes. J Exp Bot 14:3601–3618

Houde M, Dhindsa RS, Sarhan F (1992) A molecular marker to select for freezing tolerance in *Gramineae*. Mol Gen Genet 234:43–48

Hsieh TH, Lee JT, Yang PT et al (2002) Heterology expression of the *Arabidopsis* C-repeat/dehydration response element binding factor 1 gene confers elevated tolerance to chilling and oxidative stresses in transgenic tomato. Plant Physiol 129:1086–1094

Hussain SS (2006) Barley genetics and genomics: a review. Proc Pak Acad Sci 43:63–84

Imin N, Kerim T, Rolfe BG, Weinman JJ (2004) Effect of early cold stress on the maturation of rice anthers. Proteomics 4:1873–1882

Jaglo-Ottosen KR, Gilmour SJ, Zarka DG et al (1998) *Arabidopsis* CBF1 overexpression induces *COR* genes and enhances freezing tolerance. Science 280:104–106

Johnston PA, Timmerman-Vaughan GM, Farnden KJF, Pickering R (2009) Marker development and characterisation of *Hordeum bulbosum* introgression lines: a resource for barley improvement. Theor Appl Genet 118:1429–1437

Kane NA, Danyluk J, Tardif G et al (2005) *TaVRT-2*, a member of the *St*-MADS-11 clade of flowering repressors, is regulated by vernalization and photoperiod in wheat. Plant Physiol 138:2354–2363

Kane NA, Agharbaoui Z, Diallo AO et al (2007) TaVRT2 represses transcription of the wheat vernalization gene *TaVRN1*. Plant J 51:670–680

Kaplan F, Kopka J, Haskell DW et al (2004) Exploring the temperature-stress metabolome of *Arabidopsis*. Plant Physiol 136:4159–4168

Kawamura Y, Uemura M (2003) Mass spectrometric approach for identifying putative plasma membrane proteins of *Arabidopsis* leaves associated with cold acclimation. Plant J 36:141–154

Kidokoro S, Maruyama K, Nakashima K et al (2009) The phytochrome-interacting factor PIF7 negatively regulates DREB1 expression under circadian control in *Arabidopsis*. Plant Physiol 151:046–2057

Kim TS, Kim SK, Han TJ et al (2002) ABA and polyamines act independently in primary leaves of cold-stressed tomato (*Lycopersicon esculentum*). Physiol Plant 115:370–376

Knight H (2000) Calcium signaling during abiotic stress in plants. Int Rev Cytol 195:269–324

Knight H, Trewavas AJ, Knight MR (1996) Cold calcium signaling in *Arabidopsis* involves two cellular pools and a change in calcium signature after acclimation. Plant Cell 8:489–503

Knox AK, Li C, Vágújfalvi A, Galiba G et al (2008) Identification of candidate *CBF* genes for the frost tolerance locus *Fr-A^{m}2* in *Triticum monococcum*. Plant Mol Biol 67:257–270

Knox AK, Dhillon T, Cheng H et al (2010) *CBF* gene copy number variation at *Frost Resistance-2* is associated with levels of freezing tolerance in temperate-climate cereals. Theor Appl Genet 121:21–35

Kobayashi F, Takumi S, Kume S et al (2005) Regulation by *Vrn-1*/*Fr-1* chromosomal intervals of CBF-mediated *Cor*/*Lea* gene expression and freezing tolerance in common wheat. J Exp Bot 56:887–895

Kocsy G, Galiba G, Brunold C (2001) Role of glutathione in adaptation and signalling during chilling and cold acclimation in plants. Physiol Plant 113:158–164

Kocsy G, Szalai G, Vágújfalvi A et al (2000) Genetic study of glutathione accumulation during cold hardening in wheat. Planta 210:295–301

Kocsy G, Athmer B, Perovic D et al (2010) Regulation of gene expression by chromosome 5A during cold hardening in wheat. Mol Genet Genomics 283:351–363

Kosmala A, Bocian A, Rapacz M et al (2009) Identification of leaf proteins differentially accumulated during cold acclimation between *Festuca pratensis* plants with distinct levels of frost tolerance. J Exp Bot 60:3595–3609

Kosová K, Vítámvás P, Prášil IT (2007) The role of dehydrins in plant response to cold. Biol Plant 51:601–617

Kosová K, Holková L, Prášil IT et al (2008a) Expression of dehydrin 5 during the development of frost tolerance in barley (*Hordeum vulgare*). J Plant Physiol 165:1142–1151

Kosová K, Prášil IT, Vítámvás P (2008b) The relationship between vernalization- and photoperiodically-regulated genes and the development of frost tolerance in wheat and barley. Biol Plant 52:601–615

Kosová K, Prášil IT, Vítámvás P (2010) Role of dehydrins in plant stress response. In: Pessarakli M (ed) Handbook of plant and crop stress. 3rd edn. CRC Press, Taylor & Francis, Boca Raton pp 239–285

Kosová K, Vítámvás P, Prášil IT, Renaut J (2011) Plant proteome changes under abiotic stress—contribution of proteomics studies to understanding plant stress response. J Proteomics 74:1301–1322

Kosová K, Prášil IT, Vítámvás P et al (2012) Complex phytohormone responses during the cold acclimation of two wheat cultivars differing in cold tolerance, winter Samanta and spring Sandra. J Plant Physiol 169:567–576

Kovács Z, Simon-Sarkadi L, Sovány C et al (2011) Differential effects of cold acclimation and abscisic acid on free amino acid composition in wheat. Plant Sci 180:61–68

Kovács Z, Simon-Sarkadi L, Szücs A, Kocsy G (2010) Differential effects of cold, osmotic stress and abscisic acid on polyamine accumulation in wheat. Amino Acids 38:623–631

Kume S, Kobayashi F, Ishibashi M et al (2005) Differential and coordinated expression of *Cbf* and *Cor/Lea* genes during long-term cold acclimation in two wheat cultivars showing distinct levels of freezing tolerance. Genes Genet Syst 80:185–197

Kumar S, Bink M, Volz RK et al (2012) Towards genomic selection in apple (Malus × domestica Borkh.) breeding programmes: Prospects, challenges and strategies. Tree Genetics Genomes 8:1–14

Laudencia-Chingcuanco D, Ganeshan S, You F et al (2011) Genome-wide gene expression analysis supports a developmental model of low temperature tolerance gene regulation in wheat (*Triticum aestivum* L.). BMC Genomics 12:299

Laurie DA, Pratchett N, Bezant JH, Snape JW (1994) Genetic analysis of a photoperiod response gene on the short arm of chromosome 2(2 H) of barley (*Hordeum vulgare* L.). Heredity 72:619–627

Laurie DA, Pratchett N, Bezant JH, Snape JW (1995) RFLP mapping of five major genes and eight quantitative trait loci controlling flowering time in a winter × spring barley (*Hordeum vulgare* L.) cross. Genome 38:575–585

Limin A, Corey A, Hayes P, Fowler DB (2007) Low-temperature acclimation of barley cultivars used as parents in mapping populations: response to photoperiod, vernalization and phenological development. Planta 226:139–146

Limin AE, Fowler DB (1991) Breeding for cold hardiness in winter wheat: problems, progress and alien gene expression. Field Crop Res 27:201–218

Limin AE, Fowler DB (1993) Inheritance of cold hardiness in *Triticum aestivum* × synthetic hexaploid wheat crosses. Plant Breed 110:103–108

Loukoianov A, Yan L, Blechl A et al (2005) Regulation of *VRN-1* vernalization genes in normal and transgenic polyploid wheat. Plant Physiol 138:2364–2373

Macháčková I, Hanišová A, Krekule J (2006) Levels of ethylene, ACC, MACC, ABA and proline as indicators of cold hardening and frost resistance in winter wheat. Physiol Plant 76:603–607

Mago R, Lawrence GJ, Ellis JG (2011) The application of DNA marker and doubled-haploid technology for stacking multiple stem rust resistance genes in wheat. Mol Breed 27:329–335

Maruyama K, Takeda M, Kidokoro S et al (2009) Metabolic pathways involved in cold acclimation identified by integrated analysis of metabolites and transcripts regulated by DREB1A and DREB2A. Plant Physiol 150:1972–1980

Mayer KFX, Waugh R, Langridge P et al (2012) A physical, genetic and functional sequence assembly of the barley genome. Nature 491:711–716

Mazzucotelli E, Tartari A, Cattivelli L, Forlani G (2006) Metabolism of GABA acid during cold acclimation and freezing and its relationship to frost tolerance in barley and wheat. J Exp Bot 57:3755–3766

Meuwissen THE, Karlsen A, Lien S et al (2001) Fine mapping of a quantitative trait locus for twinning rate using combined linkage and linkage disequilibrium mapping. Genetics 161:373–379

Miller AK, Galiba G, Dubcovsky J (2006) A cluster of 11 *CBF* transcription factors is located at the frost tolerance locus *Fr-A^{m}2* in *Triticum monococcum*. Mol Genet Genomics 275:193–203

Miura K, Jin JB, Lee J et al (2007) SIZ1-mediated sumoylation of ICE1 controls CBF3/DREB1A expression and freezing tolerance in *Arabidopsis*. Plant Cell 19:1403–1414

Mochida K, Uehara-Yamaguchi Y, Yoshida T et al (2011) Global landscape of a co-expressed gene network in Barley and its application to gene discovery in triticeae crops. Plant Cell Physiol 52:785–803

Monroy AF, Labbé E, Dhindsa RS (1997) Low temperature perception in plants: effects of cold on protein phosphorylation in cell-free extracts. FEBS Lett 410:206–209

Monroy AF, Dryanova A, Malette B et al (2007) Regulatory gene candidates and gene expression analysis of cold acclimation in winter and spring wheat. Plant Mol Biol 64:409–423

Murata N, Los DA (1997) Membrane fluidity and temperature perception. Plant Physiol 115:875–879

Nadeau P, Delaney S, Chouinard L (1987) Effects of cold hardening on the regulation of polyamine levels in wheat (*Triticum aestivum* L.) and alfalfa (*Medicago sativa* L.). Plant Physiol 84:73–77

Nakashima K, Yamaguchi-Shinozaki K (2006) Regulons involved in osmotic stress-responsive and cold stress-responsive gene expression in plants. Physiol Plant 126:62–71

Nakashima K, Ito Y, Yamaguchi-Shinozaki K (2009) Transcriptional regulatory networks in response to abiotic stresses in *Arabidopsis* and grasses. Plant Physiol 149:88–95

Neilson KA, Gammulla CG, Mizraei M et al (2010) Proteomic analysis of temperature stress in plants. Proteomics 10:828–845

Nevo E (2011) *Triticum*. In: Kole C (ed) Wild crop relatives: Genomic and breeding resources: Cereals. Springer-Verlag, Berlin, pp 407–456

Oliver SN, Finnegan EJ, Dennis ES et al (2009) Vernalization-induced flowering in cereals is associated with changes in histone methylation at the *VERNALIZATION1* gene. Proc Natl Acad Sci USA 106:8386–8391

Ouellet F, Vazquez-Tello A, Sarhan F (1998) The wheat *wcs120* promoter is cold-inducible in both monocotyledonous and dicotyledonous species. FEBS Lett 423:324–328

Paux E, Sourdille P, Mackay I, Feuillet C (2012) Sequence-based marker development in wheat: Advances and applications to breeding. Biotechnol Adv 30:1071–1088

Pecchioni N, Milc JA, Pasquariello M, Francia E (2012) Barley: Omics approaches for abiotic stress tolerance. In: Tuteja N, Gill SS, Tiburcio AF, Tuteja R (eds) Improving crop resistance to abiotic stress. Wiley-VCH, Weinheim, pp 777–882

Penfield S (2008) Temperature perception and signal transduction in plants. New Phytol 179:615–628

Pennycooke JC, Cheng H, Roberts SM et al (2008) The low temperature-responsive, *Solanum CBF1* genes maintain high identity in their upstream regions in a genomic environment undergoing gene duplications, deletions, and rearrangements. Plant Mol Biol 67:483–497

Pickering R, Johnston PA (2005) Recent progress in barley improvement using wild species of *Hordeum*. Cytogenet Genome Res 109:344–349

Pourkheirandish M, Komatsuda T (2007) The importance of barley genetics and domestication in a global perspective. Ann Bot 100:999–1008

Queitsch C, Sangster TA, Lindquist S (2002) Hsp90 as a capacitor of phenotypic variation. Nature 417:618–624

Quesada V, Dean C, Simpson CG (2005) Regulated RNA processing in the control of *Arabidopsis* flowering. Int J Dev Biol 49:773–780

Rabbani A, Maruyama K, Abe H et al (2003) Monitoring expression profiles of rice genes under cold, drought, and high-salinity stresses and abscisic acid application using cDNA microarray and RNA gel-blot analyses. Plant Physiol 133:1755–1767

Ramalingam J, Pathan MS, Feril O et al (2006) Structural and functional analysis of the wheat genomes based on expressed sequence tags (ESTs) related to abiotic stresses. Genome 49:1324–1340

Rapacz M, Tyrka M, Gut M, Mikulski W (2010) Associations of PCR markers with freezing tolerance and photosynthetic acclimation to cold in winter barley. Euphytica 175:293–301

Reddy ASN, Ali GSA, Celesnik H, Day IS (2011) Coping with stresses: roles of calcium- and calcium/calmodulin-regulated gene expression. Plant Cell 23:2010–2032

Reinheimer JL, Barr AR, Eglinton JK (2004) QTL mapping of chromosomal regions conferring reproductive frost tolerance in barley (*Hordeum vulgare* L.). Theor Appl Genet 109:1267–1274

Renaut J, Hausman JF, Bassett C et al (2008) Quantitative proteomic analysis of short photoperiod and low-temperature responses in bark tissues of peach (*Prunus persica* L. Batsch). Tree Genet Genomes 4:589–600

Rinalducci S, Egidi MG, Karimzadeh G et al (2011a) Proteomic analysis of a spring wheat cultivar in response to prolonged cold stress. Electrophoresis 32:1807–1818

Rinalducci S, Egidi MG, Mahfoozi S et al (2011b) The influence of temperature on plant development in a vernalization-requiring winter wheat: a 2-DE based proteomic investigation. J Proteomics 74:643–659

Rizza F, Baldi P, Cattivelli L, Delogu G (1997) Cold hardening in triticale in comparison with rye and wheat. Cereal Res Commun 25:947–954

Rizza F, Pagani D, Gut M et al (2011) Diversity in the response to low temperature in representative barley genotypes cultivated in Europe. Crop Sci 51:2759–2779

Roberts DWA (1990) Identification of loci on chromosome 5A of wheat involved in control of cold hardiness, vernalization, leaf length, rosette growth habit, and height of hardened plants. Genome 33:247–259

Rutkoski JE, Heffner EL, Sorrells ME (2011) Genomic selection for durable stem rust resistance in wheat. Euphytica 179:161–173

Sabatini E, Beretta M, Sala T et al (2013) Molecular breeding. In: Liedl BE, Labate JA, Stommel JR, Slade A, Kole C (eds) Tomato, Genomics of fruit and vegetable crops series. Science Publishers Inc., Enfield

Sandre SR, Kosmala A, Rudi H et al (2011) Molecular mechanisms underlying frost tolerance in perennial grasses adapted to cold climates. Plant Sci 180:69–77

Sarhadi E, Mahfoozi S, Hosseini SA, Salekdeh GH (2010) Cold acclimation proteome analysis reveals close link between the up-regulation of low-temperature associated proteins and vernalization fulfillment. J Proteome Res 9:5658–5667

Sasaki E, Takahashi C, Asami T, Shimada Y (2010) AtCAST, a tool for exploring gene expression similarities among DNA microarray experiments using networks. Plant Cell Physiol 52:169–180

Savitch LV, Harney T, Huner NPA (2000) Sucrose metabolism in spring and winter wheat in response to high irradiance, cold stress and cold acclimation. Physiol Plant 108:270–278

Sears ER (1953) Nullisomic analysis in common wheat. Am Nat 87:245–252

Seki M, Narusaka M, Abe H, Kasuga M et al (2001) Monitoring the expression pattern of 1300 *Arabidopsis* genes under drought and cold stresses by using a full length cDNA microarray. Plant Cell 13:61–72

Shitsukawa N, Ikari C, Shimada S et al (2007) The einkorn wheat (*Triticum monococcum*) mutant, *maintained vegetative phase*, is caused by a deletion in the *VRN1* gene. Genes Genet Syst 82:167–170

Skinner DZ (2009) Post-acclimation transcriptome adjustment is a major factor in freezing tolerance of winter wheat. Funct Integr Genomics 9:513–523

Skinner JS, von Zitzewitz J, Szucs P et al (2005) Structural, functional, and phylogenetic characterization of a large CBF gene family in barley. Plant Mol Biol 59:533–551

Skinner JS, Szücs P, von Zitzewitz J et al (2006) Mapping of barley homologs to genes that regulate low temperature tolerance in *Arabidopsis*. Theor Appl Genet 112:832–842

Sreenivasulu N, Sopory SK, Kavi Kishor PB (2007) Deciphering the regulatory mechanisms of abiotic stress tolerance in plants by genomic approaches. Gene 388:1–13

Steffenson BJ, Olivera P, Roy JK et al (2007) A walk on the wild side: mining wild wheat and barley collections for rust resistance genes. Aust J Agr Res 58:532–544

Stockinger EJ, Gilmour SJ, Thomashow MF (1997) *Arabidopsis thaliana* CBF1 encodes an AP2 domain-containing transcriptional activator that binds to the C-repeat/DRE, a cis-acting DNA regulatory element that stimulates transcription in response to low temperature and water deficit. Proc Natl Acad Sci USA 94:1035–1040

Stockinger EJ, Skinner JS, Gardner KG et al (2007) Expression levels of barley *Cbf* genes at the Frost resistance-*H2* locus are dependent upon alleles at *Fr-H1* and *Fr-H2*. Plant J 51:308–321

Sutka J (1994) Genetic control of frost tolerance in wheat. Euphytica 77:277–282

Suzuki I, Los DA, Kanesaki Y et al (2000) The pathway for perception and transduction of low-temperature signals in *Synechocystis*. EMBO J 19:1327–1334

Svensson JT, Crosatti C, Campoli C et al (2006) Transcriptome analysis of cold acclimation in barley *Albina* and *Xantha* mutants. Plant Physiol 141:257–270

Thomashow MF (1999) Plant cold acclimation: freezing tolerance genes and regulatory mechanisms. Annu Rev Plant Physiol Plant Mol Biol 50:571–599

Thomashow MF (2010) Molecular basis of plant cold acclimation: insights gained from studying the CBF cold response pathway. Plant Physiol 154:571–577

Tondelli A, Pagani D, Rizza F et al (2009) Association mapping of frost tolerance QTL in barley. 19th International Triticeae Mapping Initiative 3rd COST Tritigen Clermont-Ferrand, France August 31st—September 4th 2009

Tondelli A, Francia E, Barabaschi D et al (2011) Inside the *CBF* locus in *Poaceae*. Plant Sci 180:39–45

Tóth B, Galiba G, Fehér E et al (2003) Mapping genes affecting flowering time and frost resistance on chromosome 5B of wheat. Theor Appl Genet 107:509–514

Trevaskis B, Hemming MN, Peacock WJ, Dennis ES (2006) *HvVRN2* responds to daylength, whereas *HvVRN1* is regulated by vernalization and developmental status. Plant Physiol 140:1397–1405

Tuberosa R, Galiba G, Sanguineti MC et al (1997) Identification of QTL influencing freezing tolerance in barley. Acta Agron Hung 45:413–417

Tuberosa R, Gill BS, Quarrie S (2002) Cereal genomics: ushering in a brave new world. Plant Mol Biol 48:445–449

Tunnacliffe A, Wise MJ (2007) The continuing conundrum of the LEA proteins. Naturwissenschaften 94:791–812

Turner A, Beales J, Faure S et al (2005) The pseudo-response regulator Ppd-H1 provides adaptation to photoperiod in barley. Science 310:1031–1034

Vágújfalvi A, Kerepesi I, Galiba G et al (1999) Frost hardiness depending on carbohydrate changes during cold acclimation in wheat. Plant Sci 144:85–92

Vágújfalvi A, Crosatti C, Galiba G et al (2000) Two loci on wheat chromosome 5A regulate the differential cold-dependent expression of the *cor14b* gene in frost-tolerant and frost-sensitive genotypes. Mol Gen Genet 263:194–200

Vágújfalvi A, Galiba G, Cattivelli L et al (2003) The cold-regulated transcriptional activator *Cbf3* is linked to the frost-tolerance gene *Fr-A2* on wheat chromosome 5A. Mol Genet Genom 269:60–67

Vágújfalvi A, Aprile A, Miller A et al (2005) The expression of several *Cbf* genes at the *Fr-A2* locus linked to frost resistance in wheat. Mol Genet Genom 274:506–514

Van Buskirk HA, Thomashow MF (2006) *Arabidopsis* transcription factors regulating cold acclimation. Physiol Plant 126:72–80

Van derSC, Rinne PL (2011) Dormancy cycling at the shoot apical meristem: transitioning between self-organization and self-arrest. Plant Sci 180:120–131

Vannini C, Locatelli F, Bracale M et al (2004) Overexpression of the rice *Osmyb4* gene increases chilling and freezing tolerance of *Arabidopsis thaliana* plants. Plant J 37:115–127

Varshney RK, Graner A, Sorrells ME (2005) Genomics-assisted breeding for crop improvement. Trends Plant Sci 10:621–630

Vij S, Tyagi AK (2007) Emerging trends in the functional genomics of the abiotic stress response in crop plants. Plant Biotech J 3:361–380

Vítámvás P, Prášil IT (2008) WCS120 protein family and frost tolerance during cold acclimation, deacclimation and reacclimation of winter wheat. Plant Physiol Biochem 46:970–976
Vítámvás P, Saalbach G, Prášil IT et al (2007) WCS120 protein family and proteins soluble upon boiling in cold-acclimated winter wheat. J Plant Physiol 164:1197–1207
Vítámvás P, Prášil IT, Kosová K et al (2012) Analysis of proteome and frost tolerance in chromosome 5A and 5B reciprocal substitution lines between two winter wheats during long-term cold acclimation. Proteomics 12:68–85
Vlachonasios KE, Thomashow MF, Triezenberg SJ (2003) Disruption Mutations of *ADA2b* and *GCN5* Transcriptional Adaptor Genes Dramatically Affect *Arabidopsis* Growth, Development, and Gene Expression. Plant Cell 15:626–638
Vogel JT, Zarka DG, van Buskirk HA et al (2005) Roles of the CBF2 and ZAT12 transcription factors in configuring the low temperature transcriptome of *Arabidopsis*. Plant J 41:195–211
von Zitzewitz J, Szücs P, Dubcovsky J et al (2005) Molecular and structural characterization of barley vernalization genes. Plant Mol Biol 59:449–467
Winfield MO, Lu C, Wilson ID et al (2009) Cold- and light-induced changes in the transcriptome of wheat leading to phase transition from vegetative to reproductive growth. BMC Plant Biol 9:55
Winfield MO, Lu C, Wilson ID et al (2010) Plant responses to cold: transcriptome analysis of wheat. Plant Bitechnol J 8:749–771
Xin Z, Mandaokar A, Chen J et al (2007) *Arabidopsis* ESK1 encodes a novel regulator of freezing tolerance. Plant J 21:786–799
Xing L, Li J, Xu Y et al (2009) Phosphorylation modification of wheat lectin VER2 is associated with vernalization-induced O-GlcNAc signaling and intracellular motility. PLoS ONE 4:e4854
Xu ZS, Xia LQ, Chen M et al (2007) Isolation and molecular characterization of the *Triticum aestivum* L. ethylene-responsive factor 1 (TaERF1) that increases multiple stress tolerance. Plant Mol Biol 65:719–732
Xue GP (2003) The DNA-binding activity of an AP2 transcriptional activator HvCBF2 involved in regulation of low-temperature responsive genes in barley is modulated by temperature. Plant J 33:373–383
Yamaguchi-Shinozaki K, Shinozaki K (2006) Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses. Annu Rev Plant Biol 57:781–803
Yan L, Loukoianov A, Blechl A et al (2004) The wheat *VRN2* gene is a flowering repressor down-regulated by vernalization. Science 303:1640–1644
Yan SP, Zhang QY, Tang ZC et al (2006) Comparative proteomic analysis provides new insights into chilling stress responses in rice. Mol Cell Proteomics 5:484–496
Zarka DG, Vogel JT, Cook D et al (2003) Cold induction of *Arabidopsis CBF* genes involves multiple *ICE1* (inducer of *CBF* expression 1) promoter elements and a cold-regulatory circuit that is desenzitized by low temperature. Plant Physiol 133:910–918
Zhang JZ, Creelman RA, Zhu JK (2004) From laboratory to field. Using Information from Arabidopsis to Engineer Salt, Cold, and Drought Tolerance in Crops. Plant Physiol 135:615–621
Zhu J, Shi H, Lee B et al (2004) An *Arabidopsis* homeodomain transcription factor gene, HOS9, mediates cold tolerance through a CBF-independent pathway. Proc Natl Acad Sci USA 26:9873–9878
Zhu J, Verslues PE, Zheng X et al (2005) HOS10 encodes an R2R3-type MYB transcription factor essential for cold acclimation in plants. Proc Natl Acad Sci USA 102:9966–9971

Chapter 7
Bridging Conventional Breeding and Genomics for A More Sustainable Wheat Production

P. Stephen Baenziger, Ali Bakhsh, Aaron Lorenz and Harkamal Walia

Abstract Conventional breeding has a long history of success, while the field of genomics and its applications offer new ways to more efficiently breed new cultivars. In this chapter, using wheat (*Triticum spp*.) as an example, we review the current understanding of wheat breeding and two of its key aspects: The use and creation of genetic diversity (intercrossing elite lines or crossing elite lines with unadapted lines or wild species; creating mutations; and inserting transgenes), and the ability to phenotypically or genetically select for useful combinations of alleles to create improved lines. The strengths of conventional breeding and genomic –assisted breeding are described, as are their limitations, and how a modern breeding program will adapt and integrate new breeding tools with proven methods. The introduction of new alleles into a population will require estimates of their breeding values and epistasis before they can be effectively selected using genomic selection. Similarly, the importance of genotype–by environment interactions will require extensive field testing and more sophisticating genomic selection models to identify lines for the target set of environments.

Once the breeding objective has been chosen, there are three critical phases of a wheat (*Triticum ssp*). breeding program, namely, the introduction of variation, inbreeding often coupled with continuous or later generation selection, and extensive evaluation to determine where the resultant advanced lines should or should not be grown (Baenziger et al. 2006b; Baenziger and DePauw 2009). For the purpose of this chapter, the breeding objective will be to develop wheat cultivars that are suitable for sustainable wheat production. Though it is simple to write the breeding objective of

P. S. Baenziger (✉) · A. Lorenz · H. Walia
Department of Agronomy and Horticulture, University of Nebraska-Lincoln,
Lincoln, NE, USA 68583–0915
e-mail: pbaenziger1@unl.edu

A. Bakhsh
College of Agriculture Dera Ghazi Khan Sub-campus
University of Agriculture Faisalabad, Pakistan
Formerly at Department of Agronomy and Horticulture,
University of Nebraska-Lincoln, Lincoln, NE, USA 68583–0915

R. Tuberosa et al. (eds.), *Genomics of Plant Genetic Resources,*
DOI 10.1007/978-94-007-7575-6_7,

developing new cultivars, in practice there are many approaches to achieve this common objective. As this chapter is part of book discussing the importance of genetic resources, the most important distinction will be between prebreeding (developing the improved parental lines that are needed to more efficiently create new cultivars) and applied breeding which actually creates new cultivars. Prebreeding is the bridge between the vast germplasm collections which store the genetic resources for wheat and the cultivar development programs whose goal is to release those extremely rare lines that combine the best attributes from those resources into a single line which becomes a cultivar or an inbred parent for hybrid wheat.

Wheat is the most widely grown crop in the world and clearly sustainability will have different definitions (social, economic, and ecological) depending upon the region where wheat is grown and produced. Regardless of the precise definition of sustainability, the core breeding methodologies will be similar and the three phases remain the same. In fact, the greatest determinant of how wheat breeders do their work is often the resources available for them to do their work. For example, the pedigree breeding method is preferred, but labor intensive, hence is used widely by programs that have ample labor resources. While other programs that have excellent machinery, but for which labor is very expensive, often breed using the bulk breeding program with its rapid planting and harvesting. Where time is critical, plant breeders use single seed descent or doubled haploidy that can greatly reduce the time needed for inbreeding particularly in winter wheat.

7.1 The Importance of Genetic Variation and Genetic Resources

To understand the importance of genetic variation there is no more important equation than the equation that estimates the predicted genetic gain from selection (or the response from selection; Fehr 1987; Falconer and McKay 1996; Moose and Mumm 2008):

$$\Delta G = R = ih^2\sigma_p = ih\sigma_a = i\sigma_a^2/\sigma_p$$

Where ΔG is the genetic gain and synonymous with R, the response to selection; i = selection differential, h^2 is the narrow sense heritability of the trait and equals the additive genetic variance (σ_a^2) divided by the phenotypic variance (σ_p^2). The relationship between the additive genetic variance and the phenotypic genetic variance is:

$$\sigma_p^2 = \sigma_a^2 + \sigma_d^2 + \sigma_I^2 + \sigma_e^2$$

Where σ_d^2 is the dominance genetic variance, σ_I^2 is the epistatic genetic variance, and σ_e^2 is the error variance. The additive genetic variance refers to the effect of the allele at a locus. The dominance genetic variance refers to the interaction of two alleles at a locus (e.g. AA, Aa, or aa in a diploid or allopolyploid). The epistatic genetic variance refers to how two or more loci interact with each other. Finally the error variance relates to how well you can measure the trait. In self-pollinated crops, the additive genetic variance is the most important because it estimates the effect of a

single allele (only one allele is passed from a parent to the progeny in a diploid or allopolyploid) followed by epistatic variation. Through inbreeding, the variation due to the heterozygote (Aa) approaches zero, hence the dominance genetic variation is only important in the early generations of selection when there may be considerable heterozgosity. The above equation is simplistic and can be expanded to include terms for measuring trait(s) in multiple environments and their interaction with the above genetic variance. The key point of the explaining the phenotypic variance is that it is based upon the genetic variances and the environmental and error variances.

The genetic gain equation includes various mathematical manipulations and the last part is the most illustrative. In simple terms, there can be no genetic improvement without genetic variation, and the ability to select is determined by how much of the phenotypic variation is due to genetic variation. The genetic gain equation explains why breeders are so interested in and so dependent upon genetic variation.

Two final aspects of the genetic gain equation are important to understand before we can discuss the importance and use of genetic variation. The first aspect is that **ΔG** is the difference from the progeny of the selected population mean and the progeny of the unselected population mean. Hence to develop improved cultivars, the variation is important, but also the population mean. Herein lays the tradeoff that all breeders face. To maximize genetic variation, one would cross very diverse parents. However, diverse parents invariably reduce the population mean (Moose and Mumm 2008). To maintain a high population mean, one would cross less diverse (narrow) parents, but you limit the potential for genetic gain. Finding the right amount of genetic diversity relative to maintaining the high population mean is critical and the basis for prebreeding as it relates to applied wheat breeding. The second aspect is that the above discussion is the theory of plant breeding. In practice, wheat breeders are limited by their resources, which in turn limit the number of populations they can create, effectively select in, and the number of replications or environments they can use for testing (Bos 1983; Huehn 1996). For example, the smallest complete population (the population that represents all the genotypes and in the correct proportion) in an F_2 generation is 4^m and in a doubled haploid population is 2^m where m is the number of segregating loci (Sneep 1977; Baker 1984; DePauw et al. 2007; Baenziger and DePauw 2009). If m is large, as would be expected in diverse crosses, the population size becomes enormous, the number of segregating populations that can be worked with very small, and the ability to find the best genotype beyond the levels of breeder or statistical detection. However, in these cases recurrent selection is often recommended.

7.2 Sources of Variation

7.2.1 Sexual Hybridization

There are three major ways of adding genetic variation to wheat. The first is through crossing (sexual hybridization) which is the most common method. The second is through genetic mutations which can be induced by chemical or irradiation treatments, or by genetic mechanisms (e.g. transposable elements, Feschotte et al. 2002).

The third method is by transformation where a new gene is added to the wheat genome. In this chapter, we will discuss in greatest detail genetic variation created by crossing, as well as briefly new methods involving mutations for creating and identifying allelic variation (Robertson 1985; McCallum et al. 2000, Guzman et al. 2011).

As mentioned previously wheat is the most widely grown crop in the world which is a testament to its genetic diversity and genetic resources. Common or bread wheat (*T. aestivum* L.) is a hexaploid (2n = 6x = 42 and contains three genomes, AABBDD). Durum wheat (*T. durum* L.) is a tetraploid (2n = 4x = 28 and contains two genomes, AABB). The genetic resources of wheat (reviewed by Peng et al. 2011) include the cultivated species, the progenitor diploid species (those that contributed the A, B, or D genome), and their relatives which include barley (*Hordeum vulgare* L., 2n = 2x = 14 and contains one genome, HH) and rye (*Secale cereale* L., 2n = 2x = 14 and contains one genome, RR). These genetic resources are accessed or introgressed through crossing, though for some crosses additional tools such as embryo rescue or tissue culture are required. The origin of hexaploid and tetraploid wheat occurred over time and it is estimated that the progenitor of durum wheat (the hybrid of the A and B genomes) occurred about 500,000 years ago. The origin of common wheat is much more recent and it is believed to have occurred between 8 and 10,000 years ago (the hybrid of AABB and the DD; Huang et al. 2002). Furthermore, as the hybrid was recent and there were few successful crosses in nature, the genetic variation in the D genome is less that of the A and B genomes. For this reason, modern plant breeders have recreated hexaploid wheat by crossing durum wheat with D genome progenitors and doubling the resulting hybrid to become AABBDD. These hybrids are known as synthetic wheat (Mujeeb-Kazi et al. 1996).

Probably the most important aspect of using the vast genetic resources found in wheat is chromosome pairing. If the chromosomes pair, then genes can be easily moved into the cultivated wheat genome. However, if the chromosomes do not pair as might be expected in intergeneric crosses, then some form of chromosomal breakage must occur to transfer the genes to wheat chromosomes where they can be incorporated into cultivars. Chromosome breakage and the formation of translocations is usually accomplished using lines deficient in the pairing inhibitor genes as a parent or irradiation mutagenesis (e.g. Sears 1976, 1993).

An example of chromosome breakage leading to a translocation is that many of the modern bread wheat cultivars have a centric rye-wheat translocation 1RS.1BL instead of the normal chromosome 1B (Braun et al. 1998). The rye translocation is a good instance of beneficial alleles or genes derived from rye. The 1RS.1BL translocation is known to provide resistance to several fungi and insects (Zeller and Hsam 1984). In addition, recent work has shown that 1RS.1BL also enhanced yield most likely through increased root biomass (Ehdaie et al. 2003). There are several other instances of beneficial traits incorporated into wheat cultivars from related species. However, the challenge for genomics and plant breeders remains to identify the specific gene(s) that are contributing to these desired traits. Due to the relatively large size of translocations, genomic efforts have led to identification of genes that are introgressed due the translocation. One such approach used wheat microarray platform for predicting the genes that are introduced by the 1RS.1BL

segment (Bhat et al. 2007). The array-based mapping approach was shown to be more cost effective than some of the PCR-based methods. With the advent of next-generation sequencing, the potential to identify alien genes contributing to a given trait will increase tremendously. Because of the challenges in assembly of short reads from genomic DNA, the transcriptome can be used as proxy for ascertaining the expressed genes derived from the introduced translocation. Such information will not only be valuable in providing functional basis of trait improvement, but also can be leveraged to develop sequence based markers for breeding programs that will be needed if the translocation must be reduced in size to eliminate linked deleterious genes.

7.2.2 *Modern Mutation Approaches for Creating Genetic Variation: TILLING (Targeting Induced Local Lesions IN Genomes)*

Though there are many ways of creating and detecting mutations in plants (Uauy et al. 2009), TILLING is rapidly becoming one favored by genomicists. TILLING is a reverse genetic technique that combines chemical mutagenesis to create a mutant population that is coupled with one of the several variants of high-throughput screening to detect mutations in a desired gene. Traditionally this approach has been used for gene function analysis. The increased capacity to sequence DNA and decreasing cost has lead to TILLING being used in several species including both bread (*T. aestivum* L.) and durum wheat (Uauy et al. 2009). Because of the ability of wheat to tolerate a large mutational load due to polyploidy (homeologous genes among the genomes), both tetraploid and hexaploid wheat species have high mutation densities enabling creation of multiple alleles for the target gene. Importantly, the mutant population and the TILLING technology are available as a public resource and a service respectively for the wheat community to share.

Besides the gene function analysis, TILLING can be used to create rare alleles for breeding programs when allelic diversity is absent in the breeding program. For example, TILLING has been used to generate novel alleles for starch branching enzyme II (*SBEIIa*) genes that regulate amylase content in wheat (Botticella et al. 2011). The challenge with using TILLING to incorporate rare genetic variation in the population is the high level of non-targeted background mutations that require several backcrosses to "clean-up" the desired mutant line (keep the desired mutation while removing the undesirable or unwanted mutations). The strategy for introducing TILLING based genetic variation in the breeding program needs to be tested, but there is no theoretical basis to believe it will not become an important source of genetic variation. It is likely to become increasingly important as gene cloning and functional characterization of wheat genes becomes more frequent due to improving genomic resources. An added advantage of TILLED allele is that the polymorphism can be used to develop SNP-based marker that is tightly linked the trait conferring gene and can be included in wheat diversity panel studies. From a practical point, genotypes with alleles generated by TILLING do not have to go through the regulatory process that is required for transgenic material and hence are cost effective.

7.2.3 *Using Transgenic Approaches for Wheat Improvement*

Although polyploidy in wheat increases its suitability for mutagenic approaches for functional analysis of genes and inducing genetic variation, polyploidy poses several challenges for using transgenic approaches for wheat improvement. Transgenic approaches typically involve altering the expression of a gene associated with a particular trait. In a diploid, insertion of T-DNA to knock-out the gene of interest or develop an overexpression line that expresses a specific gene or beneficial allele or splice variant is relatively straightforward. However, silencing expression of particular genes in wheat requires a different approach that involves RNA interference (RNAi) (reviewed by Frizzi and Huang 2010; Travella et al. 2006) or virus induced gene silencing (VIGS) to repress the expression of all three homeologs (Fu et al. 2007; Scofield and Nelson 2009; Bennypaul et al. 2011). VIGS involves transient expression or repression of the target gene and is less cumbersome than transgenic approaches for functional analysis. However, the use of VIGS has been limited in functional analysis. RNAi is of great importance for suppressing the expression of homeologous genes in a polyploid like wheat. Phenotypic consequences of RNAi lines correlate with level of suppression. In most instances, all three homeologous copies are suppressed equally as measured from mRNA quantification. Unlike VIGS, RNAi is stably inherited so is very useful for functional analysis of genes in wheat (Travella et al. 2006; Uauy et al. 2006). To effectively use RNAi technology in wheat for functional genomics analysis, two components are important: (1) the ability to silence the homoeologous genes; (2) stable inheritance of the RNAi-induced phenotype. Ability to alter gene expression in wheat using RNAi for will become increasingly important for introducing commercial traits in wheat in the near future.

7.3 Measuring and Using Genetic Diversity

As mentioned previously, genetic diversity whether it is from intraspecific, interspecific, or intergeneric crosses, mutation, or transformation is critical to plant breeding to make improvements in the crop. Accurate measurements of genetic diversity have always been invaluable for crop breeding because it provides an estimate of genetic variability in cultivars (Smith 1984; Cox et al. 1986), helps breeders to make diverse parental combinations to obtain segregating populations with higher genetic variability for selection (Barrett and Kidwell 1998) and informs the breeder on subpopulations or lineages that might benefit most from the introduction of desirable genes from diverse sources (Thompson and Nelson 1998). Knowledge of relationship among pure lines and inbred lines is useful for planning crosses, identification of heterotic groups and for precise identity of genotypes for plant varietal protection (Hallauer and Miranda 1988).

Methods of measuring diversity within and between populations use morphological traits, pedigree information, and recently molecular marker data. Morphological traits are a simple and easy way to measure genetic diversity using data readily found

within a breeding program, hence did not require sophisticated tools. The disadvantage of this method was that the number of morphological traits is limited and many may be affected by the changes in the environment. Some morphological traits may be influenced by gene interactions such as epistasis and pleiotropy (van Beuningen and Busch 1997). Pedigree information is used for computing pairwise coefficients of parentage (COP), which provides an estimation of genetic diversity for cultivars of self-pollinating species with known pedigrees (Almanza-Pinzon et al. 2003; Murphy et al. 1986; Souza and Sorrells 1989). However COP estimates are limited by the assumption that ancestral lines are unrelated and analysis of genetic similarity between the cultivars may be under- or overestimated due to selection and reselection biases (Almanza-Pinzon et al. 2003; Souza and Sorrells 1989; Fufa et al. 2005).

As biochemical tools became available, breeders developed markers using proteins and DNA. Protein markers were mostly seed storage proteins and isozymes. Asins and Corbonell (1988) used isozymes and Damania et al. (1983) used storage proteins to examine genetic diversity in durum wheat. However protein markers are also influenced by the environmental factors and change with the developmental stages of plants, while isozymes were found to be robust in morphometric analysis of variations. Use of DNA based molecular markers for measuring genetic diversity has received much attention in recent years. These include restriction fragment length polymorphism (RFLP) (Paull et al.1998), randomly amplified polymorphic DNA (RAPD) (Kumar et al. 2011), amplified fragment length polymorphism (AFLP) (Khalighi et al. 2008), simple sequence repeats (SSR) (Mohammadi et al. 2008 and Hao et al. 2011), sequence-related amplified polymorphism (SRAP) and target region amplification polymorphism (TRAP) (Al-Doss et al. 2011), inter-simple sequence repeats (ISSRs) (Carvalho et al. 2009), DNA amplification fingerprinting (DAF) (Mandoulakaniab et al. 2010), sequenced based markers derived from expressed sequence tags (EST) (Karakas et al. 2010) such as single nucleotide polymorphisms (SNPs) (Chao et al. 2009) and diversity arrays technology (DArT) (Zhang et al. 2011). These molecular markers have several advantages: (1) markers are not affected by environmental factors, (2) some markers are co-dominant, hence give a fuller picture of the genetics of lines, and (3) they may provide a better estimate of genetic diversity than phenotypic measurement can provide because they sample the genome and not the expressed aspects of the genome.

To estimate genetic diversity, numerous statistical approached were developed to analyze morphological, pedigree and molecular data. Measurements of genetic distance (GD) or genetic similarity (GS; where GS = 1 − GD), estimates the differences or similarities at genetic level (Weir 1990, p. 162). Genetic distance defined by Nei (1973) is "that difference between two entities that can be described by allelic variation." Euclidean or straight-line statistics has been used to calculate GD for morphological data. For two individuals i and j with observations on phenotypic characters (p) measured as $x_1, x_2 \ldots x_p$ and $y_1, y_2 \ldots y_p$ for i and j, respectively, the Euclidean distance (d) is calculated as:

$$d_{(i,j)} = \left[(x_1 - y_1)^2 + (x_2 - y_2)^2 + \ldots (x_p - y_p)^2\right]^{1/2}$$

Numerous genetic distance measurements have been suggested for calculating genetic diversity with molecular marker data. These include Nie and Li's (1979) coefficient (GD_{NL}), Jaccard's (1908) coefficient (GD_J), Sokal and Michener's (1958) simple matching coefficient (GD_{SM}) and Modified Roger's distance (GD_{MR}). These coefficients can be measured as follows:

$$GD_{NL} = 1 - [2N_{11}(2N_{11} + N_{10} + N_{01})]$$
$$GD_J = 1 - [N_{11}(N_{11} + N_{10} + N_{01})]$$
$$GD_{SM} = 1 - [N_{11} + N_{00}(N_{11} + N_{10} + N_{01} + N_{00})]$$
$$GD_{MR} = [(N_{10} + N_{10})/2N]^{0.5}$$

Whereas N_{11} is number of alleles present in both individuals; N_{10} denotes the number of alleles in individual i; N_{01} is the number of alleles present in individual j; N_{00} is the number of alleles absent in both individuals and N represents the total number of alleles. Each distance measure has its own properties and assumptions for their application to different types of markers for example; some markers are dominant and co-dominant, type of variable under study and the scale of measurements. Different distance measures provide different estimates of mean, minimum and maximum diversity. Mohammadi and Prasanna (2003) have discussed the above mentioned measures of genetic similarity based on molecular marker data and reported that modified Roger's distance is preferred due to its superior genetical and statistical properties.

Multivariate techniques have been widely used to study (visualize) the patterns of genetic diversity within breeding populations. Most commonly used techniques include *k*-means cluster analysis, principal component analysis (PCA), principal coordinate analysis (PCoA) and multidimensional scaling (MDS) (Melchinger 1993; Johns et al. 1997; Thompson and Nelson 1998; Brown-Guedira et al. 2000). Broadly, cluster analyses strive to group individuals to minimize the multivariate distances within groups. Unweighted Paired Group Method using Arithmetic averages (UPGMA) and k-means clustering are widely used as clustering algorithms (Mohammadi and Prasanna 2003; Saatchi et al. 2011). Principal component analysis reduces the data to clarify the relationship between two or more traits and individuals are examined for differences and possible groups are formed. Principal coordinate analysis uses the matrix of similarities and dissimilarities between a set of individuals. Multidimensional scaling represents a set of individuals in a few dimensions by using similarity/dissimilarity distance matrix. Principal component analysis, PCoA and MDS are similar but MDS is preferable when number of individuals is very large and few dimensions are needed to find out relationship between genotypes.

With all of these methods, there continues to be questions regarding their accuracy in estimating genetic diversity among populations. For example what sample size will correctly represent the total genetic diversity, how many markers are needed before little further information is obtained, and which distance measure, clustering method should be used. However, as the markers systems become much more robust and include thousands of markers, the number of markers is no longer a concern.

7.4 Genomics as an Aid for Selection

Once genetic variation has been added to the population, a critical goal is to select for the beneficial variant. For many traits this can be done easily or most effectively by phenotypic selection. For example, the alleles for spring growth habit are dominant, so the selection for winter growth habit from crosses that involve winter and spring growth habit parents can be highly effective in environments where the spring growth habit types are easily killed during the winter. Because the alleles for winter growth habit are recessive and only the homozygous recessive loci are winter growth habit are selected, in one or two generations the surviving lines are homozygous for winter growth habit and no further selection for growth habit is needed. Similarly, the selection for spring growth habit from crosses that involve winter and spring growth habit parents can be highly effective in environments where the winter growth habit plant do not vernalize, hence do not reproduce and are removed from the segregating population. However, removing segregating recessive alleles from a population can require more generations of natural selection as their allele function is hidden by the dominant spring growth habit allele. Progeny rows and breeder selection can remove those families segregating for winter growth habit and can select for homozygous spring growth habit lines. The value of simple phenotypic selection for qualitative traits is that literally millions of plants can be selected for or against with ease. Simple phenotypic selection can similarly be used very effectively for other traits such as herbicide tolerance in which the herbicide resistant plants are selected by spraying the segregating populations with the desired herbicide and killing the susceptible plants (Newhouse et al. 1992; Baenziger et al. 2006a; reviewed by Baenziger and DePauw 2009).

However, for many traits, simple phenotypic selection is difficult (e.g. traits with low heritability), time consuming (e.g. selection for adult plant traits), or costly (e.g. traits that require expensive assays). For some traits two or more of the above concerns are present. It is for these traits that indirect selection is recommended (Falconer 1952; Baenziger et al. 2011). Indirect selection is selecting for a correlated trait that can be more effectively selected for than directly selecting for the trait of interest. For indirect selection to be effective, the ratio of the correlated response (selection for the non-targeted trait) to the direct response (selection for the trait of interest) is greater than one. The ratio is estimated by the square root of the heritability of the correlated trait [h_2] divided by square root of the heritability of the direct trait [h_1] multiplied by the genetic correlation between the genotypes (r_G) or:

$$\frac{Correlated.response}{Direct.response} = \frac{h_2}{h_1} r_G$$

The above equation assumes that the selection intensity is the same and that the ease of measurement is similar between direct and indirect selection. The ease of measurement can affect the selection intensity, but for the purpose of this manuscript, we will assume that the selection intensity is the same and that the ease of measurement is similar between direct and indirect selection. To be successful in this approach,

$h_2 r_G > h_1$ which can occur whenever the square root of the heritability of the trait under direct selection is low relative to the square root of the heritability of the correlated trait and the genetic correlation. An early example of indirect selection was proposed by Shands (1946) who identified a tight linkage between a stem rust [incited by *Puccinia graminis* Pers.:Pers.) gene and loose smut (incited by *Ustilago tritici* (Pers.) Rostr.] resistance gene in barley (*Hordeum vulgare* L.). He proposed that by selecting for the stem rust resistance gene, a barley breeder could also increase the probability of having the loose smut resistance gene. The value of this process was that stem rust could be easily selected for in the seedling stage and the assay for stem rust resistance was highly reliable (hence quick and efficient), while the assay for loose smut resistance required plants be grown to near maturity to determine their susceptibility to the disease (hence time consuming). Currently indirect selection is widely used as the basis for molecular marker based selection in which the marker analysis can be done on seeds or young plants and predicts the gene or quantitative trait locus of interest (e.g. Knox and Clarke 2007, Lehmensiek et al. 2008).

Molecular markers are commonly used in backcrossing for adding one or more traits to an excellent cultivar or elite line (Frisch et al. 1999; Frisch and Melchinger 2001a, b). One of the most important recent examples of using molecular markers to improve wheat has been their use in identifying and aiding the selection for *Fhb1* and other genes and QTLs conveying tolerance to Fusarium head blight (incited primarily by *Fusarium graminearum* Schwabe, but also other by *Fusarium spp*.). Fusarium head blight is difficult, expensive, and labor intensive to screen in the field (Fuentes et al. 2005). Once sources of resistance were identified, a goal was to identify molecular markers that could be used to select for these sources of resistance. One of the earliest resistance genes, *Fhb1,* was identified in the Chinese cultivar Sumai 3. Sumai 3 is a spring wheat that meets the quality needs of the Chinese miller and baker but does not have the end-use characteristics required by American and European millers and bakers, nor is adapted to American and European production environments (Salameh et al. 2011). As such, Sumai 3 became the donor parent for *Fhb1* and other Fusarium head blight alleles in a backcrossing program to move the Fusarium head blight genes into adapted and high end-use quality wheat lines. There are a number of critical needs for an effective backcrossing program. First there must be a good donor parent (e.g. Sumai 3) and a good recurrent parent (the elite released or near release lines in a breeding program). Second there must be a way to identify those backcross lines that contain the desirable alleles. In this case, the phenotypic assay is difficult and time consuming on a plant basis, hence unsuitable for efficient backcrossing. However there are very good molecular markers for introgressing *Fhb1* and other Fusarium head blight QTLs (as both are tagged by molecular marker for the purpose of discussing introgression we will use the term gene to mean gene or QTL) into wheat (Pumphrey et al. 2007; Liu et al. 2008; Buerstmayr et al. 2009). Hence most breeding programs are using molecular markers to introgress Fusarium head blight resistance into wheat. Furthermore, by using a backcrossing procedure to move Fusarium head blight QTLs into wheat, the QTL can be separated from deleterious genes in the donor parent. It is the authors' experience that many crosses using *Fhb1* from Sumai 3 did not lead to commercial cultivars, yet the Fhb1 allele

seems to have no detrimental effect on grain yield in European cultivars (Salameh et al. 2011). Using a different approach in the U.S. Great Plains based upon lines derived from a common population, *Fhb1* does not have any deleterious effects on grain yield (Bakhsh et al. 2013). The most likely explanation of the lack of success deriving lines from previous crosses involving *Fhb1* parents is that the crosses were too wide (too diverse) and while the important gene and considerable variation was present, much of the variation was poor. The rare recombinant that combined the *Fhb1* allele with excellent adaptation genes was simply too hard to find.

The advantages of using molecular markers are that once the DNA is extracted, numerous markers can be run on the material in addition to those tracking *Fhb1*, selecting for Fusarium head blight resistance by phenotyping is difficult and costly as mentioned above, and that even the best gene or QTL is not sufficient to truly protect the plant when there are severe epidemics. Hence, most wheat breeders would like to pyramid genes that reduce the impact of Fusarium head blight. Gene pyramiding for disease or insect resistance is better done using molecular markers than by phenotyping because the disease conditions in the field are hard to precisely reproduce season after season and many of the genes have similar and small effects making them virtually indistinguishable in the field. Hence the breeder does not know if the genes have been pyramided or the assay failed to discriminate between a pyramided or single gene (e.g. Salameh et al. 2011).

One other important aspect of molecular markers is that they can help determine if novel sources of resistance/tolerance may have the same gene or QTL by determining if the line has the same molecular markers as previously identified in resistant lines. For example, *Fhb1* was introgressed from the Chinese wheat Sumai 3. Fusarium head blight resistance is common among Chinese wheat cultivars and it is important to know if the source of resistance is identical to those found in Sumai 3 or not. If the novel source of resistance has the same markers as Sumai 3 at the Fhb1 locus, then it is likely the gene in the novel source is the same as that of Sumai 3. The similarity is based upon the concept that the markers are the same due to a common parent between Sumai 3 and the novel source which contributed the same resistance gene. The markers are the same because they are identical by descent. With a perfect marker or a very closely linked maker, the chance of identical markers indicating the presence of the same gene becomes much higher. Of course with gene duplication and with multiple alleles at the same locus at, marker similarity will be an important first step in suggesting similar sources of resistance, but careful phenotyping will be needed to fully understand allele variation for a trait at a locus, duplicated or closely linked loci (e.g. Yahiaoui et al. 2004).

However, molecular markers have a far greater role in plant breeding than just being used for indirect selection and backcrossing for one or a few traits (Moose and Mumm 2008). Currently molecular markers are used for association mapping studies (e.g. Miedaner et al. 2011) and for marker assisted breeding (e.g. Gupta et al. 2010, discussed in greater detail below). The closeness of the marker to the gene of interest is important because use of molecular markers is based upon linkage disequilibrium (LD, basically does the maker(s) and the gene of interest co-segregate). If the LD is lost, the marker no longer co-segregates with the gene of interest and selecting

for the marker becomes meaningless. Returning to the indirect selection equation, in this case $r_G = 0$. A more problematic situation is when the LD is reversed. In this case for some parents, selecting for the diagnostic marker alleles will be linked to the desired gene(s), however with other parents, the diagnostic marker alleles are linked to the undesired gene. The LD remains but the diagnostic marker alleles switch. By phenotyping the parents and determining their haplotype, the correct marker alleles can be identified for selecting the gene of interest using a parent line by parent line approach.

7.4.1 The Impact of Genomics on Marker Assisted Breeding

A stark trend over the several years has been rapidly falling genotyping costs and static or increasing phenotyping costs. Dramatic advances in sequencing and genotyping technologies are making it feasible to regularly genotype entire breeding populations with huge panels of markers (Close et al. 2009; Zhao et al. 2011), and even genotype individuals by sequencing (Elshire et al. 2011). This trend has already made phenotyping more expensive than dense genotyping in many instances, and surely genotyping will be far cheaper than phenotyping for nearly all complex, difficult-to-measure traits in the coming years. For example, we currently estimate it costs about $35/plot for agronomic characterization and yield measurements, while we can genotype a single line for about $ 20–60 depending upon the marker system. Clearly for replicated studies, genotyping is now less expensive than phenotyping. What logically follows from this trend is that the cost effectiveness of breeding programs will be maximized when the use of genotypic information is maximized and phenotyping is minimized. This prediction is obviously contingent on the successful translation of genotypic data into useful information for generating and selecting lines with superior genetic value. This translation is the challenge for the breeder and geneticist: to maximize the information content of genotypic data for enhance genetic gain through selection and hybridization.

Prediction of genetic value is fundamental to breeding programs, whether it be for cultivar development or population improvement. Previously the only method of genetic value prediction available to a breeder was phenotyping. Molecular markers have created a new method for genetic value prediction. As discussed above, markers have been very useful in wheat breeding programs for introgressing major QTL alleles from exotic sources (e.g., *Fhb1*) or identifying individuals carrying alleles with a qualitative effect, especially on disease resistance (Williams et al. 2007). On the other hand, genetic gain for highly complex traits—such as grain yield and abiotic stress tolerance—have benefited very little, if at all, from the use of marker-assisted selection (MAS) to our knowledge.

Marker-assisted selection in the traditional sense has not worked well for complex traits primarily because genetic variation for these types of traits within elite breeding germplasm is very often composed of many small-effect alleles (Xu and Crouch 2008). Accurate mapping and effect estimation is difficult for alleles of small effect, and hence markers diagnostic of favorable alleles across families are unavailable.

One way around this is to identify markers having a significant effect on the trait and estimate marker effects *de novo* for each family, then apply those markers only to individuals within that family. Linkage disequilibrium is maximized within families and relatively few alleles having an effect on the trait are segregating, meaning that the genetic variation within a family is generated by a small number of effective factors. In this situation, it is possible to use a small set of markers (~ 200) to adequately capture the genetic variation for a complex trait contained within a family, make accurate predictions of genetic value, and increase response to recurrent selection on an annual basis. This approach has been successfully used to perform marker-assisted selection on agronomic traits and yield in maize *(Zea mays* L.) (Eathington et al. 2007; Johnson 2004). A disadvantage to this approach is that a fairly large number of individuals (more than 100; Bernardo and Charcosset 2006) from each family need to be genotyped and phenotyped (Hospital et al. 1997). While this is feasible in maize because of large seed quantities per plant, or more recently the availability of doubled haploids, it is less feasible in wheat. In wheat, smaller seed numbers per plant and the higher number of seeds per plot do not allow multi-environment yield evaluations until advanced generations. In between the F2 generation and yield testing, several rounds of selection are performed based on individual plant visual ratings, disease resistance, flowering time, among other traits. By the time progenies are tested for yield, few progenies remain from each original cross. This process works in wheat because it is easy to remove (cull) types in the early generations that are obviously not needed in later generations while the seed supplies are being increased. For example, in the early generations of the authors' breeding program, selection for winter survival, stem rust resistance, flowering date, plant height, maturity, and an estimate of end-use quality are made before there is sufficient seed for multi-location augmented design testing (Baenziger et al. 2011). The concept is to have as few plots enter into agronomic testing as possible that can be shown to not have potential for eventual release. Occasionally, few lines are advanced that do not have release potential because they have valuable traits for use as parents in future crosses. All of this means that in wheat, the later generations are those that are phenotyped for complex traits and progenies from multiple crosses will need to be used to estimate marker effects. The challenge here is that as the diversity of the population increases, the more quickly LD between loci diminishes. Meeting this challenge is where high-throughput genotyping enters in: today we can affordably screen large populations with enough markers to capture marker-QTL linkages conserved across breeding families.

After marker data and phenotypic data have been gathered on the breeding population, the next issue is marker effect estimation method. A least-squares multiple linear regression model combined with a variable selection technique was previously the most common approach (Hospital et al. 1997; Eathington et al. 1997). As pointed out by Meuwissen et al. (2001), this model faces two problems: (1) An arbitrary statistical threshold is used to select markers. The effect of any marker that surpasses the significance threshold is fully included in the prediction model, while the effects of any marker not reaching the threshold are completely left out of the model; (2)

Inability to model enough markers to adequately capture the small effect alleles distributed across the genome. It would be desirable to include as much of the marker information as possible in order to capture as many small effects across the genome, but ordinary least-squares cannot handle situations where the number of variables (p) is larger than the number of observations (n). This is the so-called "small n, large p" problem. Even when degrees of freedom remain, a large value of p relative to n produces high multicollinearity among the predictor variables, resulting in unstable effect estimates. Another approach to marker-assisted selection, termed genomic selection (Meuwissen et al. 2001) or genome-wide selection (Bernardo and Yu 2007), is gaining acceptance and is better to exploit information contained within large panels of markers scored on entire breeding populations. Several reviews of genomic selection in the plant breeding literature have recently been published (Heffner et al. 2009; Piepho 2009; Jannink et al. 2010; Lorenz et al. 2011), as well as on prediction in human genetics (de los Campos et al. 2010) and animal breeding (Calus 2009). It is not our intent here to provide a comprehensive review of genomic selection, but rather briefly introduce its methodology and rationale, review recent results specific to wheat improvement, and explore areas in need of research to optimize genomic selection for wheat breeding and the use of germplasm.

To reiterate, the goal of genomic selection is to predict the genetic value of breeding progenies from marker data alone with enough accuracy to allow selection based on that prediction alone. Three basic elements constituting a genomic selection will be discussed in turn: genome-wide marker data, a training population (TP), and a prediction model or estimation method.

7.4.2 Genome-Wide Marker Data

Overall accuracy of genomic selection is the product of two sources: the prediction of realized relationships among individuals in the breeding population, and marker-QTL LD (Habier et al. 2007). The first mentioned source of accuracy can be high, meaning that genomic selection can be quite accurate regardless of whether markers are in LD with QTL or not. In a simulation study with a base population of 100, Habier et al. (2007) showed that when 1000 markers were in complete linkage equilibrium with all QTL, genomic selection prediction accuracy was equivalent to that of pedigree-based predictions. Genomic prediction accuracy never surpassed that of pedigree-based predictions, illustrating that markers need to be in LD with QTL for them to be more informative than pedigree information regarding genetic value prediction. The level of LD necessary between QTL and markers for accurate genomic selection depends on the trait genetic architecture and heritability (Calus and Veerkamp 2007). The rate of decay of LD in a population depends on its effective size and hence its genetic diversity (Hedrick 2000), which dramatically varies between species and populations (Gaut and Long 2003; Hamblin et al. 2010). Linkage disequilibrium is maximized in biparental populations at the F2 generation, and decreases slowly with inbreeding because fewer double heterozygotes are present for effective recombination to take

place. This result means the number of markers needed to ensure every QTL is in LD with a marker is actually quite small. In a genomic selection context, prediction accuracy only drops off when the number of markers within biparental populations becomes extremely low. In barley and wheat, this number has been reported to be less than 128 (Lorenzana and Bernardo 2009; Heffner et al. 2011b). Across a diverse panel of wheat cultivars, the LD between markers has been reported to be $r^2 = 0.20$ out to 5 cM. If this level of LD were sufficient for accurate genomic selection, this means approximately 700 markers would be needed to cover a wheat genetic map of 3500 cM (Chao et al. 2010) which is a high estimate for genome size (e.g. Somers et al. 2004). The only two studies investigating wheat genomic selection across a broad population used approximately 1500 Diversity Array Technology (DArT) markers (Triticarte Pty. Ltd., Yarralumla, ACT, Australia) (Crossa et al. 2010; Heffner et al. 2011a). Heffner et al. (2011a) showed that within their population of soft winter wheat varieties and breeding progenies, prediction accuracy did benefit as marker number increased from 192 to 1158, but there was very little benefit between 384 and 1158. In a population of six-row barley lines adapted to the Upper Midwest, Lorenz et al. (2012) found that 1023 polymorphic markers could be reduced to 384 with little to no sacrifice in accuracy. These are population specific situations, and the marker number needed for other breeding populations and settings needs to be determined *de novo* in each particular case. It has also been shown that TP size should increase with marker number in order to take full advantage of additional markers (Muir 2007; Meuwissen 2009).

7.5 Training Population and Selection Candidates

In the context of genomic selection, a TP is a set of individuals that has been both genotyped and phenotyped for estimation of marker effects. Selection candidates are those individuals that have been genotyped but not phenotyped, and for which predictions of genetic value are made. A TP can either be very narrow (e.g., a single breeding family), or very wide (e.g., lines representing the entire genetic diversity of a species), and it should increase (scale) with the effective population size just as the marker number would increase (Meuwissen 2009). This need is easily seen by comparing results from biparental populations versus more diverse populations. High prediction accuracies can be achieved with TPs of only 100 individuals or less when the TP and selection candidates are constrained to recombinant inbred lines from a single cross (Lorenzana and Bernardo 2009; Zhao et al. 2011), whereas accuracy responds strongly to increasing TP size among more diverse populations (Heffner et al. 2011a; Lorenz et al. 2012).

Besides TP size, other key issues are the relationship between the TP and the selection candidates, and combining subpopulations into a single TP. Models developed using a TPs consisting of one distinct population cannot be used to predict the value of selection candidates belonging to another population (Lorenz et al. 2012; Hayes et al. 2009; Asoro et al. 2011), with the penalty in accuracy increasing along

with population divergence (de Roos et al. 2009). This penalty for divergence also applies to the distance between the TP and selection candidates with respect number of breeding cycles. As cycles of selection and intermating proceed, the selection candidates are diverging from the TP in allele frequency and LD pattern, reducing accuracy of prediction (Habier et al. 2007). For this reason, a subset of selection candidates needs to be selected for phenotyping as often as possible in order to update the TP and help re-establish marker-QTL associations and re-estimate marker effects in the context of a new genetic background and changing climate. This need is one of the rationales behind the "Mapping as You Go" approach (Podlich et al. 2004). Mapping as you go is also needed for those breeding situations when you do not have a TP and when new alleles are added to the breeding pool (discussed below).

There is some evidence that combining subpopulations into a single, larger TP can increase prediction accuracy (Hayes et al. 2009; Asoro et al. 2011). This finding is particularly true when a few large effect QTL are present and captured by the appropriate model, markers close to these QTL are in the same linkage phase across subpopulations, and the TP size could be significantly bolstered by combining subpopulations (Hayes et al. 2009). There have also been instances where combining subpopulations failed to increase accuracy despite much higher TP sizes (Lorenz et al. 2012; Zhao et al. 2011). This result could be caused by different segregating QTL between populations, different marker-QTL linkage phases, and epistasis. Clearly nothing can be done to remedy the first possible situation. Higher marker densities will increase the probability that marker and QTL are in the same linkage phase across more diverged populations (de Roos et al. 2009). The problem of epistasis depends upon the objective of selection. If the goal is the select the best line within a family to use as a cultivar, then it would be desirable to capture all those epistatic interactions to identify the line with the highest genotypic value, or even stack alleles across loci that interact positively. If, on the other hand, the objective is to select a parent with a superior breeding value, then the goal is to estimate additive effects of alleles in the context of all other lines it could possibly by crossed with.

7.6 Genomic Selection Models

A fairly large number of models has been developed or adapted from other fields to handle the high-dimensional marker datasets typical of genomic selection applications. These models can be largely grouped into four categories: (1) shrinkage models, (2) variable selection models, (3) dimension reduction models, and (4) nonparametric/machine learning models (for a complete review of the various models, see Lorenz et al. 2011). The models typically classified as shrinkage models, including random (or ridge) regression best linear unbiased prediction (RR-BLUP, assume all marker effects have a common variance, which effectively assumes all QTL have similar effects and are evenly distributed throughout the genome. Clearly this could be a vast over simplification in some cases. For this reason, many other models

have been developed that relax these assumptions and allow marker effects to have different variances and therefore dramatically different effects.

Variable selection models include the least shrinkage and selection operator (LASSO), BayesB, BayesCπ, among many others. The common feature of this category of models is that a proportion of the markers are assigned an effect of zero, with the effects of markers included in the model being sampled either from different distributions (BayesB) or from a common distribution (BayesCπ). This parameterization allows many markers to have zero effects, and other markers having wide ranging effects, which likely models reality better than RR-BLUP (Meuwissen et al. 2001). Dimension reduction models include partial least squares (PLS) and principal component regression (PCR). Besides their simplicity and computational efficiency, any benefit to using these models over other model options has not yet been demonstrated (Moser et al. 2009; Solberg et al. 2009). Finally, a promising, developing research area in estimation methods includes nonparametric and machine learning methods. An advantage of these types of models is their ability to model complex interactions contained within high-dimensional datasets that standard parametric models cannot capture (Gianola et al. 2006). These models have been shown to be useful in both real and simulated data, especially when epistasis is important (Gianola et al. 2006; Crossa et al. 2010).

Studies using cross validation techniques on real data often show very little difference between models in terms of prediction accuracy of overall genetic value despite differences in individual marker effect estimates (Lorenz et al. 2012; Heffner et al. 2011b; VanRaden et al. 2009; Bernardo and Yu 2007). These studies in crops, however, often use cross validation on populations of lines of the same generation and rather limited marker numbers. We expect that studies using training and validation populations separated by one or more generations of selection and recombination will show larger differences between models, with the advantage going to the variable selection models that place more weight on markers of large effect and thus more likely in high LD with QTL (Habier et al. 2007). Larger marker numbers may also differentiate models now compared to the smaller markers numbers currently being used (Meuwissen and Goddard 2010).

7.7 Future Research Needs, New Concepts

We believe the new era of inexpensive genomics relative to phenotyping will revolutionize many aspects of wheat breeding. Some areas of plant breeding that could be dramatically shifted by the genomic selection paradigm and are in need of more research include introgression of new germplasm, exploitation of genotype-by-environment interaction, and germplasm exchange. Clearly many other aspects of plant breeding will be affected, and we choose to give attention to only these three areas for brevity.

Predicting the value of new germplasm within a genomic selection program presents many challenges. If truly new alleles are being introduced, then obviously

their effects on the phenotype within the pertinent environments have never been evaluated. Even if they are not novel alleles, the probability that the markers are in the same linkage phase with the QTL between the elite and exotic populations decreases with population divergence and distance between markers (de Roos et al. 2009). It is well established that accuracy of genomic predictions is greatest when the TP and selection candidates are closely related (Lorenz et al. 2012; de Roos et al. 2009; Habier et al. 2010; Saatchi et al. 2011), meaning that one cannot expect high prediction accuracies of lines derived from exotic sources unless those populations are adequately represented in the TP. One problem is that we do not yet know what precisely constitutes adequate representation, hence the breeder may need to use the current population to estimate allele value. Intuitively, we feel that the constitution of the TP corresponding to any set of selection candidates should represent those selection candidates. For example, if the breeder desires released lines or parents that are 100 % derived from exotic germplasm, then the TP should consist of lines only from that exotic population. If, on the other hand, the breeder wants lines with a small percentage (e.g., 10 %) of exotic alleles, then it would be most appropriate to create a TP composed of recombinant inbred lines derived after a few generations of backcrossing to elite parents. This would generate new LD between markers and QTL –meaning fewer markers required– as well as allow estimation of the exotic alleles within the genetic background of the elite parents, thereby circumventing any problems with allele-by-genetic background interactions (Kramer et al. 2009; Steinhoff et al. 2011). Whether or not this approach is optimal in terms of resource use efficiency is not known. This strategy would require extra phenotyping, meaning extra expense. It would be favorable if an existing TP could simply be augmented with relatively little data on new germplasm, therefore giving value to the already accumulated training data. If and how this could be achieved are areas in need of research.

Genotype-by-environment interaction (GEI) is something that can either be minimized or exploited. Here we are considering a target set of environment where there is a reproducible set of abiotic and biotic conditions, such as a location or management system. Exploiting GEI involves identifying those lines with superior performance in particular environment. A hindrance to fully exploiting GEI has been the need to physically evaluate all lines at all environments. The routine collection of genotypic information allows easy calculation of the realized relationships among all lines in the breeding program. This relationship information, as well as the covariance between different environment effects, can be used in a mixed models framework to predict the performance of lines at environments where they have not been physically planted and evaluated (Crossa et al. 2006; Burgueno et al. 2011). How exactly this is done to fully exploit the allelic information across multiple environments varying in their similarity is yet to be determined. For example, are separate genomic selection models to be developed for each target population of environments, or should a multivariate genomic selection model be developed with performance in different environments being treated as different traits with a covariance structure? Since effects are being estimated at the level of the allele, is it better to maximize the number of lines being evaluated across environments with little to no replication of lines

within or between environments? Or should a complete set of lines be evaluated in each environment with adequate replication? These are questions pertaining more to resource allocation, but successful leverage of GEI data would be required to predict the performance of a line in an environment in which it has not been evaluated.

These issues of new germplasm and GEI lead us to consider the effect of genomic selection on germplasm exchange. Breeding programs, especially within a major seed company, share germplasm on a global basis for mutual benefit. It seems to us that germplasm could be more rationally exchanged under a genomic selection framework compared to phenotypic selection. Consider a situation where two neighboring wheat breeding programs, or programs separated by continents but in similar ecogeographic regions, are routinely genotyping all individuals entering the breeding programs and conducting genomic selection. Breeder A could obtain all the genotypic data on the lines of breeder B and calculate genetic distance statistics as described herein. If a set of lines from breeder B is sufficiently closely related to breeder A's TP, the genetic value of those lines within the environments being targeted by breeder A could be predicted simply by a few key strokes. Breeder A would then obtain those lines with high predicted genetic value and validate their performance in the field. This could make germplasm sharing more effective by saving precious field resources for only those lines with a high probability of success in the new environments. The approach would be readily available to public programs who want to exchange germplasm and maximize their impact.

7.8 It is More than Genomics and Breeding: A Glimpse at the Future

> *Dans les champs de l'observation le hasard ne favorise que les esprits préparés.* (Translated as "In the fields of observation chance favors only the prepared mind.")
> Louis Pasteur, 1854

In concluding, it must be recognized that plant breeding and crop improvement will always be a synthesis of established and evolving science. Furthermore, its success or failure will depend upon the highly skilled individuals who ask the right questions, find the right germplasm, select the best breeding method for their objective, and see the process through to new released cultivars. Modern plant breeding will bridge conventional plant breeding and genomics and it will continue to bridge to other sciences. With the massive information technology and tools that will be developed to fully take advantage of genomics and its applications to plant breeding, it will be a small step to fully recognize the need for computer simulations to help understand and further advance selection science (Sun et al. 2011), as well as attempting to model grain yields in environments that were not or could not be tested previously (Baenziger et al. 2004).

Finally, as much as we learn about the genetics of plant breeding, there will always be areas that require new understanding. As described by Rasmussen and Phillips (1997), plant breeders have been highly successful with improving cultivars

using extremely narrow gene pools. Most breeding programs have highly successful lineages where a few lines and their progeny consistently are the basis for new cultivars. Again these lineages are often quite narrow, yet genetic improvement continues. Rasmussen and Phillips (1997) attributed the continued gains to *de novo* variation and elevated epistasis. While genomic models will be able to improve their estimates of epistasis, the ability to capture or estimate *de novo* variation is impossible and will continue to be one of the reasons that plant breeders will continue to make crosses and select and evaluate in the field.

References

Al-Doss AA, Elshafei AA, Moustafa KA et al (2011) Comparative analysis of diversity based on morph-agronomic traits and molecular markers in durum wheat under heat stress. Afr J Biotech 10:3671–3681

Almanza-Pinzon MI, Warburton ML, Fox PN, Khairallah M (2003) Comparison of molecular markers and coefficients of parentage for the analysis of genetic diversity among spring bread wheat accessions. Euphytica 130:77–86

Asins MJ, Carbonell EA (1988) Distribution of variability in durum wheat world collection. Theor Appl Genet 77:287–294

Asoro FG, Newell MA, Beavis WD et al (2011) Accuracy and training population design for genomic selection on quantitative traits in elite North American oats. Plant Genome 4:132–144

Baenziger PS, DePauw RM (2009) Wheat breeding: procedures and strategies. In: Carver BF (ed) Wheat: science and trade. Wiley-Blackwell Publishing, Ames, pp 275–308

Baenziger PS, McMaster GS, Wilhelm WW et al (2004) Putting genes into genetic coefficients. Field Crop Res 90:133–143

Baenziger PS, Beecher B, Graybosch RA et al (2006) Registration of 'Infinity CL' wheat. Crop Sci 46:975–977

Baenziger PS, Russell WK, Graef GL, Campbell BT (2006b) Improving lives: 50 years of crop breeding, genetics and cytology (C-1). Crop Sci 46:2230–2244

Baenziger PS, Salah I, Little RS et al (2011) Structuring an efficient organic wheat breeding program. Sustainability 3(8):1190–1206

Bakhsh A, Mengistu N, Baenziger PS, Dweikat I, Wegulo SN, Rose D, Bai G, Eskridge KM (2013) Effect of Fusarium head blight (FHB) resistance gene *Fhbl* on agrnomic and end-use quality traits of hard red witner wheat. Crop Sci .. 53:793-801.

Baker RJ (1984) Quantitative genetic principles in plant breeding. In: Gustafson JP (ed) Gene manipulation in plant improvement I. Plenum Press, New York, pp 147–176

Barrett BA, Kidwell KK (1998) AFLP-based genetic diversity assessment among wheat cultivars from the Pacific Northwest. Crop Sci 38:1261–1271

Bennypaul HS, Mutti JS, Rustgi S, Kumar N, Okubara, PA, and Gill KS (2011) Virus-induced gene silencing (VIaS) of genes expressed in root, leaf, and meiotic tissues of wheat. Funct. Integr Genomics 12:143-158.

Bernardo R, Charcosset A (2006) Usefulness of gene information in marker-assisted recurrent selection: a simulation appraisal. Crop Sci 46:614–621

Bernardo R, Yu J (2007) Prospects for genome-wide selection for quantitative traits in maize. Crop Sci 47:1082–1090

Bhat PR, Lukaszewski A, Cui XP et al (2007) Mapping translocation breakpoints using a wheat microarray. Nucleic Acids Res 35:2936–2943

Bos I (1983) The optimal number of replications when testing lines or families on a fixed number of plots. Euphytica 32:311–318

Botticella E, Sestili F, Hernandez-Lopez A et al (2011) High resolution melting analysis for the detection of EMS induced mutations in wheat SbeIIa genes. BMC Plant Biol 11:156

Braun HJ, Payne TS, Morgounov AI et al (1998) The challenge: one billion tons of wheat by 2020. In: Proc of the Ninth Int Wheat Genet Symp, Saskatoon, Saskatchewan, Canada, 2–7 August 1998, Canada, pp 33–40

Brown-Guedira GL, Thompson JA, Nelson RL, Warburton ML (2000) Evaluation of genetic diversity of soybean introductions and North American ancestors using RAPD and SSR markers. Crop Sci 40:815–823

Buerstmayr H, Ban T, Anderson JA (2009) QTL mapping and marker assisted selection for Fusarium head blight resistance in wheat: a review. Plant Breed 128:1–26

Burgueno J, Crossa J, Cotes JM et al (2011) Prediction assessment of linear mixed models for multi-environment trials. Crop Sci 51:944–954

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

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 124:362–368

Carvalho A, Lima-Brito J, Macas B, Guedes-Pinto H (2011) Genetic diversity in old Portuguese durum wheat cultivars assessed by retrotransposon-based markers. Biochem Genet 47:276–294

Chao S, Zhang W, Akhunov E et al (2009) Analysis of gene-derived SNP marker polymorphism in US wheat (*Triticum Aestivum* L.) cultivars. Mol Breed 23:23–33

Chao S, Dubcovsky J, Dvorak J et al (2010) Population and genome-specific patterns of linkage disequilibrium and SNP variation in spring and winter wheat (*Triticum aestivum L.*) RID A-4969-2008 RID C-5600-2011. BMC Genom 11:727

Close TJ, Bhat PR, Lonardi S et al (2009) Development and implementation of high-throughput SNP genotyping in barley. BMC Genom 10:582

Cox TS, Murphy JP, Rodgers DM (1986) Changes in genetic diversity in the red and winter wheat regions on the United States. Proc Natl Acad Sci 83:5583–5586

Crossa J, Burgueno J, Cornelius PL et al (2006) Modeling genotype × environment interaction using additive genetic covariances of relatives for predicting breeding values of wheat genotypes. Crop Sci 46:1722–1733

Crossa J, de los CG, Perez P et al (2010) Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. Genetics 186:713–724

Damania AB, Porceddu E, Jakson MT (1983) A rapid method for the evaluation of variation in germplasm collections of cereals using polyacrylamide gel electrophoresis. Euphytica 32: 877–883

de LCamposG, Gianola D, Allison DB (2010) Predicting genetic predisposition in humans: the promise of whole-genome markers. Nat Rev Genet 11:880–886

DePauw RM, Knox RE, Clarke FR et al (2007) Shifting undesirable correlations. Euphytica 157:409–415

de Roos APW, Hayes BJ, Goddard ME (2009) Reliability of genomic predictions across multiple populations. Genetics 183:1545–1553

Eathington SR, Crosbie TM, Edwards MD et al (2007) Molecular markers in a commercial breeding program. Crop Sci 47:S154–S163

Eathington S, Dudley J, Rufener G (1997) Usefulness of marker-QTL associations in early generation selection. Crop Sci 37:1686–1693

Ehdaie B, Whitkus RW, Waines JG (2003) Root biomass, water-use efficiency, and performance of wheat-rye translocations of chromosomes 1 and 2 in spring bread wheat '*Pavon*'. Crop Sci 43:710–717

Elshire RJ, Glaubitz JC, Sun Q et al (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE 6:e19379

Falconer DS (1952) The problem of environment and selection. Am Nat 86:293–298

Falconer DS, Mackay TFC (1996) Interlocution to Quantitative Genetics, 4th edn. Longman and Company, Essex

Fehr WR (1987) Principles of cultivar development: theory and technique. Macmillan, New York

Feschotte C, Jiang N, Wessler SR (2002) Plant transposable elements: where genetics meets genomics. Nature Rev 3:329–341

Frisch M, Bohn M, Melchinger AE (1999) Minimum sample size and optimal positioning of flanking markers in marker-assisted backcrossing for transfer of a target gene. Crop Sci 39:967–975

Frisch M, Melchinger AE (2001a) Marker-assisted backcrossing for introgression of a recessive gene. Crop Sci 41:1485–1494
Frisch M, Melchinger AE (2001b) Marker-assisted backcrossing for simultaneous introgression of two genes. Crop Sci 41:1716–1725
Frizzi A, Huang S (2010) Tapping RNA silencing pathways for plant biotechnology. Plant Biotech J 8:655–677
Fu D, Uauy, C, Blechel A, Dubocovsky J (2007) RNA interference for wheat functional gene analysis. Transgenic Res. 16: 689-701.
Fuentes RG, Mickelson HR, Busch RH et al (2005) Resource allocation and cultivar stability in breeding for Fusarium head blight resistance in spring wheat. Crop Sci 45:1965–1972
Fufa H, Baenziger PS, Beecher BS et al (2005) Comparison of phenotypic and molecular marker-based classifications of hard red winter wheat cultivars. Euphytica 145:133–146
Gaut B, Long A (2003) The lowdown on linkage disequilibrium. Plant Cell 15:1502–1506
Gianola D, Fernando RL, Stella A (2006) Genomic-assisted prediction of genetic value with semiparametric procedures. Genetics 173:1761–1776
Gupta PK, Langridge P, Mir RR (2010) Marker-assisted wheat breeding: present status and future possibilities. Mol Breed 26:145–161
Guzman C, Caballero L, Alvarez JB (2011) Molecular characterization of the Wx-B1 allelic variants identified in cultivated emmer wheat and comparison with those of durum wheat. Mol Breed 28:402–411
Habier D, Fernando RL, Dekkers JCM (2007) The impact of genetic relationship information on genome-assisted breeding values. Genetics 177:2389–2397
Habier D, Tetens J, Seefried F et al (2010) The impact of genetic relationship information on genomic breeding values in german holstein cattle. Genet Sel Evol 42:5
Hallauer AR, Miranda JB (1988) Quantitative genetics in maize breeding. 2nd edn. Iowa State University Press, Ames
Hamblin MT, Close TJ, Bhat PR et al (2010) Population structure and linkage disequilibrium in US barley germplasm: implications for association mapping. Crop Sci 50:556–566
Hao C, Wang L, Ge H et al (2011) Genetic diversity and linkage disequilibrium in chinese bread wheat (*Triticum Aestivum* L.) Revealed by SSR Markers. PLoS ONE 6(2):e17279. doi:10.1371/journal.pone.0017279
Hayes BJ, Bowman PJ, Chamberlain AC et al (2009) Accuracy of genomic breeding values in multi-breed dairy cattle populations. Genet, Sel, Evol: GSE 41:51
Hedrick PW (2000) Genetics of populations (Second Edition). Jones and Bartlett Publishers, Sudbury, MA, USA, p. 553
Heffner EL, Sorrells ME, Jannink J (2009) Genomic selection for crop improvement. Crop Sci 49:1–12
Heffner EL, Jannink JL, Sorrells ME (2011a) Genomic selection accuracy using multifamily prediction models in a wheat breeding program. Plant Genome 4:65–75
Heffner EL, Jannink JH, Iwata E et al (2011b) Genomic selection accuracy for grain quality traits in biparental wheat populations. Crop Sci 51:2597–2606
Hospital F, Moreau L, Lacoudre F et al (1997) More on the efficiency of marker-assisted selection. Theor Appl Genet 95:1181–1189
Huang S, Sirikhachornkit A, Su X et al (2002) Genes encoding plastid *acetyl-CoA carboxylase* and *3-phosphglycerate kinase* of the *Triticum/Aegilops* complex and the evolutionary history of wheat. Proc Natl Acad Sci 99:8133–8138
Huehn M (1996) Optimum number of crosses and progeny per cross in breeding self-fertilizing crops. I. General approach and first numerical results. Euphytica 91:365–374
Jaccard P (1908) Nouvelles researches sur la distribution florale. (Translated as "New researches on floral distribution.") Bull Soc Vaud Sci Natl 44:223–270
Jannink J, Lorenz AJ, Iwata H (2010) Genomic selection in plant breeding: from theory to practice. Briefings in functional genomics and. Proteomics 9:166–177
Johns MA, Skrotch PW, Neinhuis J et al (1997) Gene pool classification of common bean landraces from Chile based on RAPD and morphological data. Crop Sci 37:605–613
Johnson GR (2004) Marker-assisted selection. Plant Breed Rev 24:293–309

Karakas O, Gurel F, Uncuoglu AA (2010) Exploiting a wheat EST database to assess genetic diversity. Genet Mol Biol 33:719–730

Khalighi M, Arzani A, Poursiahbidi MA (2008) Assessment of genetic diversity in *Triticum* spp. and *Aegilops* spp. using *AFLP* markers. Afr J Biotech 7:546–552

Knox RE, Clarke FR (2007) Molecular breeding approaches for enhanced resistance against fungal pathogens. In: Punja ZK, De Boer S, Sanfacon H (ed) Biotechnology and plant disease management. CAB Int, Oxfordshire. pp 321–357

Kramer CC, Polewicz H, Osborne TC (2009) Evaluation of QTL alleles from exotic sources for hybrid seed yield in the original and different genetic backgrounds of spring-type *Brassica Napus* L. Mol Breed 24:419–431

Kumar R, Jaiswal SK, Vishwakarma MKetal (2011) Assessment of genetic diversity and its usefulness for varietal identification in Indian Elite varieties of wheat (*Triticum Aestivum* L.) using RAPD markers. Asian J Biotech 3:460–469

Lehmensiek A, Sutherland MW, McNamara RB (2008) The use of high resolution melting (HRM) to map single nucelopide polymorphism markets lined to a covered smut resistance gene in barley. Theor Appl Genet 117:721–728

Liu S, Pumphrey MO, Gill BS et al (2008) Toward positonal cloning of *Fhb1*, a major QTL for Fusarium head blight in wheat. Cereal Res Comm 36 (Suppl B):195–201

Lorenz AJ, Chao S, Asoro FG et al (2011) Genomic selection in plant breeding: knowledge and prospects. Adv Agron 110:77–123

Lorenz AJ, Smith KP, Jannink J-L (2012) Potential and optimization of genomic selection for Fusarium head blight resistance in six-row barley. Crop Sci 52:1609–1621

Lorenzana RE, Bernardo R (2009) Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. Theor Appl Genet 120:151–161

Mandoulakaniab BA, Shahnejat-Bushehric A-A, Tabatabaeid BES et al (2010) Genetic diversity among wheat cultivars using molecular markers. J Crop Imp 24:299–309

McCallum CM, Comai L, Greene EA, Henikoff S (2000) Targeting induced local lesions in genomes (TILLING). Plant Physiol 123:439–442

Melchinger AE (1993) Use of RFLP markers for analyses of genetic relationships among breeding materials and prediction of hybrid performance. In: Buxton DR (ed) Proc of the Int Crop Sci Congress, 1st, Ames, IA. July 1992. CSSA, Madison, WI. pp 621–628

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

Meuwissen THE (2009) Accuracy of breeding values of 'unrelated' individuals predicted by dense SNP genotyping. Genet Sel Evol 41

Meuwissen T, Goddard M (2010) Accurate prediction of genetic values for complex traits by whole-genome re-sequencing. Genetics 185:623–338

Miedaner T, Wurschum T, Maurer HP et al (2011) Association mapping for Fusarium head blight resistance in European soft winter wheat. Mol Breed 28:647–655

Mohammadi SA, Prasanna BM (2003) Analysis of genetic diversity in crop plants—salient statistical tools and considerations. Crop Sci 43:1235–1248

Mohammadi SA, Khodarahmi M, Jamalirad S, Jalal Kamali MR (2008) Genetic diversity in a collection of old and new bread wheat cultivars from Iran as revealed by simple sequence repeat-based analysis. Annals Appl Biol 154:67–76

Moose SP, Mumm RH (2008) Molecular plant breeding as the foundation for the 21st century crop improvement. Plant Physiol 147:969–977

Moser G, Tier B, Crump RE et al (2009) A comparison of five methods to predict genomic breeding values of dairy bulls from genome-wide SNP markers. Genet Sel E 41:56

Muir WM (2007) Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters. J Anim Breed Genet 124:342–355

Mujeeb-Kazi A, Rosas VB, Roldan S (1996) Conservation of the genetic variation of *Triticum tauschii* (Coss.) Schmalh. (*Aeqilops squarrosa* auct Non L.) in synthetic hexaploid wheats (*T turgidum* L s lat X *T taschhii*; (2n = 6x = 42, AABBDD) and its potential utilization for wheat improvement. Genet Res Crop E 43:129–134

Murphy JP, Cox TS, Rodgers DM (1986) Cluster analysis of red winter wheat cultivars based upon coefficients of parentage. Crop Sci 26:672–676
Nei M (1973) Analysis of gene diversity in subdivided populations. Multi Proc Natl Acad Sci (USA) 70:3321–3323
Nei M, Li W (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci (USA) 76:5269–5273
Newhouse KE, Smith WA, Starrett MA et al (1992) Tolerance to imidazolinone herbicides in wheat. Plant Physiol 100:882–886
Paull JG, Chalmers KJ, Karakousis A et al (1998) Genetic diversity in Australian wheat varieties and breeding material based on RFLP data. Theor Appl Genet 96:435–446
Peng JH, Sun D, Nevo E (2011) Domestication, evolution, and genetics and genomics in wheat. Mol Breed 28:281–301
Piepho HP (2009) Ridge regression and extensions for genome-wide selection in maize. Crop Sci 49:1165–1176
Podlich D, Winkler C, Cooper M (2004) Mapping as you go: An effective approach for marker-assisted selection of complex traits. Crop Sci 44:1560–1571
Pumphrey MO, Bernardo R, Anderson JA (2007) Validating the *Fhb1 QTL* for Fusarium head blight resistance in near-isogenic wheat liens developed from breeding populations. Crop Sci 47:200–206
Rasmussen DC, Phillips RL (1997) Plant Breeding progress and genetic diversity from de novo variation and elevated epistasis. Crop Sci 37:303–310
Robertson DS (1985) A possible technique for isolating genic DNA for quantitative traits in plants. J Theor Biol 117:1–10
Saatchi M, McClure MC, McKay SD et al (2011) Accuracies of genomic breeding values in American Angus beef cattle using K-means clustering for cross-validation. Genet Sel E 43:40
Salameh A, Buerstmayr M, Steiner B et al (2011) Effects of introgression of two *QTL* for Fusarium head blight resistance from Asian spring wheat by marker-assisted backcrossing into European winter wheat on Fusarium head blight resistance, yield, and quality traits. Mol Breed 28:485–494
Scofield S, Nelson R (2009) Resources for virus-induced gene silencing in the grasses. Plant Physiology 149:152-157.
Sears ER (1976) Genetic control of chromosome pairing in wheat. Ann Rev Genet 10:31–51
Sears ER (1993) Use of radiation to transfer alien chromosome segments to wheat. Crop Sci 33:897–901
Shands RG (1946) An apparent linkage of resistance to loose smut and stem rust in barley. Agron. J. 38: 690-692.
Smith JSC (1984) Genetic variability within U.S. hybrid maize: multivariate analysis of isozyme data. Crop Sci 24:1041–1046
Sneep J (1977) Selection for yield in early generations of self-fertilizing crops. Euphytica 26:27–30
Sokal RR, Michener CD (1958) A statistical method for evaluating systematic relationships. Univ Kansas Sci Bull 38:1409–1438
Solberg TR, Sonesson AK, Woolliams JA, Meuwissen THE (2009) Reducing dimensionality for prediction of genome-wide breeding values. Genet Sel Evol 41
Somers DJ, Isaac P, Edwards K (2004) A high-density microsatellite consensus map for bread wheat (*Triticum Aestivum* L.). Theor Appl Genet 109:1105–1114
Souza E, Sorrells ME (1989) Pedigree analysis of North American oat cultivars released from 1951–1985. Crop Sci 29:595–601
Steinhoff J, Liu W, Maurer HP et al (2011) Multiple-line cross quantitative trait locus mapping in European elite maize. Crop Sci 51:2505–2516
Sun X, Peng T, Mumm RH (2011) The role and basics of computer simulation in support of critical decisions in plant breeding. Mol Breed 28:421–436
Thompson JA, Nelson RL (1998) Utilization of diverse germplasm for soybean yield improvement. Crop Sci 38:1362–1368
Travella S, Klimm TE, Keller B (2006) *RNA* interference-based gene silencing as an efficient tool for functional genomics in hexaploid bread wheat. Plant Physiol 142:6–20
Uauy C, Distelfeld A, Fahima T et al (2006) A *NAC* gene regulating senescence improves grain protein, zinc, and iron content in wheat. Science 314:1298–1301

Uauy C, Paraiso F, Colasuonno P et al (2009) A modified *TILLING* approach to detect induced mutations in tetraploid and hexaploid wheats. BMC Plant Biol 9:115

Van Beuningen LT, Busch RH (1997) Genetic diversity among North American spring wheat cultivars: III. Cluster analysis based on quantitative morphological traits. Crop Sci 37:981–988

VanRaden PM, Van Tassell CP, Wiggans GR et al (2009) Invited review: reliability of genomic predictions for north American Holstein bulls. J Dairy Sci 92:16–24

Weir BS (1990) Genetic data analysis. Sinauer Associates, Inc, Suderland, Massachusetts

William HM, Trethowan R, Crosby-Galvan EM (2007) Wheat breeding assisted by markers: *CIMMYT*'s experience. Euphytica 157:307–319

Xu YB, Crouch JH (2008) Marker-assisted selection in plant breeding: from publications to practice. Crop Sci 48:391–407

Yahiaoui N, Srichumpa P, Dudler R, Keller B (2004) Genome analysis at different ploidy levels allows cloning of the powdery mildew resistance gene *Pm3b* from hexaploid wheat. Plant J 37:528–538

Zeller FJ, Hsam SKL (1984) Broadening the genetic variability of cultivated wheat by utilizing rye chromatin. In: Sakamoto S (ed) Proc 6th Int Wheat Genet Symp, Kyoto. pp 161–173

Zhang L, Liu D, Guo X et al (2011) Investigation of genetic diversity and population structure of common wheat cultivars in northern China using *DArT* markers. BMC Genet 12:42

Zhao Y, Gowda M, Liu W et al (2011) Accuracy of genomic selection in European maize elite breeding populations. Theor Appl Genet. doi:1007/s00122-011-1745-y

Chapter 8
Genetic Dissection of Aluminium Tolerance in the Triticeae

Harsh Raman and Perry Gustafson

Abstract Aluminium (Al) toxicity is the major constraint to crop productivity on acidic soils worldwide. Members of the Triticeae such as wheat, barley, and rice, show a range of genetic variation within and between species. Among key cereals, rye displays the maximum level of Al tolerance, while barley shows the least. In the majority of species, genetic control for aluminium tolerance has been investigated using conventional genetic and molecular analyses. During the last decade, candidate and causative genes and mechanisms for Al tolerance have been identified in wheat, barley, rice and sorghum. New phenotypic and genotyping platforms were also developed in order to understand genes and their networks underlying Al tolerance comprehensively. In this chapter, we review the progress made on recent discoveries on genetic dissection of aluminium tolerance with special focus on wheat, barley, and rice.

8.1 Introduction

Soil acidity limits crop production on 30–40 % of the world arable soils that constitute approximately 60 % of tropical and sub-tropical regions (von Uexkull and Mutert 1995). The poor fertility of acid soils is generally ascribed due to a combination of toxicities of aluminium (Al^{3+}), manganese (Mn^{2+}), iron (Fe^{2+}) and deficiencies of phosphorus (P) due to its decreased availability, calcium, magnesium and potassium, in addition to low water holding capacity and susceptibility to compaction (Camargo et al. 1989, 1992; Moroni et al. 1991; von Uexkull and Mutert 1995). Under highly acidic soil conditions (pH < 5), dissolution of Al-containing compounds is enhanced and the release of toxic trivalent Al^{3+} ions into soil solution can rapidly inhibit root growth even at low concentrations (Delhaize et al. 1993a). Subsequently Al^{3+}

H. Raman (✉)
Graham Centre for Agricultural Innovation (an alliance between Charles Sturt University and NSW Department of Department of Primary Industries), Wagga Wagga Agricultural Institute, Wagga Wagga, NSW 2650, Australia
e-mail: harsh.raman@dpi.nsw.gov.au

P. Gustafson
Division of Plant Sciences, University of Missouri, Columbia, Missouri 65211, USA

R. Tuberosa et al. (eds.), *Genomics of Plant Genetic Resources,*
DOI 10.1007/978-94-007-7575-6_8,

toxicity impairs the ability of plant to acquire nutrients and water (Pan et al. 1989), thus increasing the plant's sensitivity to various stresses especially to drought and heat (Krizek and Foy 1988). Therefore, Al^{3+} toxicity has been considered as a major limiting constraint to crop productivity, although on some highly acidic soils, rhizotoxicity of H^+ can also occur (Iuchi et al. 2007). In addition to repeated surface applications of lime, improvement of crop tolerance to Al^{3+} ions has been one of the key targets of cereal breeding programs aiming to develop varieties suitable for cultivation on acidic soils. As a result, hundreds of varieties tolerant to Al^{3+} toxicity have been developed (Raman et al. 2008). Carver et al. (1993) showed the benefits of growing bread wheat (*Triticum aestivum* L) on acid soil by demonstrating that Al^{3+} tolerant lines produce 31 % more spikes, 66 % more biomass, and 68 % higher grain yield than less tolerant cultivars if grown on acid soil.

Members of the Triticeae (Poaceae) include several key food and feed crops such as bread wheat, durum wheat (*T. turgidum* ssp *durum* L.), barley (*Hordeum vulgare* L.), rice (*Oryza sativa* L.), rye (*Secale cereale* L.), oat (*Avena sativa* L), maize (*Zea mays* L.), and sorghum (*Sorghum bicolor* L.), and contribute more than 60 % of total daily caloric intake of world's population. These cereal crops display a range of genetic variation for their ability to tolerate the toxic levels of Al^{3+}. The Al^{3+} tolerance order has been reported as maize > rye > triticale (X *Triticosecale* Wittmack) > wheat > barley (Polle and Konzak 1985), rye > oat > millet (*Panicum miliaceum* L.) > bread wheat > barley > durum wheat (Bona et al. 1993), and rice > maize > pea > barley (Ishikawa et al. 2000), rice > maize > sorghum ≥ wheat (Famoso et al. 2010). Natural variation also exists among genotypes within species (cultivars, breeding lines and landrace accessions); for example, within cultivated rice, japonica cultivars are the most tolerant to Al^{3+} as compared to indica cultivars. In wheat, the highest level of Al^{3+} tolerance has been reported in genotypes originating from Brazil.

Members of the Triticeae have developed mechanisms to tolerate Al^{3+} toxicity either during events of evolution/domestication or adaptation to acid soils (Kochian 1995; Pellet et al. 1996; Larsen et al. 1998; Garvin and Carver 2003; Raman and Gustafson 2010; Ryan et al. 2010). Two main mechanisms of Al^{3+} tolerance were proposed by Ma et al. (2001): (I) external tolerance mechanisms, by which Al is excluded from plant tissues, especially the symplastic portion of the root meristems to the rhizosphere, which modifies the pH and chelate the toxic Al^{3+} ions; and, (II) internal tolerance mechanisms, allowing plants to tolerate Al^{3+} in the plant symplasm where Al that has permeated the plasmalemma is sequestered or converted into an innocuous form. Different cereal species secrete different types of organic acids from the roots in response to Al, for example, wheat, rye and maize secrete both malate and citrate ions, whereas barley secretes citrate alone (Delhaize et al. 1993b; Ryan et al. 1995a, 2009; Raman and Gustafson 2010; Ligaba et al. 2012). Galvez et al. (1991) reported that in response to Al^{3+}, an Al^{3+}-tolerant sorghum cultivar increased root organic acid content more than an Al^{3+}-sensitive cultivar. These organic anions have high affinity for Al and protect the sensitive root apices by chelating the Al^{3+} and then detoxify in the rhizosphere (Delhaize et al. 1993b; Jones 1998; Kinraide et al. 2005). Organic acids vary in their ability to complex with Al^{3+}, for example stability of the Al^{3+}-citrate complex is substantially stronger than that of the Al^{3+}-malate complex (Hue et al. 1986).

Table 8.1 Candidate/causative genes controlling Al^{3+} tolerance in cereals

Crop species	Gene	Reference
Barley (*Hordeum vulgare* L)	*HvAACT1*	Fujii et al. (2012)
Maize (*Zea mays* spp. *mays L.*)	*ZmMATE1 & ZmMATE2*	Maron et al. (2010)
	ZmALMT2	Krill et al. (2010)
		Ligaba et al. (2012)
	Malic Enzyme	Krill et al. (2010)
	S-adenosyl-L-homocysteinase	Krill et al. (2010)
Rice (*Oryza sativa* L)	*ART1*	Yamajia et al. (2009)
	OsFRDL1	Yokosho et al. (2011)
	STAR1 & STAR2	Huang et al. (2009)
	Nrat1	Xia et al. (2010)
Rye (*Secale cereale* L)	*ScALMT*	Collins et al. (2008)
	ScMATE	Yokosho et al. (2010)
Sorghum (*Sorghum bicolor* L)	*SbMATE*	Magalhaes et al. (2007)
Wheat (*Triticum aestivum* L)	*TaALMT (ALMT1)*	Sasaki et al. (2004, 2006)
	TaMATE1	Fujii et al. (2009)

Genetic analyses of several Triticeae populations showed both discrete and continuous variation for Al^{3+} tolerance. Qualitative and quantitative trait loci (QTL) associated with natural genetic variation within genotypes/species have been dissected with the conventional genetic and molecular marker analyses see reviews (Wang et al. 2006a; Raman and Gustafson 2010). Allelic variation at a single locus conditioning Al^{3+} tolerance has been observed in barley, wheat and maize (Rhue et al. 1978; Minella and Sorrells 1992; Raman et al. 2005b; Raman et al. 2008). More recently, candidate/causative genes encoding transporters for Al^{3+} -stimulated secretion of malate and citrate have been identified in wheat, barley, sorghum, and rye (Table 8.1). In rice, Al^{3+} tolerance is mediated by a novel mechanism, independent of root tip Al exclusion (Famoso et al. 2010). Genes encoding a nucleotide binding domain (designated as *STAR1*) and gene encoding for a transmembrane-spanning protein, similar to a bacterial type ATP binding cassette (ABC) transporter (designated as *STAR2*) have been shown to be regulated by Al^{3+} (Huang et al. 2009). Additionally, *ART1* and *Nrat1* genes for *Aluminium rhizotoxicity 1* and *Nramp aluminium transporter 1,* respectively, have also been cloned and characterised in rice (Yamajia et al. 2009; Xia et al. 2010). During last five years, significant progress has been made in phenotyping, genotyping and statistical methods to identify phenotypic trait-marker association (Krill et al. 2010; Famoso et al. 2011). This chapter focuses on genes (loci) and their functions, and new phenotypic methods to uncover Al^{3+} tolerance in the key cereals such as wheat, barley and rice.

8.2 Improved Methods for Germplasm Evaluation for Al^{3+} Tolerance

Several methods such as pot assays in the glasshouse, field evaluation on acidic soils and laboratory based methods: nutrient solution culture (Raman et al. 2002, 2005b, 2008) and staining roots with hematoxylin (Polle et al. 1978; Cançado et al. 1999),

eriochrome cyanine (Magalhaes et al. 2004; Gruber et al. 2006; Wang et al. 2006b; Furukawa et al. 2007), and nitroblue tetrazolium have been employed for evaluating germplasm for aluminium tolerance (Massot et al. 1992, 1999; Bennet 1997; Horst et al. 1997; Maltais and Houde 2002; Wang et al. 2006a). Laboratory and greenhouse based assays are preferred as they are efficient, reliable, generally non-destructive, and cheaper to perform even at the earlier stages of plant breeding. These methods have enabled scientists to characterise germplasm for Al^{3+} tolerance, investigate the inheritance and location of loci associated with Al^{3+} tolerance, assessed on the basis of root growth, root regrowth (RRG)/relative root length and differential patterns of root staining. Among these measures, relative growth of the longest root has been the most commonly used criteria for estimating Al^{3+} tolerance. However, Famoso et al. (2010) showed that the RRG is not an accurate predictive parameter for root growth, as the total root system is inhibited under Al^{3+} stress. This has been observed in wheat and barley as well. However, no attempt was made to use the total root system as a 'phenotype' to predict Al^{3+} tolerance and to use it as a trait for the identification of loci previously associated with Al^{3+} tolerance. In order to accurately estimate total root growth, Famoso et al. (2010) used a custom-built system based on digital photography and semi-automated measurements of primary, secondary and tertiary rice roots using RootReader2D software. These authors identified some novel loci associated with Al^{3+} tolerance, in addition to the ones that were found with RRG measurements.

Recently Delhaize et al. (2012) reported significant genetic variation in the ability of wheat to form rhizosheaths on acid soil and assessed whether differences in (Al^{3+}) tolerance of root hairs between genotypes was the physiological basis for this genetic variation. A phenotyping method was developed to rapidly screen rhizosheath size in a range of wheat genotypes. This study found that a positive correlation existed between rhizosheath size on acid soil and root hair length and concluded that greater Al^{3+} tolerance of root hairs underlies the larger rhizosheath of wheat grown on acid soil. Analysis of wheat lines with *TaALMT1* gene-based markers revealed that tolerance of the root hairs to Al^{3+} was largely independent of the *TaALMT1* gene, which suggests that different genes encode the Al^{3+} tolerance of root hairs.

8.3 Dissection of Al^{3+} Tolerance Loci

Al^{3+} tolerance has been proposed generally due to a single major locus that account for most of genotypic variation in wheat (Kerridge and Kronstad 1968; Delhaize et al. 1993a; Baier et al. 1995; Luo and Dvorak 1996; Somers et al. 1996; Johnson et al. 1997; Milla and Gustafson 2001; Raman et al. 2005b); barley (Reid et al. 1971; Tang et al. 2000; Raman et al. 2003; Ma et al. 2004; Wang et al. 2006b, 2007), rye (Zhang and Jessop 1998), oat (Wight et al. 2006), and sorghum (Gourley et al. 1990; Magalhaes et al. 2004). However, multigenic inheritance for Al^{3+} tolerance has also been observed in these crops as well as in rice and maize (Reid 1971; Berzonsky 1992; Lima et al. 1992; Nguyen et al. 2001, 2002; Echart et al. 2002; Ninamango-Cardenas et al. 2003; Raman et al. 2005b).

8.4 Molecular Marker Systems for Mapping of Al^{3+} Tolerance Loci

The development of molecular markers for the detection and further utilisation of DNA polymorphisms has been the 'hallmark' in molecular biology and modern plant breeding programs. Various molecular marker systems based upon randomly amplified polymorphic DNA- RAPDs (Philipp et al. 1994; Loarce et al. 1996; Senft and Wricke 1996; Masojć et al. 2001); restriction fragment length polymorphism-RFLPs (Riede and Anderson 1996; Tang et al. 2000); simple sequence repeat-SSR (Saal and Wricke 1999; Masojć et al. 2001; Raman et al. 2001, 2002, 2006; Ma et al. 2005; Wang et al. 2007; Cai et al. 2008); amplified fragment length polymorphisms-AFLP (Wu et al. 2000; Miftahudin et al. 2002; Raman et al. 2002); diversity arrays technology-DArT (Wenzl et al. 2006; Wang et al. 2007) and candidate gene markers (Raman et al. 2005b, 2008; Fontecha et al. 2007; Wang et al. 2007) have been utilised to tag Al^{3+} tolerance loci. In recent years, DNA sequencing technologies have increased efficiency of genotyping, particularly in species where a high quality reference genome scaffold is available (Altshuler 2000; Davey 2011). New massively parallel genotyping technologies based upon single nucleotide polymorphism (SNP) and genotyping by sequencing (GBS) have been developed in maize and other crop plants (Baird et al. 2008; Elshire et al. 2011), which allow inexpensive identification of genome–wide markers. So far, although these technologies have not been widely utilised for mapping Al^{3+} tolerance loci (Krill et al. 2010; Famoso et al. 2011), it is anticipated that they will become an integrated part of future crop breeding programs.

Both major and minor QTL loci associated with Al^{3+} tolerance have been identified using molecular markers (Table 8.2). Genetic dissection of loci controlling Al^{3+} tolerance has been accomplished following principally two approaches; (i) linkage mapping and (ii) association mapping analysis. Basically both linkage and association mapping approaches rely on linkage disequilibrium (LD). Linkage mapping (utilising qualitative/Mendelian trait data such as Al^{3+} tolerant and Al^{3+}-sensitive, and quantitative trait data such as root regrowth (mm in length) is performed in structured biparental populations (such as doubled haploid, recombinant inbred lines, intercross and backcross populations), whereas association mapping (AM) is performed in unstructured populations such as germplasm collections and genetically diverse genotypes that have diverse ancestry (Stich et al. 2007, 2008). However, in practice both types of analyses (linkage and association mapping) are done simultaneously in order to validate true trait-marker associations and reduce the false-positive associations (Raman et al. 2010). Various biological factors such as rate of recombination, gene density, reproduction system, genetic diversity within genetic pool, genetic selection, population structure, genetic drift, genetic bottleneck during and after domestication, and heterozygousity/heterogeneity affect LD in any given species (Akhunov et al. 2003; Flint-Garcia et al. 2003; Smith et al. 2005; Yu et al. 2006, 2008; Buckler et al. 2009; Yan et al. 2009; Soto-Cerda and Cloutier 2012). The major advantages of AM over linkage mapping is: (i) it allows

for the utilization of germplasm and advanced breeding lines rather than structured segregating populations, and (ii) has more power to detect LD due to the historical recombination accumulated over several years of breeding and selection (Soto-Cerda and Cloutier 2012). Buckler and Thornsberry (2002) reported that linkage mapping has strong statistical power and is useful for understanding how and to what extent allelic effects are dependent on one another, but provides low genetic resolution unless the population is very large, whereas AM evaluates a large number of alleles. This approach has been recently exploited in wheat, triticale, maize and rice (Krill et al. 2010; Raman et al. 2010; Famoso et al. 2011; Niedziela et al. 2012).

8.4.1 (1) Dissection of Al^{3+} Tolerance Genes in Wheat

Several studies have shown that Al^{3+} tolerance in wheat is a complex trait and is controlled by a number of loci localised on the chromosome 2A, 3B, 4B, and 4D (Raman and Gustafson 2010). However a majority of these studies identified the major locus for Al^{3+} tolerance on the long arm of chromosome 4D (4DL) in diverse populations irrespective of their geographic origins (Luo and Dvorak 1996; Riede and Anderson 1996; Ma et al. 2005; Raman et al. 2005b, 2006, 2008, 2009; Navakode et al. 2009a).

In a RIL population derived from Atlas 66/Century, one major QTL on the 4DL was mapped where a malate transporter gene was located (Ma et al. 2005). This QTL accounted for approximately 50 % of the phenotypic variation for Al-tolerance. Two SSR markers Xgdm125 and Xwmc331 flanked the QTL as reported in other populations (Milla and Gustafson 2001; Raman et al. 2005b).

Zhou et al. (2007a) identified two QTL for Al^{3+} tolerance localized to chromosome arms 3BL and 4DL in a RIL population from Atlas66/Chisholm. Both QTL accounted for approximately 50–57 % of the variation for net root growth and hematoxylin staining scores, however their effects were not additive because expression of the minor QTL on 3BL (accounting for approximately 11 % of the variation) had epistatic interaction by the major locus on 4DL (accounting for approximately 45 % of the variation).

Cai et al. (2008) identified three significant QTL: *Qalt.pser-2A, Qalt.pser-3BL* and *Qalt.pser-4DL*, for Al^{3+} tolerance on chromosomes 2A, 3B, and 4D respectively in an RIL (recombinant inbred line) population from a cross between Chinese wheat lines FSW (Al^{3+}-tolerant) and ND35 (Al^{3+}-sensitive). These three QTL accounted for approximately 78–82 % of the phenotypic variation for Al^{3+} tolerance measured by net root growth and hematoxylin stain score. This study also showed that both QTL on the long arms of chromosomes 4D and 3B (3BL) are additive and hence can be used to improve Al^{3+} tolerance in wheat breeding programs.

Among the genomic regions conditioning Al^{3+} tolerance identified so far, the 4DL locus is co-localised with the malate efflux 'phenotype'- a Mendelised trait, in two DH populations from Diamondbird (Al^{3+}-tolerant)/Janz (Al^{3+}-sensitive), CD87 (Al^{3+} sensitive)/Currawong (Al^{3+} tolerant) and the *A*luminium *M*alate *T*ransporter

gene in wheat designated as *TaALMT1* (Raman et al. 2008) originally named *ALMT1*, (Sasaki et al. 2004, 2006; Yamaguchi et al. 2005) in several populations derived Diamondbird/Janz, CD87/Currawong, Spica (Al^{3+} sensitive)/Maringa (Al^{3+} tolerant), Cranbrook (Al^{3+} tolerant)/Halberd (Al^{3+} sensitive), Sunco (Al^{3+} sensitive)/Tasman (Al^{3+} tolerant), Atlas66 (Al^{3+} tolerant)/Century (Al^{3+} sensitive), Atlas66/Chisholm (Al^{3+} sensitive) and FSW (Al^{3+} tolerant)/ND35 (Al^{3+} sensitive), suggesting that *TaALMT1* conditions Al^{3+} tolerance in these populations. A strong correlation between malate efflux and Al^{3+} tolerance in wheat was observed (Sasaki et al. 2006), which suggested that malate efflux is the primary mechanism for Al^{3+} tolerance, a hypothesis originally proposed by Ryan et al. (1995b). Heterologous expression of *TaALMT1* in tobacco (*Nicotiana tabacum L.*) suspension cells and barley confirmed *TaALMT1* to be an Al-tolerance gene (Delhaize et al. 2004; Sasaki et al. 2004). Homologs and paralogs of *TaALMT1* were also found in *Arabidopsis*, wheat, barley, maize, rye, lupin and in rapeseed (*Brassica napus* L.) (Gruber et al. 2006; Hoekenga et al. 2006; Ligaba et al. 2006; Delhaize et al. 2007; Fontecha et al. 2007; Collins et al. 2008; Liu et al. 2008; Pineros et al. 2008), suggesting that these genes are not only conserved in monocots.

In order to investigate sequence divergence, *TaALMT1* was isolated from genomic DNA and characterized from 12 different wheat genotypes (Sasaki et al. 2004; Raman et al. 2005b). Molecular analysis has indicated that *TaALMT1* gene is 3968-bp long and consists of six exons (1388 bp) and 5 introns ranging from 0.1 to 1.8 kb encoding a membrane-localised transporter (Yamaguchi et al. 2005). Genetically diverse wheat genotypes had two alleles *TaALMT1-1* and *TaALMT1-2* that differ by six nucleotides (SNPs) of which only two nucleotides encode for different amino acids in the predicted protein (Sasaki et al. 2004; Raman et al. 2005b). Among 12 genotypes, *TaALMT1* exhibited at least 44 SNPs or small insertions/deletions (InDels) (Raman et al. 2005b). These polymorphisms were mainly in the introns in addition to 6 SNPs in the exons. One of the 6 SNPs, in exon 4, was used to develop a CAPS marker to distinguish *TaALMT1-1* from *TaALMT1-2* (Sasaki et al. 2004). The third intron region is the largest and shows considerable allelic variability due to simple sequence repeat motifs (SSR) with variable copy numbers and InDels (Raman et al. 2005b, 2006, 2008). As expression of *TaALMT1* showed significant correlation with the level of Al^{3+} tolerance in diverse wheat genotypes (Sasaki et al. 2004; Raman et al. 2005b), therefore, upstream and downstream sequence of the *TaALMT1* was characterized to identify allelic variants in 69 wheat lines (Sasaki et al. 2006). The first 1,000 bp upstream of the *TaALMT1* coding region was more variable and six different promoter patterns could be discerned (types I–VI). Type I had the simplest structure, while the others had blocks of sequence that were duplicated or triplicated in different arrangements (Sasaki et al. 2006). Besides, allelic variants of upstream region of *TaALMT1* were also reported in highly diverse germplasm comprising wheat cultivars, subspecies and landraces of common and spelt wheat (Raman et al. 2008, 2009, 2010).

Al^{3+}-activated malate release is not the only Al^{3+} tolerance mechanism in wheat. Ryan et al. (2009) described a second mechanism and mapped the underlying locus, Xce_c responsible for citrate efflux on the long arm of chromosome 4B (4BL)

using an F_2 population derived from a cross between Carazinho (citrate and malate efflux, Al^{3+}-tolerant) and the cultivar EGA-Burke (malate efflux without citrate efflux, Al^{3+} tolerant). The Xce_c was delimited with markers GWM495 and wPT-8397 (Table 8.2). This linkage was validated in an independent F_2 population derived from Egret/Carazinho and markers predicted 91–96 % of phenotypic variation for citrate efflux. This study showed that citrate efflux not only occurred constitutively from the roots of Carazinho, but also in other Brazilian cultivars Maringa, Toropi, and Trintecinco and the expression of an expressed sequence tag, belonging to the multidrug and toxin efflux (*MATE*) gene family, correlated with the citrate efflux phenotype. The *MATE* family proteins are proposed to transport small, organic compounds (Omote et al. 2006), and are the members of a large and complex family of transporters. The homolog of this *MATE* exists in rice, sorghum (Magalhaes et al. 2007), and white lupins (*Lupinus albus L.*) (Uhde-Stone et al. 2005). Enhanced Al^{3+} tolerance is generally correlated with higher expression of these tolerance genes, regardless of whether they are from *TaALMT1 or MATE* family, (Sasaki et al. 2004; Raman et al. 2005b; Hoekenga et al. 2006; Magalhaes et al. 2007; Fujii et al. 2012).

Raman et al. (2010) performed genome wide association analysis and identified genomic regions (markers) that were significantly associated with Al^{3+} tolerance on chromosomes 1A, 1B, 2A, 2B, 2D, 3A, 3B, 4A, 4B, 4D, 5B, 6A, 6B, 7A and 7B using 1,055 accessions of common wheat from different geographic regions of the world. Some of these genomic regions corresponded to previously identified loci on chromosome 2A, 3B and 4D for Al^{3+} tolerance whereas others appear to be novel. Among the markers for *TaALMT1*, the major Al^{3+} tolerance gene located on 4D, those that targeted the promoter explained most of the phenotypic variance for Al^{3+} tolerance, which is consistent with this region controlling the level of *TaALMT1* expression. The results demonstrated that genome-wide association mapping can both confirm known Al^{3+}-tolerance loci, such as those on 3B, 4D and 4B, and also identify novel tolerance loci.

8.5 Multiple Origin of Al^{3+} Tolerance in Wheat

Previously, it was accepted that Al^{3+} tolerance is originated/evolved in Brazil, as majority of wheat cultivars tested were tolerant to acid soils and Al^{3+} toxicity, and may have originated from Brazilian landraces Polyssu and/or Alfredo Chavez 21 (Garvin and Carver 2003). However, recent studies suggest that Al^{3+} tolerance evolved from different sources originated from diverse geographic origins. For example, Stodart et al. (2007) evaluated 250 accessions of wheat landraces from 21 countries for tolerance to Al^{3+} and found 35 diverse accessions assessed using molecular markers were tolerant. These accessions were originated from Bulgaria, Croatia, India, Italy, Nepal, Spain, Tunisia and Turkey. Genetic analysis of wheat genotypes collected from various regions also indicated that Al^{3+} tolerance may have originated independently (Stodart et al. 2007; Zhou et al. 2007b; Hu et al. 2008; Raman et al. 2008) and several genes may control Al^{3+} tolerance.

Ryan et al. (2010) investigated the evolution of this trait in wheat utilising 760 accessions of *T. aestivum* (genomes AABBDD), *T. turgidum* (genome AABB), *T. monococcum* (genome AA), *T. urartu* (genome AA), *T. dicoccoides* (genome AABB), *T. timopheevii* (genomes AAGG), *T. zhukovskyi,* and *Ae. tauschii* (genome *DD*) and showed that none of the accessions with A and/or B genomes alone displayed significant Al^{3+} tolerance. However, five of the *Ae. tauschii* accessions (AUS110802, AUS21711, AUS18913, AUS110812 and AUS110668) displayed moderate levels of tolerance compared to the tolerant hexaploid genotype ET8. One origin of Al^{3+} tolerance is highly likely to be from *Ae. tauschii;* additional Al^{3+} tolerance may have arisen more recently due to mutations within the *TaALMT1* promoter (Raman et al. 2008; Ryan et al. 2010). These mutations were able to increase the Al^{3+} tolerance above the levels present in *Ae. tauschii.* These authors further demonstrated that the tandemly repeated elements act to enhance gene expression using transformation experiments. Pereira et al. (2010) showed that the Al^{3+} tolerance of wheat was increased by over-expressing *TaALMT1* through the use of stable plant transformation and a variety of strong promoter constructs.

8.6 Dissection of Al^{3+} Tolerance Genes in Barley

Genetic linkage analyses of several mapping populations have shown that a single major locus, located on the long arm of chromosome 4H (4HL) controls Al^{3+} tolerance in barley (Tang et al. 2000; Raman et al. 2001, 2002, 2003; Wang et al. 2004, 2006a, b, 2007; Inostroza-Blancheteau et al. 2010) (Table 8.2). Single gene inheritance of Al^{3+} tolerance was revealed using trisomic analysis in barley (Minella and Sorrells 1997). Loci on chromosomes 2H, 3H, 4H, 5H and 6H seem to contribute towards Al^{3+} tolerance, as indicated in two independent but preliminary studies (Raman et al. 2005a; Navakode et al. 2009b). However, these loci need to be validated through additional experimentation. As in wheat, the genetic variation for Al^{3+} tolerance in barley has been correlated with the citrate efflux from root-apices in the doubled haploid populations from Dayton (Al^{3+}-tolerant, high citrate efflux)/Gairdner (Al^{3+}-sensitive and no citrate efflux), and Murasakimochi (Al^{3+}-tolerant, high citrate efflux)/Morex (Al^{3+}-sensitive and no citrate efflux) cultivars (Ma et al. 2004; Wang et al. 2007). The secretion of citrate in barley shows a pattern distinct from wheat, characterised by rapid, non-dose-responsive and temperature dependent pattern of organic acid release (Ma et al. 2004).

In order to identify genes underlying Al^{3+}-stimulated citrate efflux, a candidate gene based approach was taken using primers targeting *MATE* genes similar to those from sorghum (*SbMATE;* Magalhaes et al. 2007). The *HvMATE* gene—homolog of *MATE* in barley, showed cosegregation with Al^{3+} tolerance in an F_2 population from Dayton/Gairdner, where Al^{3+} activated citrate release had already been demonstrated to be the method of Al^{3+} tolerance (Wang et al. 2007). This study suggested that *HvMATE* may be the causative gene for Al^{3+} tolerance. In an independent study, Furukawa et al. (2007) used both genetic mapping and microarray analyses

to identify*HvAACT1* in the Murasakimochi/Morex population, Sequence alignment indicates that *HvMATE* and *HvAACT1* represent the same locus. High correlation between *HvAACT1/HvMATE1* expression with citrate efflux and Al^{3+} tolerance was shown in barley genotypes (Furukawa et al. 2007; Wang et al. 2007), where the relative expression of the *HvMATE* gene was 30-fold greater in 'Dayton' (Al^{3+} tolerant) than the Al^{3+} sensitive cv. 'Gairdner' (Wang et al. 2007). When expressed in *Xenopus* oocytes, the *HvAACT1/HvMATE* protein mediated the efflux of citrate. The candidacy of *HvAACT1* gene was further tested by genetic transformation, transgenic tobacco (*Nicotiana tabacum*) expressing *HvAACT1* showed higher citrate secretion in the presence of Al^{3+} and exhibited higher tolerance to Al^{3+}, but the citrate secretion was not altered in the absence of Al^{3+} despite the constitutive promoter in the heterologous host (Furukawa et al. 2007).

Fujii et al. (2012) analysed the genomic sequences of *HvAACT1* belonging to the *MATE* family and found that an insertion of 1023 bp in the upstream of *HvAACT1* coding region of the Al^{3+} tolerant cultivar Murasakimochi and other 20 cultivars, however this insertion was not present in Morex and other 225 Al^{3+} sensitive cultivars belonging to cultivated *H. vulgare* ssp *vulgare* and *H. vulgare* ssp *spontaneum*. The candidacy of this *HvAACT1* containing insertion (from Murasakimochi*)* was tested using genetic transformation of Al^{3+}-sensitive barley cultivar Golden Promise. Transgenic lines showed high level of *HvAACT1* expression in the root tips, when driven by the Murasakimochi promoter, but were unaffected in those driven by the Morex (Al^{3+} sensitive) promoter. The enhanced expression of *HvAACT1* in the root zone resulted in increased citrate secretion to the rhizosphere, thereby protecting the root tips from Al^{3+} toxicity. Authors further suggested that the occurrence of the 1-Kb insertion in the upstream region of *HvAACT1* gene might originate in the Eastern region where acid soils are widely distributed.

8.7 Dissection of Al^{3+} Tolerance Genes in Rice

QTL analysis revealed that a total of 33 QTL for Al^{3+} tolerance using six different inter- and intra-specific mapping populations (Table 8.2) on 12 chromosomes (Wu et al. 2000; Nguyen et al. 2001, 2002, 2003; Ma et al. 2002; Xue et al. 2007). Recently, Famoso et al. (2011) performed a genome-wide association analysis with 383 diverse rice accessions, and QTL mapping in two bi-parental populations using three estimates of Al^{3+} tolerance based on root growth and identified forty-eight regions associated with Al^{3+} tolerance. Four of these regions co-localized with a priori candidate genes, and two highly significant regions co-localized with previously identified QTL. Three regions corresponding to induced Al^{3+} sensitive rice mutants (*ART1, STAR2, and Nrat1*) were identified through bi-parental QTL mapping or GWA to be involved in natural variation for Al^{3+} tolerance. Among the candidate genes identified so far, only a few of them are collocated at the similar positions of Al^{3+} tolerance QTL (Yamajia et al. 2009).

In rice, Al tolerance is mediated by a novel mechanism, independent of root tip Al exclusion (Iuchi et al. 2007). However, it was recently established that rice secretes citrate in much lower amounts compared to other cereal crops such as rye and wheat (Ma et al. 2002) and the citrate efflux does contribute to tolerance via an Al^{3+} activated *MATE* transporter designated *OSFRDL4* (Yokosho et al. 2011). This study characterized the functions of *OsFRDL4* that belongs to the *MATE* family in rice. Heterologous expression in *Xenopus* oocyte showed that the *OsFRDL4* protein was able to transport citrate and was activated by Al^{3+}. The expression level of the *OsFRDL4* gene in roots was very low in the absence of Al^{3+}, but was greatly enhanced by Al^{3+} after short exposure. Furthermore, the *OsFRDL4* expression was regulated by *ART1*, a C_2H_2-type zinc finger transcription factor for Al^{3+} tolerance. Transient expression of *OsFRDL4* in onion (*Allium cepa L.*) epidermal cells showed that it localized to the plasma membrane. Immunostaining showed that *OsFRDL4* was localized in all cells in the root tip. These expression patterns and cell specificity of localization of *OsFRDL4* are different from other *MATE* members identified previously. Knockout of *OsFRDL4* resulted in decreased Al^{3+} tolerance and decreased citrate secretion compared with the wild-type rice, but did not affect citrate concentration in the xylem sap. Furthermore, there is a positive correlation between *OsFRDL4* expression level and the amount of citrate secretion in rice cultivars that are differing in Al tolerance. Taken together, the results show that *OsFRDL4* is an Al^{3+}-induced citrate transporter localized at the plasma membrane of rice root cells and is one of the components of high Al^{3+} tolerance in rice.

Recently, genes encoding a nucleotide binding domain (designated as *STAR1*) and gene encoding for a transmembrane domain, of a bacterial type ATP binding cassette (ABC) transporter (designated as *STAR2*) are shown to be regulated by Al^{3+} in rice (Huang et al. 2009). Expression in onion epidermal cells, rice protoplast, and yeast (*Saccharomyces cerevisiae*) showed that *STAR1* interacts with *STAR2*. This study showed that proteins encoded by these genes form a complex that functions as a unique ATP binding cassette, which is required for Al^{3+} tolerance. While the positions of *STAR1* and *STAR2* do not coincide with the QTLs associated with Al^{3+} tolerance reported so far, a segment of chromosome 5 containing *STAR2* was identified by GWA for Al^{3+} tolerance (Famoso et al. 2011). However, at the present level of resolution with GWA does not build a compelling argument, such that additional mapping studies are required.

Yamajia et al. (2009) reported an Al^{3+} tolerance transcription factor 1 (*ART1*) that encodes a transcription factor that regulates 31 genes implicated in Al^{3+} tolerance, including *STAR1* and *STAR2* in rice. Some of these genes are implicated in both internal and external detoxification of Al^{3+} at different cellular levels.

Besides, *Nrat1* genes for *Aluminium rhizotoxicity 1* and *Nramp aluminium transporter 1* have also recently been cloned (Yamajia et al. 2009; Xia et al. 2010). Expression of *Nrat1* is shown to be up-regulated by Al^{3+} in the roots and regulated by a C_2H_2 zinc finger transcription factor 1 (*ART1*) and the protein is located on the plasma membranes of root cells near the apex. Knockout mutations of *Nrat1* decrease Al^{3+} uptake, increase sensitivity to Al^{3+}. GWA has also implicated *Nrat1* as a contributor to Al^{3+} tolerance in at least one sub-population of rice (Famoso et al. 2011).

Table 8.2 Linkage of aluminium tolerance loci with PCR-based markers suitable for marker-assisted selection in cereal crops (updated from Raman and Gustafson 2010)

Screening Method[a]	Population	Chromosome location	Linked Marker(s)	Reference
Barley (*Hordeum vulgare* L)				
RG	Yambla/WB229	4HL	Bmag353, Bmac310 HVM68	Raman et al. (2002)
	WB229/Mimosa	4HL	Bmag353, Bmac310 HVM68	Raman et al. (2002)
	Harrington/ Brindabella	4HL	Bmag353, Bmac310	Raman et al. (2001)
	Ohichi/F6ant-28B48–16	4HL	Bmag353, Bmac310	Raman et al. (2005b)
	Dayton/Zhepi2	4HL	Bmag353, HvMATE HVM68, GBM1071	Wang et al. (2007)
	Dayton/F6ant-28B48–16	4HL	Bmag353, Bmac310, HVM68	Raman et al. (2003)
RRG	Dayton/Zhepi2	4HL	HvGABP, Bmag353, HVM68, HvMATE, GBM1071	Wang et al. (2007)
Root tolerance index	OWB_{DOM}/OWB_{REC}	2H, 3H, 4H	GBR441, GBM1043, GBM1233, GBM1251, HVM3	Navakode et al. (2009b)
Hematoxylin Staining	Dayton/Harlan Hybrid	4HL	Bmag353, Bmac310, HVM68	Raman et al. (2003)
Eriochrome cyanine	Dayton/Zhepi2	4HL	HvGABP, Bmag353, HVM68, HvMATE, GBM1071	Wang et al. (2007)
	Dayton/Gairdner	4HL	HvGABP, ABG715, GWM165, Bmag353	Wang et al. (2007)
	F6ant28B48–16/Honen	4HL	Bmag353, HVM68	Wang et al. (2006b)
Root/shoot fresh wt ratio	Murasakimochi/ Morex	4HL	Bmag353	Ma et al. (2004)
Wheat (*Triticum aestivum* L)				
Hematoxylin	Diamondbird/Janz	4DL	TaALMT1, WMC331	Raman et al. (2003, 2005b, 2008)
	Currawong/CD87	4DL	TaALMT1, WMC331	Raman et al. (2005b, 2008)
	Spica/Maringa	4DL	TaALMT1, GWM165	Raman et al. (2005b, 2008)
	Atlas66/Century	4DL	WMC125, GDM125 TaALMT1	Ma et al. (2005)

Table 8.2 (continued)

Screening Method[a]	Population	Chromosome location	Linked Marker(s)	Reference
Root growth	BH1146/Anahuac	4DL	BCD1230, GDM125 TaALMT1	Riede and Anderson (1996); Milla and Gustafson (2001); Raman et al. (2008)
RRG	Diamondbird/Janz	4DL	TaALMT1, WMC331	Raman et al. (2005)
	Cranbrook/ Halberd	4DL	TaALMT1	Raman et al. (2005, 2008)
	Sunco/Tasman	4DL	TaALMT1	Raman et al. (2005, 2008)
	Atlas66/Century	4DL	WMC125, GDM125 TaALMT1	Ma et al. (2005)
	Atlas66/Chisholm	4DL	TaALMT1, WMC331, GDM125	Zhou et al. (2007a)
		3BL	BARC164	Zhou et al. (2007a)
	FSW/ND35	4DL	TaALMT1	Cai et al. (2008)
		3B	BARC164, BARC344	Cai et al. (2008)
		2A	GWM515, GWM296	Cai et al. (2008)
	EGA-Burke/ Carazhino	4BL	GWM495, GWM513	Ryan et al. (2009)
	Egret/Carazhino	4BL	GWM495, GWM513	Ryan et al. (2009)
Rice (*Oryza sativa* L)				
RG, RRG	IR64/*O. rufipogon*	QTLs	RFLP/SSR markers	Nguyen et al. (2003)
	CT9933/IR62266	QTLs	RFLP/SSR markers	Nguyen et al. (2002)
Rye (*Secale cereale* L)				
RG, RRG	Ailes/Riodeva *(Alt1)*	6RS	$ScR01_{600}$, $ScB15_{790}$	Gallego and Benito (1997)
	AR6–17, AR1–13		*SCR01600*	Gallego et al. (1998a, b)
	Ailes/Riodeva (*Alt3*)	4RL	$ScOPS17_{705}$	Benito et al. (2009)
	M39A-1–6/M77A-1 (*Alt4*)	7RS	B1, B4, B11, B25, B26, B27	Collins et al. (2008)
			BCD1230	Miftahudin et al. (2002, 2004, 2005)

Table 8.2 (continued)

Screening Method[a]	Population	Chromosome location	Linked Marker(s)	Reference
	Ailes/Riodeva *(Alt4)*	7RS	B1, B4, B26, ScALMT1	Benito et al. (2009); Fontecha et al. (2007)
Oats (*Avena sativa* L)	CIav2921/CIav9011		SCA08 and calretB1_3	Wight et al. (2006)
Maize *(Zea mays L.)*	L53/L1327	QTLs	SSR	Ninamango-Cardenas et al. (2003)
RG	Association Panel	1, 4, 6 and 10	SNPs	Krill et al. (2010)

[a] *RRG/RRE* Relative root growth, *RG/RE* Root growth/Root elongation, *RFLP* Restriction fragment length polymorphism

8.8 Molecular Breeding for Al Tolerance Using Marker-Assisted Selection (MAS)

Al^{3+} toxicity can be ameliorated by the repeated applications of lime, however it is not cost effective option especially in the developing countries. Molecular markers would enhance the development of varieties tolerant to Al^{3+} toxicity, and as a surrogate trait for tolerance to soil acidity. In order to increase selection efficiency of MAS for Al^{3+} tolerance, diagnostic markers based on functionally-associated variation in the candidate genes such as *TaALMT1, HvMATE*, and *SbMATE* have been developed. Unfortunately, some of these markers (e.g. for *TaALMT1*) were not found to be 'diagnostic' as a number of alleles were identified in different Al^{3+} tolerant genotypes (Raman et al. 2005b, 2008). Nevertheless these candidate gene-based functional markers are preferred for AM studies and MAS as compared with linked markers. Table 8.2 describes the linkage between markers based upon RFLP, AFLP, SSR, DArT, and SNP, and Al^{3+} tolerance loci in different cereals. Some of these Al^{3+} tolerance-marker associations have been validated in different genetic backgrounds (Raman et al. 2002, 2005b, 2008; Wang et al. 2006b, 2007).

8.9 Conclusions

During the last decade, significant achievements have been made in the understanding of genetic variation, physiological and molecular mechanisms, and candidate/causative genes controlling Al^{3+} tolerance. Genotypic variation in Al^{3+} tolerance for wheat, barley, sorghum, rye, and rice is positively correlated with the level of *ALMT* and *MATE* gene expression in root apices. Molecular markers within

these candidate genes have been identified and will provide excellent tools to enrich desirable Al^{3+} tolerance alleles into the superior genotypes. Molecular markers have also been utilised to reveal allelic diversity as well as trace origin and transmittance of genes in germplasm collections, landraces, and contemporary cultivars. This has provided new insights into the evolution and existence of novel alleles for Al^{3+} tolerance, still largely unexplored in the natural cereal germplasm including in traditional landrace cultivars. It has also been shown that aluminium tolerance lines enhance phosphorus nutrition and grain production when grown on acid soil, suggesting that Al^{3+} tolerant cultivars that secrete organic acids such as malate may enhance bioavailability of phosphorus to the plant.

References

Akhunov ED, Goodyear AW, Geng S et al (2003) The organization and rate of evolution of wheat genomes are correlated with recombination rates along chromosome arms. Genome Res 13: 753–763

Altshuler D, Pollara VJ, Cowles CR et al (2000) An SNP map of the human genome genome generated by reduced representation shotgun sequencing. Nature 407:513–516

Baier AC, Somers DJ, Gustafson JP (1995) Aluminium tolerance in wheat: correlating hydroponic evaluations with field and soil performances. Plant Breed 114:291–296

Baird NA, Etter PD, Atwood TS et al (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE 3:e3376

Benito C, Silva-Navas J, Fontecha G et al (2009) From the rye *Alt3* and *Alt4* aluminum tolerance loci to orthologous genes in other cereals. Plant Soil 327:107–120

Bennet RJ (1997) The response of lucerne and red clover roots to aluminium/Hematoxylin: How universal is the hematoxylin test for aluminium? S Afr J Plant Soil 14:120–125

Berzonsky WA (1992) The genomic inheritance of aluminium tolerance in 'Atlas 66' wheat. Genome 35:689–693

Bona L, Wright RJ, Baligar VC, Matuz J (1993) Screening wheat and other small grains for acid soil tolerance. Landsc Urban Plan 27:175–178

Buckler ES, Thornsberry JM (2002) Plant molecular diversity and applications to genomics. Curr Opin Plant Biol 5:107–111

Buckler ES, Holland JB, Bradbury PJ et al (2009) The genetic architecture of maize flowering time. Science 325:714–718

Cai SB, Bai GH, Zhang DD (2008) Quantitative trait loci for aluminum resistance in Chinese wheat landrace FSW. Theor Appl Genet 117:49–56

Camargo CEdO, Felicio JC, Ferreira F (1989) Wheat breeding: XXI. Evaluation of inbred lines in different regions of the state of São Paulo, Brazil. Bragantia 48:53–71

Camargo CEdO, Filho F, Penteado AW et al (1992) Wheat breeding: XXVII. Variance, heritability and correlations in hybrid populations for grain yield, tolerance to aluminum toxicity and plant height. Bragantia 51:21–30

Cançado GMA, Loguercio LL, Martins PR et al (1999) Hematoxylin staining as a phenotypic index for aluminum tolerance selection in tropical maize (*Zea mays* L.). Theor Appl Genet 99:747–754

Carver BF, Whitmore WE, Smith EL, Bona L (1993) Registration of four aluminum-tolerant winter wheat germplasms and two susceptible near-isolines. Crop Sci 33:1113–1114

Collins NC, Shirley NJ, Saeed M et al (2008) An *ALMT1* gene cluster controlling aluminum tolerance at the *Alt4* locus of rye (*Secale cereale* L.). Genetics 179:669–682

Davey JW, Hohenlohe PA, Etter PD et al (2011) Genome-wide genetic marker discovery and genotyping using next generation sequencing. Nat Rev Genet 12:499–510

Delhaize E, Craig S, Beaton CD et al (1993a) Aluminum tolerance in wheat (*Triticum aestivum* L.) I. Uptake and distribution of aluminum in root apices. Plant Physiol 103:685–693

Delhaize E, Ryan PR, Hebb DM, Yamamoto Y, Sasaki T, Matsumoto H (2004) Engineering high-level aluminum tolerance in barley with the *ALMT1* gene. PNAS 101: 15249–15254

Delhaize E, Ryan PR, Randall PJ (1993b) Aluminum tolerance in wheat (*Triticum aestivum* L.) II. Aluminum-stimulated excretion of malic acid from root apices. Plant Physiol 103:695–702

Delhaize E, Gruber BD, Ryan PR (2007) The roles of organic anion permeases in aluminium resistance and mineral nutrition. FEBS Lett 581:2255–2262

Delhaize E, James RA, Ryan PR (2012) Aluminium tolerance of root hairs underlies genotypic differences in rhizosheath size of wheat (*Triticum aestivum*) grown on acid soil. New Phytol 195:609–619

Echart CL, Barbosa-Neto JF, Garvin DF et al (2002) Aluminum tolerance in barley: methods for screening and genetic analysis. Euphytica 126:309–313

Elshire RJ, Glaubitz JC, Sun Q et al (2011) A robust, simple Genotyping-by-Sequencing (GBS) approach for high diversity species. PLoS ONE 6:e19379

Famoso AN, Clark RT, Shaff JE et al (2010) Development of a novel aluminum tolerance phenotyping platform used for comparisons of cereal aluminum tolerance and investigations into rice aluminum tolerance mechanisms. Plant Physiol 153:1678–1691

Famoso AN, Zhao K, Clark RT et al (2011) Genetic architecture of aluminium tolerance in rice (*Oryza sativa*) determined through genome-wide association analysis and QTL mapping. PLoS Genet 7(8):e1002221

Flint-Garcia SA, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrium in plants. Annu Rev Plant Biol 54:357–374

Fontecha G, Silva-Navas J, Benito C et al (2007) Candidate gene identification of an aluminum-activated organic acid transporter gene at the *Alt4* locus for aluminum tolerance in rye (*Secale cereale* L.). Theor Appl Genet 114:249–260

Fujii M, Yamaji N, Sato K, Ma JF. 2009. Mechanism regulating HvAACT1 expression in barley. In: Liao H, Yan X, Kochian LV, eds. Plant–soil interactions at low pH: a nutriomic approach – Proceedings of the 7th International Symposium of Plant–Soil Interactions at Low pH. Guangzhou: South China University of Technology Press, 165–166.

Fujii M, Yokosho K, Yamaji N et al (2012) Acquisition of aluminium tolerance by modification of a single gene in barley. Nat Commun 3:713

Furukawa J, Yamaji N, Wang H et al (2007) An aluminum-activated citrate transporter in barley. Plant Cell Physiol 48:1081–1091

Gallego FJ, Benito C (1997) Genetic control of aluminium tolerance in rye (*Secale cereale* L.). Theor Appl Genet 95:393–399

Gallego FJ, Calles B, Benito C (1998a) Molecular markers linked to the aluminium tolerance gene *Alt1* in rye (*Secale cereale* L.). Theor Appl Genet 97:1104–1109

Gallego FJ, Lopez-Solanilla ELÃ, Figueiras AM, Benito C (1998b) Chromosomal location of PCR fragments as a source of DNA markers linked to aluminium tolerance genes in rye. Theor Appl Genet 96:426–434

Galvez L, Clark RB, Klepper LA, Hansen L (1991) Organic acid and free proline accumulation and nitrate reductase activity in sorghum (*Sorghum bicolor*) genotypes differing in aluminum tolerance. In: 'Plant-Soil Interaction at Low PH', Wright RJ, Baligar VC, Murrmann RP (eds) Kluwer Academic Publishers, Dordrecht, pp 859–867

Garvin DF, Carver BF (2003) Role of the genotype in tolerance of acidity and aluminium toxicity. In: Rengel Z (ed) Handbook of soil acidity. Marcel Dekker, Inc., New York, pp 387–406

Gourley LM, Rogers SA, Ruiz-Gomez C, Clark RB (1990) Genetic aspects of aluminium tolerance in sorghum. Plant Soil 123:211–216

Gruber B, Ryan P, Richardson A et al (2006) The identification and characterisation of *ALMT1* homologs in the Triticeae. Proceedings of 8th International Congress of Plant Molecular Biology, Adelaide, Australia, p 185

Hoekenga OA, Maron LG, Cancado GMA et al (2006) *AtALMT1*, which encodes a malate transporter, is identified as one of several genes critical for aluminum tolerance in *Arabidopsis*. Proc Natl Acad Sci USA 103:9738–9743

Horst WJ, Puschel AK, Schmohl N (1997) Induction of callose formation is a sensitive marker for genotypic aluminium sensitivity in maize. Plant Soil 192:23–30
Hu SW, Bai GH, Carver BF, Zhang DD (2008) Diverse origins of aluminium-resistance sources in wheat. Theor Appl Genet 118:29–41
Huang CF, Yamaji N, Mitani N et al (2009) A bacterial-type ABC transporter is involved in aluminum tolerance in rice. Plant Cell 21:655–667
Hue NV, Craddock GR, Adams F (1986) Effect of organic acids on aluminum toxicity in subsoils. Soil Sci Soc Am J 50:28–34
Inostroza-Blancheteau C, Soto B, Ibáñez C et al (2010) Mapping aluminum tolerance loci in cereals: a tool available for crop breeding. Electron J Biotech 13:doi:10.2225/vol2213-issue2224-fulltext-2224
Ishikawa S, Wagatsuma T, Sasaki R, Ofei-Manu P (2000) Comparison of the amount of citric and malic acids in Al media of seven plant species and two cultivars each in five plant species. Soil Sci Plant Nutr 46:751–758
Iuchi S, Koyama H, Iuchi A et al (2007) Zinc finger protein STOP1 is critical for proton tolerance in *Arabidopsis* and coregulates a key gene in aluminum tolerance. Proc Nat Acad Sci 104: 9900–9905
Johnson JP, Carver BF, Baligar VC (1997) Expression of aluminum tolerance transferred from Atlas 66 to hard winter wheat. Crop Sci 37:103–108
Jones DL (1998) Organic acids in the rhizosphere—a crtitical review. Plant Soil 205:25–44
Kerridge PC, Kronstad WE (1968) Evidence of genetic resistance to aluminium toxicity in wheat. Agronomy J 60:710–711
Kinraide TB, Parker DR, Zobel RW (2005) Organic acid secretion as a mechanism of aluminium resistance: a model incorporating the root cortex, epidermis, and the external unstirred layer. J Exp Bot 56:1853–1865
Kochian LV (1995) Cellular mechanisms of aluminum toxicity and resistance in plants. Annu Rev Plant Physiol Plant Mol Biol 46:237–260
Krill AM, Kirst M, Kochian LV et al (2010) Association and linkage analysis of aluminium tolerance genes in maize. PLoS ONE 5:e9958
Krizek DT, Foy CD (1988) Role of water stress in differential aluminium tolerance of two barley cultivar in acid soil. J Plant Nutr 11:351–367
Larsen PB, Degenhardt J, Tai CY et al (1998) Aluminum-resistant Arabidopsis mutants that exhibit altered patterns of aluminum accumulation and organic acid release from roots. Plant Physiol 117:9–18
Ligaba A, Katsuhara M, Ryan PR et al (2006) The *BnALMT1* and *BnALMT2* genes from rape encode aluminum-activated malate transporters that enhance the aluminum resistance of plant cells. Plant Physiol 142:1294–1303
Ligaba A, Maron L, Shaff J et al (2012) Maize ZmALMT2 is a root anion transporter that mediates constitutive root malate efflux. Plant Cell Environ 35:1185–1200
Lima M, Furlani PR, Miranda-Filho JB de (1992) Divergent selection for aluminium tolerance in a maize (*Zea mays* L.) population. Maydica 37:123–132
Liu J, Magalhaes JV, Shaff J et al (2008) Aluminum-activated citrate and malate transporters from the *MATE* and *ALMT* families function independently to confer Arabidopsis aluminum tolerance. Plant J 389–399
Loarce Y, Hueros G, Ferrer E (1996) A molecular linkage map of rye. Theor Appl Genet 93: 1112–1118
Luo MC, Dvorak J (1996) Molecular mapping of an aluminium tolerance locus on chromosome 4D of Chinese Spring wheat. Euphytica 91:31–35
Ma HX, Bai GH, Carver B, Zhou LL (2005) Molecular mapping of a quantitative trait locus for aluminum tolerance in wheat cultivar Atlas 66. Theor Appl Genet 112:51–57
Ma JF, Ryan PR, Delhaize E (2001) Aluminium tolerance in plants and the complexing role of organic acids. Trends Plant Sci 6:273–278
Ma JF, Shen R, Zhao Z et al (2002) Response of rice to Al stress and identification of quantitative trait loci for Al tolerance. Plant Cell Physiol 43:652–659
Ma JF, Nagao S, Sato K et al (2004) Molecular mapping of a gene responsible for Al-activated secretion of citrate in barley. J Exp Bot 55:1335–1341

Magalhaes JV, Garvin DF, Wang YH et al (2004) Comparative mapping of a major aluminum tolerance gene in sorghum and other species in the Poaceae. Genetics 167:1905–1914
Magalhaes JV, Liu J, Guimaraes CT et al (2007) A gene in the multidrug and toxic compound extrusion (*MATE*) family confers aluminum tolerance in sorghum. Nat Genet 39:1156–1161
Maltais K, Houde M (2002) A new biochemical marker for aluminium tolerance in plants. Physiol Plant 115:81–86
Maron LG, Pineros MA, Guimaraes CT et al (2010) Two functionally distinct members of the MATE (multi-drug and toxic compound extrusion) family of transporters potentially underlie two major aluminum tolerance QTLs in maize. Plant J 61:728–740
Masojć P, Mysków B, Milczarski P (2001) Extending a RFLP-based genetic map of rye using random amplified polymorphic DNA (RAPD) and isozyme markers. Theor Appl Genet 102:1273–1279
Massot N, Poschenrieder C, Barcelo J (1992) Differential response of three beans (*Phaselous vulagaris*) cultivars to aluminium. Acta Botanica Neerlandica 41:293–298
Massot N, Llugany M, Poschenrieder C, Barcelo J (1999) Callose production as indicator of aluminum toxicity in bean cultivars. J Plant Nutr 22:1–10
Miftahudin T, Scoles GJ, Gustafson JP (2002) AFLP markers tightly linked to the aluminum-tolerance gene *Alt3* in rye (*Secale cereale* L.). Theor Appl Genet 104:626–631
Miftahudin T, Scoles GJ, Gustafson JP (2004) Development of PCR-based codominant markers flanking the *Alt3* gene in rye. Genome 47:231–238
Miftahudin T, Chikmawati T, Ross K et al (2005) Targeting the aluminum tolerance gene *Alt3* region in rye, using rice/rye micro-colinearity. Theor Appl Genet 110:906–913
Milla R, Gustafson JP (2001) Genetic and physical characterization of chromosome 4DL in wheat. Genome 44:883–892
Minella E, Sorrells ME (1992) Aluminum tolerance in barley: Genetic relationships among genotypes of diverse origin. Crop Sci 32:593–598
Minella E, Sorrells ME (1997) Inheritance and chromosome location of *Alp*, a gene controlling aluminium tolerance in 'Dayton' barley. Plant Breed 116:465–469
Moroni JS, Briggs KG, Taylor GJ (1991) Pedigree analysis of the origin of manganese tolerance in Canadian spring wheat (*Triticum aestivum* L.) cultivars. Euphytica 56:107–120
Navakode S, Weidner A, Lohwasser U et al (2009a) Molecular mapping of quantitative trait loci (QTLs) controlling aluminium tolerance in bread wheat. Euphytica 166:283–290
Navakode S, Weidner A, Varshney RK et al (2009b) A QTL analysis of aluminium tolerance in barley, using gene-based markers. Cereal Res Commun 37:531–540
Nguyen BD, Brar DS, Bui BC et al (2003) Identification and mapping of QTL for aluminum tolerance introgressed from the new source, *Oryza rufipogon* Griff., to indica rice (*Oryza sativa* L.). Theor Appl Genet 106:583–593
Nguyen VT, Burrow MD, Nguyen HT et al (2001) Molecular mapping of genes conferring aluminium tolerance in rice (*Oryza sativa* L.). Theor Appl Genet 102:1002–1010
Nguyen VT, Nguyen BD, Sarkarung S et al (2002) Mapping of genes controlling aluminum tolerance in rice: comparison of different genetic backgrounds. Mol Genet Gen 267:772–780
Niedziela A, Bednarek P, Cichy H et al (2012) Aluminum tolerance association mapping in triticale. BMC Genomics 13:67
Ninamango-Cardenas FE, Guimaraes CT, Martins PR et al (2003) Mapping QTLs for aluminum tolerance in maize. Euphytica 130:223–232
Omote H, Hiasa M, Matsumoto T et al (2006) The MATE proteins as fundamental transporters of metabolic and xenobiotic organic cations. Trends Pharmacol Sci 27:587–593
Pan W, Hopkins A, Jackson W (1989) Aluminum inhibition of shoot lateral branches of *Glycine max* and reversal by exogenous cytokinin. Plant Soil 120:1–9
Pellet DM, Papernik LA, Kochian LV (1996) Multiple aluminum-resistance mechanisms in wheat. Roles of root apical phosphate and malate exudation. Plant Physiol 112:591–597
Pereira JF, Zhou GF, Delhaize E et al (2010) Engineering greater aluminium resistance in wheat by over-expressing *TaALMT1*. Ann Bot 106:205–214
Philipp U, Wehling P, Wricke G (1994) A linkage map of rye. Theor Appl Genet 88:243–248

Pineros MA, Cancado GMA, Maron LG et al (2008) Not all *ALMT1*-type transporters mediate aluminum-activated organic acid responses: the case of *ZmALMT1*—an anion-selective transporter. Plant J 53:352–367

Polle E, Konzak CF (1985) A single scale for determining Al tolerance levels in cereals. Agronomy Abstracts 67, ASA, Madison, USA

Polle E, Konzak CF, Kittrick JA (1978) Visual detection of aluminum tolerance levels in wheat by hematoxylin staining of seedling roots. Crop Sci 18:823–827

Raman H, Gustafson P (2010) Molecular breeding for aluminium tolerance in cereals. In: Root Genomics (A Costa de Oliveira and R.K. Varshney eds). Springer, pp 251–288

Raman H, Moroni S, Raman R et al (2001) A genomic region associated with aluminium tolerance in barley. Proceedings of the 10th Australian Barley Technical Symposium. (http://wwwregionalorgau/au/abts/2001/t3/indexhtm#TopOfPage), Canberra

Raman H, Moroni JS, Sato K et al (2002) Identification of AFLP and microsatellite markers linked with an aluminium tolerance gene in barley (*Hordeum vulgare* L.). Theor Appl Genet 105: 458–464

Raman H, Karakousis A, Moroni JS et al (2003) Development and allele diversity of microsatellite markers linked to the aluminium tolerance gene *Alp* in barley. Aust J Agric Res 54:1315–1321

Raman H, Wang JP, Read B et al (2005a) Molecular mapping of resistance to aluminium toxicity in barley. Proceedings of Plant and Animal Genome XIII Conference, San Diego, p 154

Raman H, Zhang K, Cakir M et al (2005b) Molecular characterization and mapping of *ALMT1*, the aluminium-tolerance gene of bread wheat (*Triticum aestivum* L.). Genome 48:781–791

Raman H, Raman R, Wood R, Martin P (2006) Repetitive indel markers within the *ALMT1* gene conditioning aluminium tolerance in wheat (*Triticum aestivum* L.). Mol Breed 18:171–183

Raman H, Ryan PR, Raman R et al (2008) Analysis of *TaALMT1* traces the transmission of aluminum resistance in cultivated common wheat (*Triticum aestivum* L.). Theor Appl Genet 116:343–354

Raman H, Raman R, Luckett D et al (2009) Characterisation of genetic variation for aluminium resistance and polyphenol oxidase activity in genebank accessions of spelt wheat. Breed Sci 59:373–381

Raman H, Stodart B, Ryan PR et al (2010) Genome wide association analyses of common wheat (*Triticum aestivum* L) germplasm identifies multiple loci for aluminium resistance. Genome 53:957–966

Reid DA (1971) Genetic control of reaction to aluminum in winter barley. In: Nilan RA (ed) Proceedings of the 2nd International Barley Genetics Symposium (1969). Washington State University Press, Pullman, WA, pp 409–413

Reid DA, Fleming AL, Foy CD (1971) A method for determining aluminum response of barley in nutrient solution in comparison to response in Al-toxic soil. Agronomy J 63:600–603

Rhue RD, Grogan CO, Stockmeyer EW, Everett HL (1978) Genetic control of aluminium tolerance in corn. Crop Sci 18:1063–1067

Riede CR, Anderson JA (1996) Linkage of RFLP markers to an aluminum tolerance gene in wheat. Crop Sci 36:905–909

Ryan P, Raman H, Gupta S et al (2010) The multiple origins of aluminium resistance in hexaploid wheat include *Aegilops tauschii* and from more recent *cis* mutations to *TaALMT1*. Plant J 64:446–455

Ryan PR, Delhaize E, Randall PJ (1995a) Characterization of Al-stimulated efflux of malate from apices of Al-tolerant wheat roots. Planta 196:103–110

Ryan PR, Delhaize E, Randall PJ (1995b) Malate efflux from root apices and tolerance to aluminium are highly correlated in wheat. Aust J Plant Physiol 22:531–536

Ryan PR, Raman H, Gupta S et al (2009) A second mechanism for aluminum resistance in wheat relies on the constitutive efflux of citrate from roots. Plant Physiol 149:340–351

Saal B, Wricke G (1999) Development of simple sequence repeat markers in rye (*Secale cereale* L.). Genome 42:964–972

Sasaki T, Yamamoto Y, Ezaki B et al (2004) A wheat gene encoding an aluminum-activated malate transporter. Plant J 37:645–653

Sasaki T, Ryan PR, Delhaize E et al (2006) Sequence upstream of the wheat (*Triticum aestivum* L.) *ALMT1* gene and its relationship to aluminum resistance. Plant Cell Physiol 47:1343–1354

Senft P, Wricke G (1996) An extended genetic map of rye (*Secale cereale* L.). Plant Breed 115: 508–510

Smith AV, Thomas DJ, Munro HM, Abecasis GR (2005) Sequence features in regions of weak and strong linkage disequilibrium. Genome Res 15:1519–1534

Somers DJ, Briggs KG, Gustafson JP (1996) Aluminum stress and protein synthesis in near isogenic lines of *Triticum aestivum* differing in aluminum tolerance. Physiol Plant 97:694–700

Soto-Cerda BJ, Cloutier S (2012) Association Mapping in Plant Genomes. Genetic Diversity in Plants. Prof Mahmut Caliskan (Ed), ISBN: 978–953-51–0185-7, InTec. http://wwwintechopencom/books/genetic-diversity-in-plants/association-mapping-in-plant-genomes

Stich B, Melchinger AE, Piepho HP et al (2007) Potential causes of linkage disequilibrium in a European maize breeding program investigated with computer simulations. Theor Appl Genet 115:529–536

Stich B, Mohring J, Piepho HP et al (2008) Comparison of mixed-model approaches for association mapping. Genetics 178:1745–1754

Stodart BJ, Raman H, Coombes N, Mackay M (2007) Evaluating landraces of bread wheat *Triticum aestivum* L. for tolerance to aluminium under low pH conditions. Genet Res Crop Evol 54: 759–766

Tang Y, Sorrells ME, Kochian LV, Garvin DF (2000) Identification of RFLP markers linked to the barley aluminum tolerance gene *Alp*. Crop Sci 40:778–782

Uhde-Stone C, Liu J, Zinn KE et al (2005) Transgenic proteiod roots of white lupin: a vehicle for characterisation and silencing root genes involved in adapation to P stress. Plant J 44:840–853

von Uexkull HR, Mutert E (1995) Global extent, development and economic impact of acid soils. Plant Soil 171:1–15

Wang JP, Raman H, Read B et al (2004) Comparison of root staining and root elongation in predicting aluminium tolerance using SSR markers in barley. Proceeding of 4th International Crop ScienceCongress. http://wwwcropscienceorgau/icsc2004/poster/3/6/4/1168_wangjhtm, Brisbane

Wang J, Raman H, Zhang G-P et al (2006a) Aluminium tolerance in barley (*Hordeum vulgare* L.): physiological mechanisms and screening methods. J Zhejiang Univ Sci 7:769–787

Wang JP, Raman H, Read B et al (2006b) Validation of an *Alt* locus for aluminium tolerance scored with eriochrome cyanine R staining method in barley cultivar Honen (*Hordeum vulgare* L.). Aust J Agric Res 57:113–118

Wang JP, Raman H, Zhou MX et al (2007) High-resolution mapping of the *Alp* locus and identification of a candidate gene *HvMATE* controlling aluminium tolerance in barley (*Hordeum vulgare* L.). Theor Appl Genet 115:265–276

Wenzl P, Li H, Carling J et al (2006) A high-density consensus map of barley linking DArT markers to SSR, RFLP and STS loci and agricultural traits. BMC Genomics 7:206

Wight CP, Kibite S, Tinker NA, Molnar SJ (2006) Identification of molecular markers for aluminium tolerance in diploid oat through comparative mapping and QTL analysis. Theor Appl Genet 112:222–231

Wu P, Liao CY, Hu B, Yi et al (2000) QTLs and epistasis for aluminum tolerance in rice (*Oryza sativa* L.) at different seedling stages. Theor Appl Genet 100:1295–1303

Xia J, Yamaji N, Kasai T, Ma JF (2010) Plasma membrane-localized transporter for aluminum in rice. Proc Natl Acad Sci USA 107:18381–18385

Xue Y, Jiang L, Su N et al (2007) The genetic basic and fine-mapping of a stable quantitative-trait loci for aluminium tolerance in rice. Planta 227:255–262

Yamaguchi M, Sasaki T, Sivaguru M et al (2005) Evidence for the plasma membrane localization of Al-activated malate transporter (*ALMT1*). Plant Cell Physiol 46:812–816

Yamajia N, Huang CF, Nagao S et al (2009) A zinc finger transcription factor ART1 regulates multiple genes implicated in aluminum tolerance in rice. Plant Cell 21:3339–3349

Yan J, Shah T, Warburton ML et al (2009) Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. PLoS ONE 4:e8451

Yokosho K, Yamaji N, Ma JF (2010) Isolation and characterisation of two *MATE* genes in rye. Funct Plant Biol 37:296–303

Yokosho K, Yamaji N, Ma JF (2011) An Al-inducible *MATE* gene is involved in external detoxification of Al in rice. Plant J 68:1061–1069

Yu J, Pressoir G, Briggs WH et al (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203–208

Yu JM, Holland JB, McMullen MD, Buckler ES (2008) Genetic design and statistical power of nested association mapping in maize. Genetics 178:539–551

Zhang X, Jessop RS (1998) Analysis of genetic variability of aluminium tolerance response in triticale. Euphytica 102:177–182

Zhou LL, Bai G-H, Ma HX, Carver BF (2007a) Quantitative trait loci for aluminum resistance in wheat. Mol Breed 19:153–161

Zhou LL, Bai GH, Carver BF, Zhang DD (2007b) Identification of new sources of aluminium resistance in wheat. Plant Soil 297:105–118

Chapter 9
Maintaining Food Value of Wild Rice (*Zizania palustris* L.) Using Comparative Genomics

Alexander L. Kahler, Anthony J. Kern, Raymond A. Porter and Ronald L. Phillips

Abstract Wild rice (*Zizania palustris*) is a naturally-occurring, aquatic plant species that is important to wildlife, aquatic biological systems and humans. Populations of *Z. palustris* across the geographic range continue to decline in their natural habitat. In some cases, natural populations are being lost. Sometimes referred to as American wild rice, it is genetically similar to *Oryza sativa*, or cultivated Asian rice. This similarity coupled with modern advances in rice genomics have allowed for comparative genetics and genomics studies between *Z. palustris* and *O. sativa* and genetic diversity studies, which have been useful for characterizing the available natural genetic resources. The ongoing wild rice breeding program has been successful in supporting and expanding the cultivated wild rice industry. The incorporation of modern molecular genetics approaches to selection have improved the ability to breed for cultivated wild rice varieties that are more resistant to seed shattering, which has contributed to increased grain production.

9.1 Introduction

Wild rice (*Zizania palustris*) is an annual, aquatic grass that belongs to the Poaceae (Watson and Dallwitz 1992). It most often occurs in shallow-water lakes, small rivers, and coastal regions in the north-central United States and southern Canada (Oelke 2007). The earliest recorded description of wild rice was by a British explorer Peter Bond in 1775 (Fannucchi et al. 1986). The plant provides habitat for numerous species of birds, mammals, juvenile fish, and invertebrates. Wild rice is involved in aquatic nutrient cycling and aids in stabilizing sediments in coastal riverine wetlands (Meeker 1996).

In optimal habitats, wild rice is a prolific producer of edible grain, which has been a dietary staple of indigenous peoples for centuries (Johnson 1969) and serves as an important food source for both migratory and resident waterfowl (Minnesota Department of Natural Resources 2008). Wild rice is culturally important to several

A. L. Kahler (✉) · A. J. Kern · R. A. Porter · R. L. Phillips
Department of Agronomy and Plant Genetics, University of Minnesota, Minneapolis, MN, USA
e-mail: kahl0041@umn.edu

R. Tuberosa et al. (eds.), *Genomics of Plant Genetic Resources,*
DOI 10.1007/978-94-007-7575-6_9,

Native American tribes, especially the Ojibwe, who consider the plant as sacred and use it for food, medicine and in ceremony (Oelke and McClellan 1992).

Historically, the wild rice grain has been economically important. Still today, the sale of hand-harvested wild rice continues to be of considerable economic importance to these tribes and non-natives alike (Vennum 2000; Wisconsin Department of Natural Resources 2003). Tribal rights to hand-harvest wild rice on- and off-reservation are legally protected in most of northern Michigan, northern Wisconsin, and northeastern Minnesota (Great Lakes Indian Fish and Wildlife Commission 2006).

Zizania palustris is the wild rice species most commonly used for food. Its flower is imperfect and exists in a branched panicle structure that is monoecious (Goldman 1990). One wild rice plant can produce several tillers each containing a panicle. The pistillate florets occur along the top one-third of the panicle and above the staminate florets. This floret arrangement favors cross-pollination as the stigmas become receptive as the pistillate florets emerge prior to the staminate florets becoming mature. In addition, pollen is released downward from the anther and away from the pistillate florets or is blown away from the source plant by the wind.

Within the last century, wild rice populations across the natural range have declined in fitness and, in some cases, have been extirpated (Minnesota Department of Natural Resources 2008). The major causes of this decline include changes to natural hydrologic regimes in wild rice waters and invasions of wild rice habitat by exotic species (Fannucchi et al. 1986).

Within numerous locations along coastal areas of the Great Lakes, wild rice populations are imperiled or extinct where they were once abundant. One such area is the St. Louis River estuary, where historically-large (> 100 ha) populations have become highly fragmented and greatly reduced in size. Contemporary populations of wild rice within most of the St. Louis River estuary sites are generally characterized as being small (often < 1 ha), highly fragmented and usually confined to the heads of protected bays. However, wild rice habitat damage in some of these areas has been mitigated and considerable interest now exists in restoring wild rice to additional historic areas (Lu et al. 2005). Such effort is ongoing on certain inland lakes where wild rice has been extirpated (Weaver et al. 2005; Wisconsin Department of Natural Resources 2006). There is growing consensus within the restoration community that understanding the genetic "structure" of remnant native populations is an important component in guiding restoration protocols.

Such understanding may increase restoration success through introducing only local and/or adapted genotypes (Edmands and Timmerman 2003; Hufford and Mazer 2003; Laikre et al. 2010).

9.2 Wild Rice Genetics

Wild rice has 15 pairs of chromosomes ($2n = 2x = 30$) that pair at meiosis as either eight rod and seven ring bivalents or seven rod and eight ring bivalents (Grombacher 1997). This is three chromosome pairs more than its cultivated rice relative, *Oryza*

sativa. Kennard et al. (2000) hypothesized that wild rice chromosomes one, four and nine may contain duplicated segments of the homologous rice chromosomes. The total content of genomic DNA in wild rice is approximately 860 Mb (Kennard et al. 2000); this is about twice the amount of DNA contained in the *Oryza sativa* genome (430 Mb) (Arumunagathan and Earle 1991). This observation along with the three chromosome duplications may be evidence of a genome-wide duplication event that led to autotetraploidization of *Oryza* followed by loss of homologous chromosomes (Kennard et al. 2000). *Z. palustris* possibly retained three of the duplicated chromosomes as the genome stabilized.

Work by Hass et al. (2003) indicated that part of the large difference in genome size between rice and wild rice can be attributed to gene duplication. For example, rice is known to have two copies of the *Adh* gene (*Adh1* and *Adh2*). DNA sequence comparison showed that both copies of *Adh* are duplicated in wild rice (*Adh1a*, *Adh1b*, *Adh2a* and *Adh2b*).

Based on comparison to the *Oryza sativa* model genome (Goff 1999), widespread colinearity among grass genomes has been well documented (Bennetzen and Freeling 1993; Paterson et al. 1995; Gale and Devos 1998). *Zizania palustris* and *O. sativa* are taxonomically grouped into the subfamily *Oryzoideae* and the tribe *Oryzeae* (Duvall et al. 1993). Significant genome colinearity has been observed between wild rice and *O. sativa* (Kennard et al. 2000; Hass et al. 2003). Early genome comparisons based on DNA hybridization between wild rice and other grass genomes verified that wild rice is highly homologous to *O. sativa* (Kennard et al. 2000).

Kennard et al. (2000) constructed an RFLP (restriction fragment length polymorphism)-based linkage map of wild rice consisting of 121 loci; all of the RFLP markers had been mapped previously in *Oryza sativa* and many markers were designed from cDNA representing gene sequences. These markers allowed for comparative mapping and identification of syntenous genome regions between *Zizania palustris* and *O. sativa*. While not perfect, the mapped markers were 82 % colinear between wild rice and rice. As has been observed in other species (Odland et al. 2006), some genes were not represented in both syntenic regions of rice and wild rice.

A significant correlation between a molecular marker genotype and the phenotype of a quantitative trait in the same population identifies a region on a chromosome as a quantitative trait locus (QTL). The wild rice linkage mapping data were used to elucidate QTLs for agronomic traits (Kennard et al. 2002). Of particular interest and importance was the discovery of four QTLs that explained varying amounts of seed shattering variability in the mapping population (Kennard et al. 2002). Two of the wild rice shattering QTLs (marked by UMC305 and CDO244) mapped to the general colinear regions of the rice genome containing seed shattering QTLs. However, neither of the wild rice seed shattering QTLs mapped directly on top of the known rice shattering QTLs (Kennard et al. 2002). Three of the wild rice QTLs were confirmed in the F_3 generation (Imle 2001). Observed segregation of the three confirmed QTLs in the F_2 population fit a three-locus model for explaining the observed variability (Kennard et al. 2002). The major QTL accounted for approximately 38 % of the shattering variation in the F_2 mapping population.

The RFLP-based marker for the major shattering QTL has not been adopted by the wild rice breeding program. The process for using RFLP markers is too slow to make early selections in the nursery before the plant enters anthesis. Further, the large amount of genomic DNA needed for RFLP analysis requires that most of the wild rice plant be stripped of its leaves leading to poor growth or even plant death. Additionally, a limitation of QTL data is that the alleles present and segregating in one population may not be present and/or segregating in another population. This is especially problematic when genotypes that work well for mapping are not favorable for breeding or germplasm improvement due to having poor agronomic traits.

Phillips et al. (2005) emphasized that the rice genome is an ideal model genome for plant genomics studies and that, in particular, the available rice genome DNA sequence and other molecular tool sets provide a basis for comparative genetics and genomics projects. A set of 60 *Oryza sativa* simple sequence repeat (SSR) markers representing all 12 rice chromosomes was evaluated by Gao et al. (2005) for their usefulness among wild relatives with varying genome constitutions. The focus of the wild rice molecular genetics program turned to the development of PCR (polymerase chain reaction)-based markers. Six-hundred SSR markers developed in *O. sativa* (Temnykh et al. 2000; McCouch et al. 2002) were tested using *Zizania palustris* DNA. About 60 % of the markers successfully amplified wild rice DNA and approximately 50 % of those were polymorphic in an F_2 mapping population. This led to the addition of 64 rice SSR markers to the *Z. palustris* RFLP linkage map (Kahler et al. 2006; Kahler 2007).

SSR markers were chosen from the *Oryza sativa* chromosome region that was syntenic to the RFLP-based *Zizania palustris* shattering QTL. These markers were tested on a *Z. palustris* population that segregated for seed shattering. This led to the identification of one polymorphic marker (RM106) linked to the non-shattering phenotype (Kahler 2007). The use of RM106 will be discussed in a subsequent section of this chapter.

9.3 Wild Rice Genetic Diversity

Despite its economic, social, and ecological significance, little is known about wild rice genetic diversity or how the plant's genetic structure and patterns of reproduction affect plant performance. Natural wild rice populations occur in various sizes. Two hypotheses have been set forth in studies of natural wild rice populations. The first hypothesis is that wild rice requires significant stability in water levels to thrive (Minnesota Department of Natural Resources 2008). This hypothesis may explain why, on larger lakes, small wild rice populations may be found in shallow water close to undisturbed shorelines. Land development and water recreation on wild rice-bearing lakes may sufficiently disturb these shallow water areas leading to reduction in population size or even complete loss of wild rice populations. The second hypothesis is that small, isolated wild rice populations are more likely to inbreed and become highly adapted to their specific mini-environment. Such inbred populations also may lose vigor.

This may lead to two condtions: (1) inhibition of migration from the mini-environment and (2) out-breeding effects from migration between the mini-environments (A. Kern, personal communication).

One study used isozyme markers to study genetic variability in several natural wild rice populations in northern Wisconsin (Lu et al. 2005). It was reported that wild rice genetic differentiation was high and gene flow was low between different wild rice populations. The characterized populations were generally isolated from each other and these results may indicate that wild rice habitat needs are specific. If this is, indeed the case, current observations of wild rice habitat areas raise serious concern about genetic erosion caused by declining habitats and reductions in wild rice population size (Waller et al. 2000).

In addition to the set of *Oryza sativa* SSR markers discussed previously, a set of highly polymorphic wild rice SSR markers was produced in *Zizania texana* (Richards et al. 2004). This marker set was tested using *Z. palustris* DNA and new DNA primers were designed from the PCR amplicon sequences (Kern 2009; Kahler 2010). These markers were unlinked based on linkage mapping data, are highly reproducible and have been optimized for high-throughput, population-level analysis (Kahler 2010). The marker set has exhibited high power of discrimination among wild rice populations with many marker PIC (polymorphism information content) values > 0.65 among populations tested (Kahler et al. 2011; Kern et al. 2011). This set of SSR markers was used to measure the genetic diversity among natural stands of wild rice in Minnesota and populations from the University of Minnesota wild rice breeding program (Kahler et al. 2012). Similar to the findings of Lu et al. (2005), genetic diversity, measured by F_{ST} was high among natural wild rice populations (Figs. 9.1a, 9.1b and Table 9.1), suggesting that many wild rice populations are genetically distinct from each other. In marked contrast, most of the breeding lines show very low levels of genetic differentiation (Fig. 9.2 and Table 9.1), which may be due to shared ancestry in the breeding program or pollen flow between populations.

Unexpectedly, the microsatellite analyses indicated high levels of inbreeding (f) in some natural populations of wild rice (Table 9.2). As previously discussed, wild rice is predominantly accepted to be a wind-pollinated, out-crossing species. However, in most habitats, wild rice typically produces multiple tillers and it is common for several inflorescences to develop on multiple culms of the same plant. These secondary stems often mature one-to-three weeks later than the primary stem making it possible for self-fertilization to occur between inflorescences on different culms. Indeed, this asynchronous tiller development is commonly used in the University of Minnesota's wild rice breeding program to generate self-fertilized individuals (Porter, personal communication). Although no study to date has investigated the natural patterns of between-inflorescence self-pollination in American wild rice, previous studies have shown this type of self-pollination to play a significant role in the reproduction of the closely-related, critically-endangered Texas wild rice, *Zizania texana*. Self-pollination between related inflorescences may be just as likely to occur as cross-pollination with a genetically-distinct individual (Power and Oxley 2004), particularly in stands where plants are present in low densities, resulting in greater numbers of tillers per plant with fewer neighbors in close proximity. Plants

	BRL	D	F	L	LAL	LOL	LRL	M
BRL	0							
D	0.25	0						
F	0.17	0.21	0					
L	0.2	0.23	0.04	0				
LAL	0.18	0.19	0.15	0.17	0			
LOL	0.23	0.25	0.2	0.2	0.1	0		
LRL	0.17	0.25	0.18	0.17	0.15	0.12	0	
M	0.17	0.22	0.02	0.02	0.15	0.21	0.17	0
N	0.21	0.25	0.06	0.09	0.17	0.25	0.2	0.07
P	0.21	0.24	0.01	0.02	0.17	0.23	0.2	0.02
PBMC	0.14	0.21	0.16	0.17	0.15	0.21	0.2	0.16
RLNWR-E	0.18	0.23	0.06	0.06	0.15	0.21	0.18	0.05
RLNWR-R	0.16	0.17	0.05	0.04	0.09	0.18	0.17	0.04
RLNWR-W	0.18	0.21	0.08	0.08	0.15	0.2	0.16	0.07
TNWR-DL	0.17	0.19	0.07	0.06	0.14	0.16	0.14	0.07
TNWR-RL	0.15	0.17	0.07	0.08	0.11	0.14	0.12	0.09
TNWR-TL	0.18	0.21	0.09	0.08	0.15	0.18	0.14	0.1

a

	N	P	PBMC	RLNWR-E	RLNWR-R	RLNWR-W	TNWR-DL	TNWR-RL	TNWR-TL
BRL									
D									
F									
L									
LAL									
LOL									
LRL									
M									
N	0								
P	0.07	0							
PBMC	0.21	0.19	0						
RLNWR-E	0.09	0.07	0.17	0					
RLNWR-R	0.1	0.05	0.1	0.05	0				
RLNWR-W	0.12	0.09	0.16	0	0.05	0			
TNWR-DL	0.14	0.08	0.14	0.08	0.07	0.08	0		
TNWR-RL	0.15	0.11	0.12	0.1	0.06	0.09	0.04	0	
TNWR-TL	0.17	0.13	0.16	0.09	0.09	0.07	0.01	0.04	0

b

Fig. 9.1 **a** F_{ST} matrix for eight natural and breeding populations of wild rice calculated using SSR marker allele frequency data. **b** F_{ST} matrix for nine natural and breeding populations of wild rice calculated using SSR marker allele frequency data

Table 9.1 Description of natural and wild rice breeding populations referenced in Figs. 9.1 and 9.2

Population ID	Population name	Population description
D	Dawn SR	Selected from cv. K2 for nonshattering using an SSR marker linked to the major shattering QTL, and for early maturity. Released in 2008
F	FY-C16	Selected for 16 cycles from cv. Franklin (which was derived from K2) for yield and shattering resistance
L	PLar-C13	Selected for pistillate (gynomonoecious) panicle, and resistance to shattering, lodging, and foliar disease, for 13 cycles. Originally from Lake LaRonge, SK, Canada
M	PM3E-C17	Selected for 17 cycles from Pistillate M3 germplasm for pistillate panicle, earliness, and resistance to shattering, lodging, and foliar disease. Selected out of cv. M3, developed by Manomin Development Corp
N	NeBR-C8a	Selected for 8 cycles from Netum for resistance to *Bipolaris* spp., shattering, and lodging. Derived from cv. Netum, released in 1978 from several lake sources
P	PBM-C15	Selected for 14 cycles from cv. Petrowske Bottlebrush (which was derived from K2) for yield, "bottlebrush" panicles, long seeds, and resistance to foliar disease and stem rot
BRL	Big Rice Lake	Natural population—St. Louis Co., MN
LRL	Little Rice Lake	Natural population—St. Louis Co., MN
PBMC	Pokegama Bay Main Channel	Natural population—St. Louis Co., MN
LOL	Low's Lake	Natural population—Crow Wing, Co., MN
LAL	Laura Lake	Natural population—Cass Co., MN
RLNWR	Rice Lake National Wildlife Refuge	Natural population—Aitkin Co., MN
TNWR	Tamarack National Wildlife Refuge	Natural population—Becker Co., MN

resulting from seeds produced by self-pollination are generally weak and survive at low rates both in nature and in the cultivated paddy. The plants are typically shorter than normal and produce fewer seeds. Further, the seeds that are produced often have poor seedling vigor when germinated. These traits are typical of inbreeding depression that is thought to be due to the accumulation of deleterious, homozygous recessive genotypes at loci that are normally heterozygous (Fehr 1987).

9.4 Wild Rice Breeding

Wild rice is categorized and labeled by law as either "natural lake or river" or "cultivated" in Minnesota (Minnesota statute 30.49, State of Minnesota 2006). Many Native American tribal reservations in northern Minnesota market hand-harvested, lake wild rice as a Minnesota specialty food. Minnesota residents may purchase a wild rice harvesting permit from the Department of Natural Resources which allows them to hand-harvest wild rice from public (non tribal) waters. Roadside stands selling hand-harvested wild rice are common in northern Minnesota in the fall.

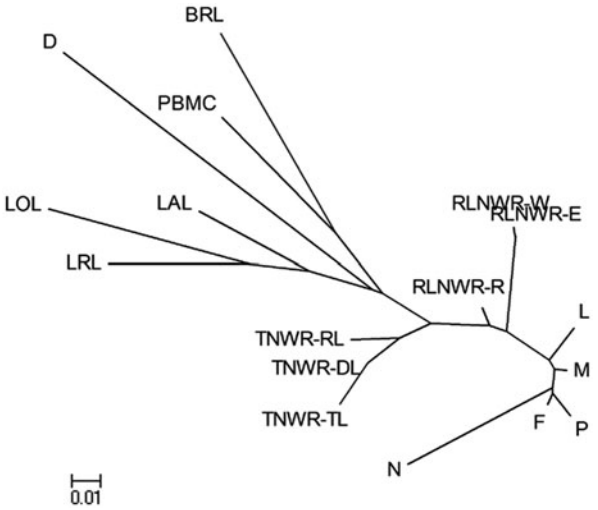

Fig. 9.2 Unrooted radiation dendogram of natural and breeding populations of wild rice based on F_{ST} values from Fig. 9.1 above

About 1853 it was suggested that wild rice could be cultivated as a crop (Aiken 1988). The first successful paddy of wild rice was cultivated and harvested in Northern Minnesota in 1950 (Oelke 2007). Around 1962, the Uncle Ben's company contracted with Minnesota wild rice growers to purchase and market wild rice.

By 1972, several farmers were cultivating wild rice in paddies but they had been largely unsuccessful in improving the natural collections for cultivation. The group of wild rice growers asked the Department of Agronomy and Plant Genetics at the

Table 9.2 Summary population statistics calculated using SSR marker genotype and allele frequency data

Population	# of Genotypes	# of Alleles	Gene diversity	Heterozygosity	PIC	F
BRL	15	9	0.72	0.71	0.68	0.04
D	18	14	0.71	0.49	0.69	0.32
F	18	12	0.77	0.65	0.75	0.17
L	17	9	0.74	0.65	0.71	0.13
LAL	22	17	0.80	0.69	0.77	0.16
LOL	14	10	0.73	0.84	0.69	−0.14
LRL	15	11	0.71	0.67	0.67	0.08
M	18	10	0.75	0.67	0.72	0.12
N	17	11	0.67	0.58	0.65	0.15
P	15	9	0.71	0.63	0.68	0.12
PBMC	24	15	0.78	0.67	0.77	0.16
RLNWR-E	16	12	0.73	0.54	0.70	0.27
RLNWR-R	25	16	0.82	0.67	0.80	0.19
RLNWR-W	17	12	0.74	0.49	0.71	0.36
TNWR-DL	17	12	0.79	0.58	0.76	0.27
TNWR-RL	28	16	0.86	0.75	0.85	0.13
TNWR-TL	20	12	0.75	0.60	0.72	0.22

University of Minnesota to help with improving wild rice as a crop. A relationship was formed which led to the organization of a formal, public wild rice breeding program, that continues today after 40 years.

Successful cultivation relies first on the ability to effectively flood and drain the paddy to maintain the correct water level. Wild rice paddies are generally constructed near natural water sources (*i.e.,* rivers and lakes), from which water is pumped. Prior to flooding, the paddy is often leveled using precise, laser-guided equipment for efficient water drainage, which is facilitated via drainage-tile lines. Wild rice production requires nitrogen fertilization, which is applied as ammonia prior to planting and then via aerial top-dressing during flowering. Further, reducing yield loss due to weeds, insect pests and foliar disease also requires relatively expensive aerial application of pesticides and herbicides.

Wild rice is an important cultivated niche crop in the United States. The majority of its production occurs in Minnesota and California. The northern California counties where wild rice is produced do not normally have high winds or thunderstorms as is common in the Midwest. Such storms cause mature seeds to shatter prior to harvest.

In typical years, 15–20 million pounds of grain are produced between the two states on about 30,000 acres. In commercial production schemes, the plant is typically grown in diked, flooded paddies in a manner very similar to most varieties of Asian white rice (*Oryza sativa*). A significant portion of the wild rice breeding program at the University of Minnesota utilizes conventional approaches to plant improvement by selecting for specific phenotypic traits that would likely significantly improve yield and production efficiency in the cultivated varieties, as well as the continued development of important germplasm resources (Grombacher et al. 1997).

Progress in crop improvement has been successful using the phenotypic mass recurrent selection method of plant breeding and has led to releases of improved wild rice varieties (Kennard et al. 2000). Phenotypic recurrent selection is carried out through the selection and bulking of seed from the phenotypically elite individuals in one breeding generation for advancement to the next generation of selection (Fehr 1987). The goal of this selection method is to increase the frequency of favorable alleles in the breeding population in successive generations. The advantage of this method to wild rice breeding is that it allows for intermating of individuals which reduces the effect of inbreeding depression in this out-crossing species. The disadvantage is that it takes a relatively long time to produce significant amounts of seed if the selection threshold is maintained at a high level ($< 10\,\%$). Conversely, it takes a long time to sufficiently increase the frequency of favorable alleles in the population if the selection threshold is kept low to facilitate greater amounts of seed.

Compared to other crop species, wild rice has had, to date, a short timeframe of commercial production and organized breeding effort. This fact coupled with certain aspects inherent to the plant's reproductive biology, have resulted in the currently available cultivars and varieties being extremely heterogeneous. They can best be described as semi-wild and in the very earliest stages of domestication. They have not yet been fixed for several "domestication traits" including: nonshattering seed (a natural seed dispersal mechanism of plants), reduced plant height, uniform seed maturity and seed nondormancy (De Wet and Oelke 1978; Paterson et al. 1995). Seed

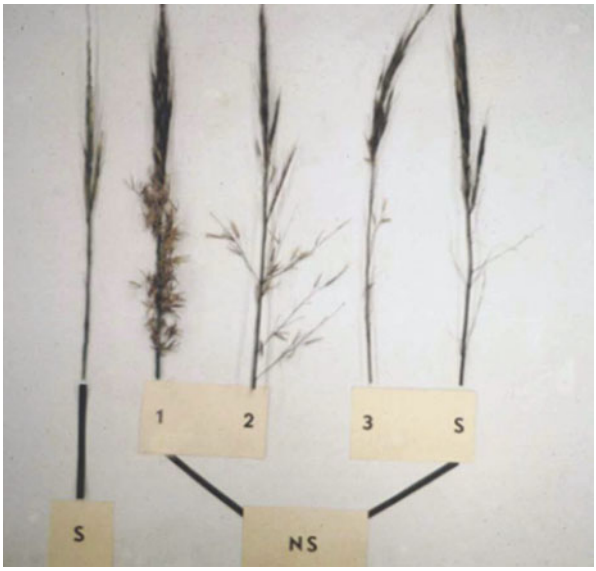

Fig. 9.3 Illustration of a shattering wild rice panicle (S on far *left*) and other wild rice panicles exhibiting quantitative degrees of nonshattering based on male floret and seed retention (1-S)

shattering is, perhaps the most important wild rice production trait as a single windy day or poorly-timed thunderstorm during the harvest season can shatter over 50 % causing huge reductions in harvestable yields (Oelke, personal communication).

Although selection for non-shattering varieties has been an ongoing focus of the breeding program since the early 1970s, the problem is exacerbated by a necessary production technique in northern climates. Wild rice seed requires several months of cold stratification to break seed dormancy (Simpson 1966; Kovach and Bradford 1992). It is economically most feasible for a producer to allow a production paddy to re-seed itself from seeds that shatter prior to or during fall harvest, which obviates the need for re-seeding every year. This technique greatly reduces seed costs, but invariably imparts a strong selection pressure that favors the shattering phenotype in subsequent crops, resulting in inconsistent yields due to shattering losses.

To reduce the production loss, most paddies are rotated out of wild rice production once every two-to-four years (to destroy the seed bank) with soybean, or potatoes, or other crops, and then replanted with wild rice varieties selected for increased resistance to seed shattering. Cultivated wild rice production would greatly benefit from the ability to effectively select for true-breeding, non-shattering genotypes, which would not only increase yield for producers, but would also directly benefit the breeding program by allowing for more efficient gains from selection for other important agronomic traits if breeding lines were "fixed" for the nonshattering trait (Kahler 2007).

Seed shattering is a quantitative trait in wild rice; there are varying degrees to which a wild rice panicle shatters its seeds (Fig. 9.3). A plant that shatters all of its seeds is scored as shattering. A plant that shatters most of its seeds, but not all, is scored as shattering resistant. A plant that shatters almost none of its seeds is scored as nonshattering. Kennard et al. (2002) reported a 90 % correlation between the retention of wild rice staminate florets and retention of seeds.

9.5 Case Study: Breeding Wild Rice Using Marker-Assisted Breeding for Nonshattering

The history of wild rice domestication began in the second half of the 20th Century. Although the first paddy to produce a successful crop was seeded in 1951 (Oelke 2007), it made use of wild-type seed collected from a natural stand. It wasn't until 1968 that the first variety with the "nonshattering"[1] trait was developed. It was named 'Johnson' after wild rice grower Algot Johnson, in whose paddy its nonshattering progenitors were first found. 'K2' was developed by growers Franklin and Harold Kosbau, who made it available to wild rice growers in 1972 as a higher-yielding variety with mostly nonshattering plant types. The variety had never been fixed for the nonshattering phenotype, but because of its superiority to Johnson, it eventually became the dominant variety grown in Minnesota.

Because shattering and strong seed dormancy have characterized wild rice cultivars since the beginning of the cultivation of wild rice, producers found it difficult to change varieties once a paddy had been in production for at least a year; viable seeds became abundant in the soil. Continuous production for many years was easier than fallowing the crop, especially since two or more years of fallowing would be needed to minimize the impact of surviving shattered, dormant seed on a re-seeded paddy. Since the release of K2, growers generally practiced continuous production of this variety, occasionally taking a paddy out of production (fallowing) or planting other crops, and sometimes seeding newly developed paddies with K2 seeds. Because of this practice growers and researchers observed that shattering increased over time.

To estimate the degree and rate of reversion to shattering, an experiment was conducted in 1990 using several sources of K2. These seeds had been collected the previous year from growers' paddies that had been in continuous production for 2, 4, 10, and 15 years. The results showed that the percentage of wild-type shattering plants went up as time under continuous production increased.

This percentage for the four sources was 9, 33, 53, and 68 % respectively (Porter unpublished). The paddy in production for only 2 years had been seeded from a seed paddy that, in turn, was newly seeded, etc. for several years. This percentage was probably close to what the originally released version of K2 exhibited. The source of the 15-years-of-continuous-production seeds used in this trial was a paddy that remained in continuous productions for several more years afterward.

When the same paddy was observed by the wild rice breeder in 1997, this population of K2 that had "reverted" to the wild shattering type also exhibited a very early flowering phenotype. This was thought to have come about due to the unintended selection pressure by the grower's practice of having to harvest the crop at an earlier point in time than in the previous year; seeds that matured and shattered to the ground prior to harvest (putatively earlier maturing) were allowed to produce the next year's

[1] "Nonshattering" is used to denote the phenotype in which male florets and seeds are retained and not readily lost (as in the wild type), regardless of the fact that some seeds may still shatter when they reach maturity.

Table 9.3 Timeline for the development of the cultivar Dawn SR

Year-season	Activity	Result
1990-field	Sources of the variety 'K2' grown under continuous cultivation for 2, 4, 10, and 15 years were compared; a lake population that had been grown continuously for 20 years was also included, from the same farm as the 15-year K2. Proximity of lake population paddy to K2 paddy on this farm was unknown. The 15-year K2 was designated K2-Fig.	68 % of K2-Fig plants were wild-type (shattering) compared to 9 % shattering for K2 (2 years)
1997-field	K2 seeds collected from the same location where K2-Fig had been obtained for the previous experiment. By this time, production had been continuous for 23 years	K2-Fig
1998-field	K2-Fig grown in isolation (8″ pots in stock tank) to increase seed	K2-Fig (405 g)
1999-field	K2-Fig grown in paddy 6 in a population block of 80′ × 45′; 63 g of seeds from non-shattering plants were collected (estimated 10–20 % of plants selected)	K2EF-C1 (63 g)
2000-field	K2EF-C1 planted at end of paddy 1B (approx 500 sq ft); early flowering non-shattering plants phenotypically selected (∼30 % of 400–800 plants)	K2EF-C2 (1435 g)
2001-field	K2EF-C2 planted at end of paddy 2 (approx 3000 sq ft); early flowering non-shattering plants phenotypically selected (∼ 1–5 % of 2000–4000 plants)	K2EF-C3 (515 g)
2002-field	24 rows (10′ in length) of K2EF-C3 planted in nursery selfing block; out of 137 plants phenotyped, 4 were shattering; tissue was harvested to identify a marker co-segregating with the shattering phenotype; selfed seeds were harvested from 39 of these	39 S1 lines of K2EF-C3
2003-grnhse	Selfed seeds were germinated; 29 of the lines were planted in the greenhouse and phenotyped; of these one family from a selfed shattering parent (family 205) had two surviving progeny, both shattering; leaf tissue was sampled for DNA to confirm the association with RM106	
2003-field	72 pots of K2EF-C3 remnant seed (open-pollinated) from 2001 were planted in isolation (lined boxes outside of seed room); plants were allowed to open-pollinate	K2EF-C3 (1 qt)

Table 9.3 (countinued)

Year-season	Activity	Result
2004-field	K2EF-C3 seeds from 2003 were planted in paddy 169B; tissue was collected from 370 plants, 151 of which produced selfed seeds; 53 of these were selected based on the RM106 marker, early flowering, and plant/panicle robustness	K2EF-C4 (53 S1 families)
2005-grnhse	A total of 824 plants from 51 of the 53 selected families were grown in pots in the greenhouse; tissue was sampled to genotype the plants for the RM106 marker; none of the progeny showed a shattering phenotype, but families that had at least one plant with a shattering RM106 allele were removed. Only 124 plants from nine non-segregating S1 families were allowed to inter-pollinate	K2EF-C5 (9 intermated S1 families)
2006-grnhse	A total of 592 plants from 77 sub-families of the 9 selected families were grown in pots in the greenhouse; tissue was sampled as before to screen using the RM106 marker. Only 28 plants had the marker allele indicating shattering, and these were removed from the population prior to flowering. Open-pollinated seeds from the remaining plants were collected and bulked.	K2EF-C6 (3 qts)
2006-field	Seeds were planted in a 0.48 acre isolation paddy on-station. RM106 marker was applied to 152 randomly collected plants to detect presence of shattering marker: none found. No shattering phenotypes observed in field increase	K2EF-C6 (210 lb)
2007-field	Since larger on-farm increase paddies were not available, a second on-station increase was planted in isolation (1/2 acre), using some of the first increase seed	K2EF-C6 (220 lb)
2008-field	A single 14 acre on-farm seed increase paddy was planted; in a transect survey across the increase paddy, shattering plants were observed at an estimated frequency of 0.0001. U of M Crop Variety Review Committee approved release of the variety as the first wildrice variety developed using marker-assisted selection; 23,000 lb. of seeds were harvested and distributed to growers	Dawn SR

crop. If this early version of K2 could be fixed for nonshattering, researchers theorized that growers armed with such a variety could esablish it in paddies that had a prevalence of wild-type shattered seeds in the soil, harvest the early variety before shattering could occur, and so clean up their old paddies of the shattering types.

Seeds were collected from this K2 population in 1997 in hopes of re-selecting it for nonshattering types and re-releasing it as an early-maturing nonshattering variety. After an isolated increase in 1998 of the small quantity of seeds collected, phenotypic selection for earliness and nonshattering phenotypes was initiated in 1999 (Table 9.3). Phenotypic recurrent selection for earliness and nonshattering was repeated in this way for a total of three cycles (K2EF-C1 through K2EF-C3).

At that point, an SSR marker had been identified that was potentially associated with one of the three previously identified shattering QTLs. The association of this marker, RM106, with the shattering trait was confirmed in the S_1 progeny of K2EF-C3. RM106 was then applied to K2EF-C3 grown from remnant seeds. Selfed seeds of agronomically superior plants that indicated a nonshattering genotype were kept for the next cycle. Fifty-one K2EF-C4 S_1 families were grown in the greenhouse and screened using RM106. Of the 824 progeny that were grown out, plants from nine S_1 families that had no shattering genotypes were isolated and intermated by open pollination to produce K2EF-C5 seeds. These were planted again in the greenhouse the following year and screened again with the same marker. Of 592 plants screened, only 28 had the RM106 shattering allele, and these were removed. The remaining plants were open pollinated, and seeds were kept as K2EF-C6. After several increases in isolation, it was released as 'Dawn SR.'

This varietal release represented the successful application of marker-assisted selection to wild rice. Similar use of genomics resources and comparative genomics approaches may be useful for accelerating the breeding of other niche crops.

References

Aiken SG (1988) Wild rice in Canada. NC Press in cooperation with Agriculture Canada and the Canadian Govt. Pub. Centre, Supply and Services Canada

Arumunagathan K, Earle ED (1991) Nuclear DNA content of some important plant species. Plant Mol Biol Rep 9:208–219

Bennetzen JL, Freeling M (1993) Grasses as a single genetic system: genome composition, collinearity and compatibility. Trends Genet 9:259–261

De Wet JM, Oelke EA (1978) Domestication of American wild rice (*Zizania palustris* L., *Gramineae*). J Agric Tradit Bot Appl 25:67–84

Duvall MR, Peterson PM, Terrell EE, Christensen AH (1993) Phylogeny of North American oryzoid grasses as construed from maps of plastid DNA restriction sites. Am J Bot 80:83–88

Edmands S, Timmerman CC (2003) Modeling factors affecting the severity of outbreeding depression. Conserv Biol 17:883–892

Fannucchi GT, Fannucchi WA, Craven S (1986) Wild rice in Wisconsin: its ecology and cultivation. Publ G3372. Univ. Wisconsin Coop. Ext. Serv., Madison

Fehr WR (1987) Principles of cultivar development. Macmillan Publishing Company, New York

Gale MD, Devos KM (1998) Comparative genetics in the grasses. Proc Natl Acad Sci USA 95:1971–1974

Gao LZ, Zhang CH, Jia JZ (2005) Cross-species transferability of rice microsatellites in its wild relatives and the potential for conservation genetic studies. Genet Res Crop Evol 52:931–940

Goff SA (1999) Rice as a model for cereal genomics. Curr Opin Plant Biol 2:86–89

Goldman DA (1990) The reproductive biology of a monoecious grass, *Zizania palustris* L. PhD thesis, University of Illinois at Urbana-Champaign

Great Lakes Indian Fish and Wildlife Commission (2006) A guide to understanding Ojibwe treaty rights. Great Lakes Indian Fish and Wildlife Commission Public Information Office, Odanah

Grombacher AW, Porter RA, Everett LA (1997) Breeding wild rice. Plant Breed Rev 14:237–265

Hass BL, Pires JC, Porter R et al (2003) Comparative genetics at the gene and chromosome levels between rice (*Oryza sativa*) and wildrice (*Zizania palustris*). Theor Appl Genet 107:773–782

Hufford K, Mazer SJ (2003) Plant ecotypes: genetic differentiation in the age of ecological restoration. Trends Ecol Evol 18:147–155

Imle PT (2001) QTL verification and testcross analysis of seed shattering in wild rice (*Zizania palustris* L.). MS thesis, University of Minnesota, St. Paul

Johnson E (1969) Archeological evidence for utilization of wild rice. Science 163:276–277

Kahler AL (2007) Adapting rice (*Oryza sativa*) Simple Sequence Repeat (SSR) markers for linkage mapping and marker-assisted selection of wild rice (*Zizania palustris*). MS thesis, University of Minnesota, St. Paul, MN

Kahler AL (2010) Genome organization and genetic diversity of wildrice (*Zizania palustris* L.). PhD thesis, University of Minnesota, St. Paul

Kahler AL, Porter RA, Phillips RL (2006) Expanding the molecular genetic map of American wildrice (*Zizania palustris*) through addition of SSR markers from rice (*Oryza sativa*). In: Abstracts of the plant and animal genome conference XIV, San Diego

Kahler AL, Kern AJ, Porter RA, Phillips RL (2011) Using SSR markers to characterize genetic diversity of wildrice populations. In: Abstracts of the plant and animal genome conference IXX, San Diego

Kahler AL, Kern AJ, Porter RA, Phillips RL (2012) Measuring genetic diversity of cultivated wild rice (*Zizania palustris*) populations. In: Abstracts of the plant and animal genome conference XX, San Diego

Kennard WC, Phillips RL, Porter RA, Grombacher AW (2000) A comparative map of wild rice (*Zizania palustris* L. $2n = 2x = 30$). Theor Appl Genet 101:677–684

Kennard WC, Phillips RL, Porter RA, Grombacher AW (2002) Genetic dissection of seed shattering, agronomic, and color traits in American wildrice (*Zizania palustris* var. L) with a comparative map. Theor Appl Genet 105:1075–1086

Kern AJ (2009) Developing microsatellite markers to understand the genetic diversity of wild rice in Lake Superior coastal habitats. Proc West Great Lakes Res Conf VII:34

Kern AJ, Kahler AL, Phillips RL (2011) Using microsatellite markers to characterize the genetic diversity of wild rice in Great Lakes coastal habitats: implications for restoration. Proc Am Water Resour Assoc (Wis Sect) 35:62

Kovach DA, Bradford KJ (1992) Temperature dependence of viability and dormancy of *Zizania palustris* var. interior seeds stored at high moisture contents. Ann Bot 66:297–301

Laikre L, Schwartz MK, Waples RS, Ryman N (2010) Compromising genetic diversity in the wild: unmonitored large-scale release of plants and animals. Trends Ecol Evol 25:520–529

Lu YD, Waller M, David P (2005) Genetic variability is correlated with population size and reproduction in American wild-rice (*Zizania palustris* var. *palustris*, Poaceae) populations. Am J Bot 92:990–997

McCouch SR, Teytelman L, Xu Y et al (2002) Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.). DNA Res 9:199–207

Meeker J (1996) Wild-rice and sedimentation processes in a Lake Superior coastal wetland. Wetlands 16(2):219–231

Minnesota Department of Natural Resources (2008) Wild rice in Minnesota: a wild rice study document submitted to the Minnesota Legislature by the Minnesota Department of Natural Resources. http://files.dnr.state.mn.us/fish_wildlife/legislativereports/20080215_wildricestudy.pdf

Odland WE, Baumgarten A, Phillips RL (2006) Ancestral rice blocks define multiple related regions in the maize genome. Plant Genome 1:41–48

Oelke EA (2007) Saga of the grain: a tribute to Minnesota cultivated wild rice growers. Hobar, Lakeville

Oelke EA, McClellan M (1992) Wild rice production research. In: Minnesota wild rice research 1991. Minnesota Agr Exp Sta, St. Paul

Paterson AH, Lin YR, Li Z et al (1995) Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. Science 269:1714–1718

Phillips RL, Odland WE, Kahler AL (2005) Rice as a model genome and more. Fifth International Rice Genetics Conference, Manila

Power P, Oxley FM (2004) Assessment of factors influencing Texas wild-rice (*Zizania texana*) sexual and asexual reproduction. U.S. Fish and Wildlife Service Report to the Edwards Aquifer Authority, San Antonio

Richards CM, Reilley A, Touchell DH, Antolin MF et al (2004) Microsatellite primers for Texas wild rice (*Zizania texana*), and a prelimary test of the impact of cryogenic storage on the allele frequency at these loci. Conserv Genet 5:853–859

Simpson GM (1966) A Study of germination in the seed of wild rice (*Zizania aquatica*). Can J Bot 44:1–9

State of Minnesota (2006) Wild rice labeling. Office of the revisor of statutes. State Statut 30:49

Temnykh S, Park WD, Ayres NM, Cartinhour S et al (2000) Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa* L.). Theor Appl Genet 100:697–712

Vennum T (2000) The traditional and social context of ricing. In: Williamson LS, Dlutkowski LA, McCammon Soltis AP (eds) Proceedings of the wild rice research and management conference, Great Lakes Indian Fish and Wildlife Commission, Odanah, pp 1–7

Waller DM, Lu Y, David P (2000) Population genetic variation among wild rice populations in northern Wisconsin. In: Williammson LS, Dlutkowski LA, McCammon Soltis AP (eds) Proceedings of the wild rice research and management conference. Great Lakes Indian Fish and Wildlife Commission, Odanah

Watson L, Dallwitz MJ (1992 onwards) The grass genera of the world: descriptions, illustrations, identification, and information retrieval; including synonyms, morphology, anatomy, physiology, phytochemistry, cytology, classification, pathogens, world and local distribution, and references. Version: 11 Feb 2012. http://delta-intkey.com

Weaver TL, Neff BP, Ellis JM (2005) Water quality and hydrology of the Lac Vieux Desert watershed, Gogebic County, Michigan, and Vilas County Wisconsin, 2002–2004. United States Geological Survey Scientific Investigations Report 2005–5237. U.S. Department of the Interior, Washington, DC

Wisconsin Department of Natural Resources (2003) Protecting the bays, streams, and Everglades of the north. In: Wisconsin Natural Resources Magazine. Wisconsin Department of Natural Resources, Madison

Wisconsin Department of Natural Resources (2006) A river under repair: the Fox fights back. In: Wisconsin Natural Resources Magazine. Wisconsin Department of Natural Resources, Madison

Part II
Genomics-Assisted Crop Improvement for Food Security

Chapter 10
Genomics-Assisted Allele Mining and its Integration Into Rice Breeding

Toshio Yamamoto, Yusaku Uga and Masahiro Yano

Abstract Understanding the association between nucleotide changes and phenotypic changes is necessary for germplasm enhancement but has been a significant challenge in the molecular genetics and breeding of rice. In this article, we summarize our efforts to develop plant materials such as chromosome segment substitution lines to enhance the genetic analysis of traits of interest. The power of genetic dissection of phenotypic traits by use of novel populations is illustrated by our genetic analysis of heading date. We also present examples of the discovery of useful alleles involved in disease resistance and drought avoidance. Finally, we describe the discovery of genome-wide single-nucleotide polymorphism, which facilitate genetic analysis. This new type of genetic marker has allowed us to uncover the genome architecture of modern cultivars in Japan. These areas of progress will gradually change the landscape of selection in rice breeding.

10.1 Introduction

It has been more than 8 years since the whole genome sequence of rice was published (IRGSP 2005). This information has contributed to improved selection strategies for rice breeding (marker-assisted selection: MAS) as well as to the functional analysis of rice genes. Elucidation of the associations between nucleotide changes and phenotypic changes is necessary to most effectively use MAS but has been a significant challenge in the molecular genetics and breeding of rice. Over the last decade, many studies have been performed to clarify the relationships between sequence variation and phenotypic changes (Yamamoto et al. 2009; Yonemaru et al. 2010). To enhance our ability to genetically dissect complex phenotypes, appropriate plant materials are needed. Discovery of genes with economic value can be achieved most effectively through the use of plant materials specifically designed for use in molecular analysis (Fukuoka et al. 2010a). These efforts have already resulted in the discovery and use of several new genes in modern rice breeding programs

M. Yano (✉) · T. Yamamoto · Y. Uga
National Institute of Agrobiological Sciences, Kannondai 2-1-2,
Tsukuba, Ibaraki 305-8602, Japan
e-mail: myano@affrc.go.jp

R. Tuberosa et al. (eds.), *Genomics of Plant Genetic Resources,*
DOI 10.1007/978-94-007-7575-6_10,

(Fukuoka et al. 2009; Uga et al. 2011). To facilitate allele mining using the novel plant materials that have been developed, several studies have also been undertaken for genome-wide discovery of single-nucleotide polymorphisms (SNPs) (McNally et al. 2009; Huang et al. 2010; Ebana et al. 2010; Nagasaki et al. 2010; Yamamoto et al. 2010; Arai-Kichise et al. 2011; Zhao et al. 2011). These SNPs facilitate cost- and labor-effective analysis of a large number of individuals in genetic studies and breeding programs (Meuwissen et al. 2001; Bernardo and Yu 2007; Nordborg and Weigel 2008; Huang et al. 2010; Zhao et al. 2011). These advances in plant materials and molecular markers are acting as a force to change the landscape of selection in rice breeding. In this article, we summarize our recent activity in developing resources such as plant materials and genetic markers, and we discuss the application of these tools to rice breeding. In the last decade, new technologies and enabling platforms have revolutionised our ability to dissect genomes and decipher gene function not only in rice, but also other crop species (Yano and Tuberosa 2009). Therefore, we believe that the strategy described in this paper can also be integrated with other novel approaches enhancing allele mining and its application to improve rice and other crops.

10.2 Discovery of Natural Variation for Use in Rice Breeding

10.2.1 Developing Advanced Plant Materials from Diverse Rice Accessions

Genetic dissection of the wide range of naturally occurring variation in rice has progressed substantially through the use of quantitative trait locus (QTL) analysis, resulting in molecular cloning of genes and loci of biological and agronomic interest. The success of these analyses depends strongly on the plant materials used. Several types of plant materials, such as recombinant inbred lines (RILs), backcrossed inbred lines (BILs), introgression lines (ILs), and near isogenic lines (NILs), can be used in molecular genetic analysis (Fig. 10.1). Each type of population has its own advantages and disadvantages for use in genetic analysis. For example, F_2 and BC_1F_1 populations are most frequently used in genetic analysis because of the relatively short time needed for their development, but they also have several disadvantages in terms of the reproducibility and reliability of phenotyping. In such populations, the phenotype must be measured in individual plants, without replication, which sometimes compromises the reliability of genetic analyses such as the detection of QTLs. RILs can also be used for QTL mapping: these lines allow replication, thereby providing more statistical power for the genetic dissection of complex traits. Although RILs can be used for reliable phenotyping of many traits because they can be used in replicated trials, one major problem with RILs as well as other primary mapping populations is the variation in heading date. In general, progeny derived from a cross between diverse accessions often exhibit wide variation in heading date, and transgressive segregation is often observed. Several morphological and physiological

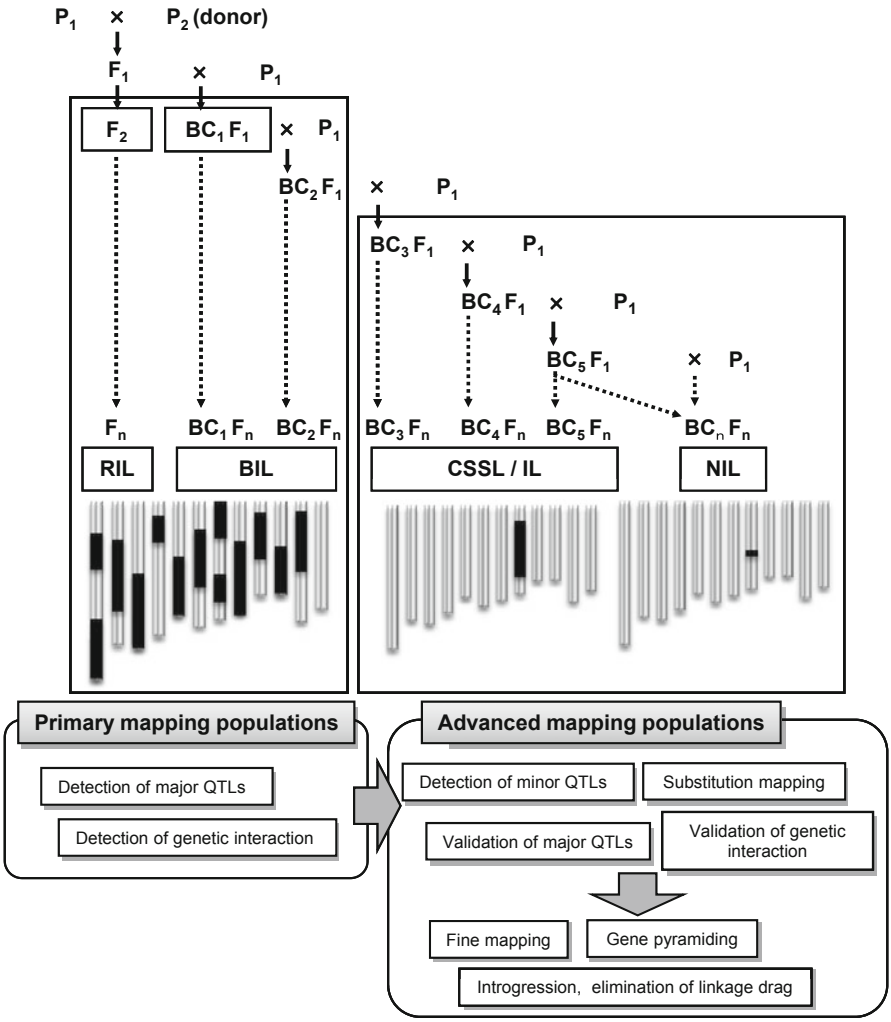

Fig. 10.1 Plant materials used for the genetic dissection of naturally occurring variation. *RIL* recombinant inbred line; *BIL* backcross inbred line; *CSSL* chromosomal segment substitution line; *IL* introgression line; *NIL* near-isogenic line

traits of agronomic interest, such as yield potential, eating quality, culm length, cold tolerance at the booting stage, and source (photosynthetic) ability, can be affected by heading date (reviewed by Fukuoka et al. 2010a). Therefore, it may be difficult to precisely evaluate such traits among segregants with large variation in heading date.

To more effectively detect relatively small allelic variations of agronomic value, it will be necessary to use mapping populations with a small range of variation in heading date and a more uniform genetic background. Eshed and Zamir (1995) proposed the novel concept of designing plant materials for genetic dissection based on their experience with a series of tomato introgression lines (ILs), each with one

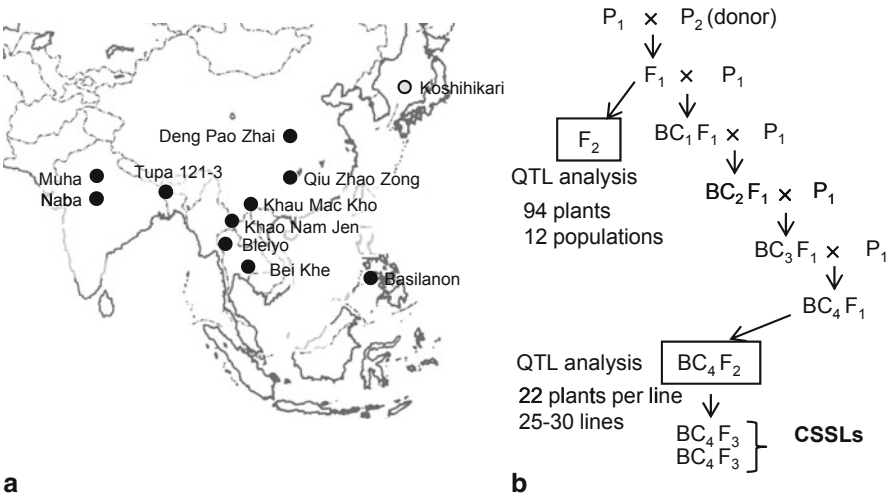

Fig. 10.2 (**a**) Geographical origin of Asian rice accessions. Koshihikari (*gray circle*) was used as a common parental strain for crosses with the 10 other accessions (*black circles*). (**b**) Pedigree of populations used in the analysis of heading date

or a few segment substitutions. So far, similar types of mapping populations have been developed as ILs and chromosome segment substitution lines (CSSLs) or single segment substitution lines (SSSLs) in rice (reviewed by Fukuoka et al. 2010a).

To develop CSSLs from multiple crosses, we selected 10 accessions from a rice core collection to comprehensively characterize the diversity of Asian cultivated rice (Kojima et al. 2005) (Fig. 10.2a). We selected these accessions based on the presence of sequence variation detected by means of restriction fragment length polymorphisms (RFLPs) and on their geographical origin (Kojima et al. 2005). The accessions originated from different regions of Asia and belong to three different cultivar groups: *japonica*, *aus*, and *indica* (Fukuoka et al. 2010a). Most of the plant materials are now at the stage of advanced backcross generations such as BC_4F_3 or BC_4F_4, and a set of CSSLs from these crosses will soon be available for public use. Once we have developed a series of CSSLs, a large-scale phenotyping experiment will be designed. By using CSSLs, the effects of variation in heading date should be minimized. Precise and reliable phenotyping of these lines will provide us more comprehensive understanding of rice morphological and physiological traits, allowing for a more effective mining of different types of alleles in diverse germplasms.

10.2.2 Genetic Architecture of Heading Date Revealed by Analysis of Novel Plant Materials

As noted above, a wide range of variation in heading date has been observed among rice cultivars. Although information on the genetic control of rice heading has accumulated rapidly (Izawa 2007; Tsuji et al. 2008), the genetic architecture underlying

heading date variation still remains to be clarified because of the limited number of cross combinations used in the analyses. It has also been very difficult to precisely compare QTLs detected in different studies because of the different level of mapping resolution in each study (Yonemaru et al. 2010). In general, because QTLs are detected based on the allelic differences between parental lines, it is difficult to determine whether a particular QTL is shared among different cross combinations. Therefore, to understand the genetic factors controlling heading date in diverse germplasms, a common parental line should be used for the development of mapping populations.

To understand the genetic control of heading date, we previously performed a QTL analysis in 12 populations derived from crosses of the *japonica* cultivar Koshihikari, as a common parental line, with diverse cultivars that originate from various regions in Asia; 10 of these accessions have been used as donor cultivars for CSSL development (Fig. 10.2a). QTL analyses using multiple crosses revealed a comprehensive series of loci involved in natural variation in flowering time. Interestingly, the chromosomal locations of those QTLs corresponded well with the locations of QTLs detected in other studies, such as *Hd1, Hd3a, Hd6, RFT1*, *Ghd7*, and *DTH8* (*Ghd8*) (Xue et al. 2008; Wei et al. 2010; Ebana et al. 2011; Yan et al. 2011). Candidate genes for several of these QTLs have been cloned and sequenced, confirming the involvement of these genes in phenotypic variation. Taking together the allelic differences found in the QTL analyses and the sequence variation within candidate genes for heading date, Ebana et al. (2011) concluded that a large portion of the wide range of phenotypic variation in heading date and daylength response could be generated by combinations of different alleles, possibly representing both loss and gain of function, of the QTLs *Hd1*, *Hd2*, *RFT1*, *Ghd7*, *DTH8* (*Ghd8*), and *Hd6*/*Hd16* (Ebana et al. 2011; Shibaya et al. 2011). Although we successfully detected several major QTLs by using F_2 populations, it is very likely that some additional QTLs with minor effects might be also involved in the phenotypic variation in these F_2 populations. To detect QTLs with minor effects, we have been performing genetic analysis using advanced backcross progeny (BC_4F_2) (Fig. 10.2b). Preliminary results have clearly demonstrated that a limited number of additional QTLs may be involved in natural variation for heading date (unpublished data). Once we complete the development of the CSSLs, we will attempt to verify the results obtained in the analysis of the BC_4F_2 populations. Furthermore, we might discover additional QTLs with minor or epistatic effects under diverse environmental conditions using these CSSLs, leading to more comprehensive understanding of the genetic factors that control natural variation in the heading date of Asian cultivated rice.

As mentioned above in our discussion of heading date, CSSLs derived from different donor cultivars will be valuable resources for the genetic dissection of naturally occurring phenotypic variations in diverse germplasms. In addition, Koshihikari, the recurrent parent in the CSSLs, is a current leading cultivar in Japan. Therefore, if we identify traits of agricultural and economical value in a particular CSSL, the line itself can be used as new breeding material.

10.3 Genetic Dissection of Agriculturally Important Traits

10.3.1 Characterization of Durable Resistance Genes for Rice Blast

Rice blast caused by the fungal pathogen *Magnaporthe oryzae* is the most serious disease in rice. During the last decade, many studies have been performed to detect and clone major genes involved in race-specific resistance to rice blast, and a series of race-specific resistance genes have been isolated by map-based cloning. However, breeding efforts using race-specific resistance have not usually led to the development of varieties with durable resistance because of the rapid breakdown of resistance caused by the emergence of new blast fungus races (Fukuoka et al. 2009). Therefore, many geneticists and breeders have been interested in using genes for partial resistance to rice blast, which might be more durable in the long term.

To understand the genes and biological mechanisms involved in durable resistance, two genes for rice blast resistance, *pi21* and *Pb1*, have been cloned by map-based cloning (Fukuoka et al. 2009; Hayashi et al. 2010). *Pb1* is considered to be a durable resistance gene, and recently it has been clarified to encode a typical plant disease resistance gene of the NBS-LRR type. Introduction of *Pb1* from *indica* cultivars into *japonica* cultivars has been performed, and resistance has not yet broken down after 30 years. Japanese upland rice cultivars exhibit a high level of resistance to rice blast, and this resistance is considered to be durable. Although many efforts have been made to introduce the genes involved in such resistance into lowland rice cultivars, no success in developing new cultivars has been achieved. The main reason for the failure of these efforts has been linkage of blast resistance and low eating quality. The blast resistance found in upland rice cultivars has been genetically dissected (Fukuoka and Okuno 2001), and one of the QTLs, *pi21*, has been cloned (Fukuoka et al. 2009). *Pi21* encodes a putative heavy metal binding protein with proline-rich sequences. A combination of two deletions, each within a putative proline-rich motif thought to be important for protein–protein interaction, causes a loss-of-function mutation in the *Pi21* gene, which normally suppresses defense response. It should be noted that the rapid onset but slow development of the *pi21*-mediated defense response may be beneficial both for the optimization of disease control and for reducing detrimental effects on agricultural traits.

Although the molecular mechanism of durable resistance generated by *pi21* needs to be clarified further, the resistant allele of *pi21* found in a limited group of *japonica* rice cultivars is a potential source for improving the blast resistance of rice worldwide. These efforts have opened up the possibility of introducing durable resistance into Japanese lowland cultivars without simultaneously introducing low eating quality. The durable blast-resistance gene *pi21* was found to be closely linked (within a 40-kb distance) with one or more genes associated with inferior eating quality (Fukuoka et al. 2009). In that study, desirable recombinants between *pi21* and the genes conferring inferior eating quality were successfully selected from a large breeding population by using DNA markers for the region around *pi21*.

The availability of DNA markers closely linked to genes of interest has enabled a breakthrough in the development of durably resistant cultivars that had not been achieved during 80 years of conventional breeding in Japan. This is a clear example of the power of MAS in breeding that could not have been achieved without the determination of the precise location of the gene of interest.

10.3.2 Genetic Control of Rice Root Morphology

Rice roots play a key role in the absorption and translocation of water and nutrients, but breeding for improvement in root traits poses unique challenges. Quantification of root traits in fields, especially paddy fields, is laborious and time consuming because the root spreads in a complicated pattern underground. Nevertheless, many researchers have applied the best available evaluation methods under several environmental conditions to detect QTLs for root morphological and anatomical traits such as maximum length, thickness, volume, and distribution. Environments used in these studies have included field plots (Li et al. 2005; Yue et al. 2005; Uga et al. 2008), cylinders or pots (Champoux et al. 1995; Yadav et al. 1997; Ali et al. 2000; Price et al. 2000; Zheng et al. 2000, 2003; Zhang et al. 2001; Kamoshita et al. 2002a, b; Price et al. 2002; Venuprasad et al. 2002; Courtois et al. 2003; Yue et al. 2006), and hydroponic culture (Price and Tomos 1997). As a result, a large number of QTLs have been identified as candidates underlying the natural variation of root traits. Courtois et al. (2009) summarized 675 QTLs for root traits that were reported in previous studies.

Why have researchers focused so intently on the exploration of root QTLs? The main reason may be that root traits appear to enable the plant to avoid drought stress by absorbing water deposited in deep soil layers (Yoshida and Hasegawa 1982). The wide extent of natural variation of root traits in rice has been revealed in previous studies (O'Toole and Bland 1987; Lafitte et al. 2001; Uga et al. 2009). For example, upland rice typically shows thicker and deeper rooting than lowland rice (O'Toole and Bland 1987). This diverse natural variation in rice could be a useful resource for improving drought avoidance in the field (Yoshida and Hasegawa 1982; Fukai and Cooper 1995). In fact, some root QTLs have been used for MAS in drought-resistance breeding (Shen et al. 2001; Steele et al. 2006). Steele et al. (2006) performed MAS to introduce the Azucena (upland *japonica*) allele at several QTLs for root traits into Kalinga III (upland *indica*), which had not been used in previous QTL analyses. Some advanced progenies having the selected QTL alleles showed significantly improved root traits, although other lines containing these alleles did not show positive effect for root phenotype. Therefore, QTLs affecting root traits across multiple genetic backgrounds have yet to be identified.

Recently, several QTLs for root traits have been fine-mapped as single loci by using advanced progeny: *qRL6.1*, a QTL for root length, on chromosome 6 (Obara et al. 2010); *Sta1*, a QTL for stele transversal area, on chromosome 9 (Uga et al. 2010); *Dro1*, a QTL for deep rooting, on chromosome 9 (Uga et al. 2011); *qFSR4*,

a QTL for root volume per tiller, on chromosome 4 (Ding et al. 2011); and *qSOR1*, a QTL associated with soil-surface rooting in paddy fields, on chromosome 7 (Uga et al. 2012). These root QTLs may be useful gene resources for improving rice production. For example, *DRO1* is expected to contribute to drought avoidance. *DRO1* was detected in RILs derived from a cross between the lowland cultivar IR64, with shallow rooting, and the upland cultivar Kinandang Patong (KP), with deep rooting (Uga et al. 2011; Uga et al. 2013). A near-isogenic line homozygous for the KP allele of *DRO1* (Dro1-NIL) in the IR64 genetic background showed significantly deeper roots than that of IR64 under upland field conditions. Moreover, the Dro1-NIL plants had a significantly larger panicle weight than that of IR64 under upland, drought-stressed conditions, suggesting that *DRO1* is involved in drought avoidance. Thus, QTL cloning and development of advanced progeny containing target QTLs within a defined genetic background can elucidate the relationships between root QTLs and abiotic stresses such as drought. In the future, it is expected that these QTLs will be isolated as single genes and their biological functions analyzed. Gene isolation is also beneficial from the point of view of breeding programs. If haplotype data of many rice accessions at a target QTL are obtained by DNA sequencing, we can detect the presence of favorable root QTL alleles in the different varieties. Such information will become a powerful tool for the breeding of root traits, which are difficult to investigate in the field.

10.3.3 Isolation and Pyramiding of Yield-Related Genes

Yield is the most important agronomic trait and the most common objective of breeding programs, but its genetic basis is so complex that, with only few exceptions, allele mining of yield-related traits is far behind that of other traits. Yield is the final mass of harvested organs, such as grain or straw, and is the product of numerous morphological and physiological traits. Moreover, each component trait is controlled by the effects of numerous loci and their interactions with environmental conditions. However, with the aid of recent progress of rice genomics research, some important genes involved in high yield have been revealed.

In crop physiological studies, yield is divided into three major conceptual components: sink size, i.e. the size and number of harvested organs; source strength, the ability to produce sucrose via photosynthesis; and translocation, the delivery of source products to sink organs. Among these three, the most QTLs have been identified for sink size because of the relative ease of sink measurement. In particular, some QTLs involved in culm length (Sasaki et al 2002), panicle architecture (Huang et al. 2009a; Miura et al. 2010), seed number (Ashikari et al. 2005), and grain size (Song et al. 2007; Shomura et al. 2008) have been isolated by map-based cloning, and their biological mechanisms have been clarified. Most of these genes are discovered independently in current high-yielding rice varieties, so enhancing the current level of yield potential will require combining these alleles by MAS. However, rice breeders and geneticists understand that there are trade-offs among

yield-related traits that prevent the expected increases from being fully attained. For example, Ohsumi et al. (2010) reported that gene pyramiding of two major sink-size genes, *Gn1a* (Ashikari et al. 2005) and *APO1/SCM2* (Ookawa et al. 2010), did not increase total grain yield because the increase in grain number was offset by changes in the ripening ratio, panicle number, and single-grain weight. The authors (Ohsumi et al. 2010) also reported that enlargement of sink size might enhance translocation of carbohydrate stored in the stem. This finding suggests that identification of major genes for source and translocation followed by combination with sink-size genes would be a promising way to improve rice yield.

Compared to the recent developments on sink traits, there has been substantially less progress toward isolation of genes involved in source and translocation traits except for reports of QTLs involved in grain filling (Wang et al. 2008), non-photochemical quenching, which regulates energy conversion in photosystem II (Kasajima et al. 2011) and leaf photosynthesis rate (Takai et al. 2013). Unlike the relatively straightforward methods that can be used to assess sink-size traits, the crop physiological evaluation systems for source and translocation traits that have been used for many years as the authoritative standards are difficult to apply to genetic analysis. For example, the most precise open gas-exchange system (LI-6400, Li-Cor Inc., Lincoln, NE, USA) for measuring leaf photosynthesis ability takes more than 10 min per sample for stabilizing the chamber environment. Given that a QTL analysis requires about 100 individuals derived from each cross, the length required for each measurement and the need for constant environmental conditions during the measurements makes it impractical to assess a population of this size in the field. Nevertheless, some QTLs related to source traits have been reported (Takai et al. 2009, 2010a; Xu et al. 2009; Adachi et al. 2010, 2011a, b; Gu et al. 2011), and One of such QTLs has recently been identified (Takai et al. 2013). Technological innovations are making field-scale measurements easier. For example, Takai et al. (2010b) revealed that leaf temperature measured by an infrared thermograph provides a promising indirect evaluation of stomatal conductance of the expanded leaf at the maximum tillering stage. We are currently applying this technique to detect QTLs involved in canopy photosynthesis.

For the future, it is important to pursue identification and functional analysis of QTLs for source and translocation. At this point, we still do not know whether genes that promote final yield (biomass) without affecting other traits actually exist. If such candidates are identified, it will be interesting to see whether these QTLs effectively increase source strength without having any pleiotropic effects on other growth-related traits.

10.4 Application of Genome-Wide SNP Analysis

The main problem in current MAS has been the shortage of available DNA markers showing high frequency of polymorphism among breeding varieties. Because of the time and labor required for genotyping experiments, it is unrealistic to develop a custom-made marker set for each breeding population. As the second best

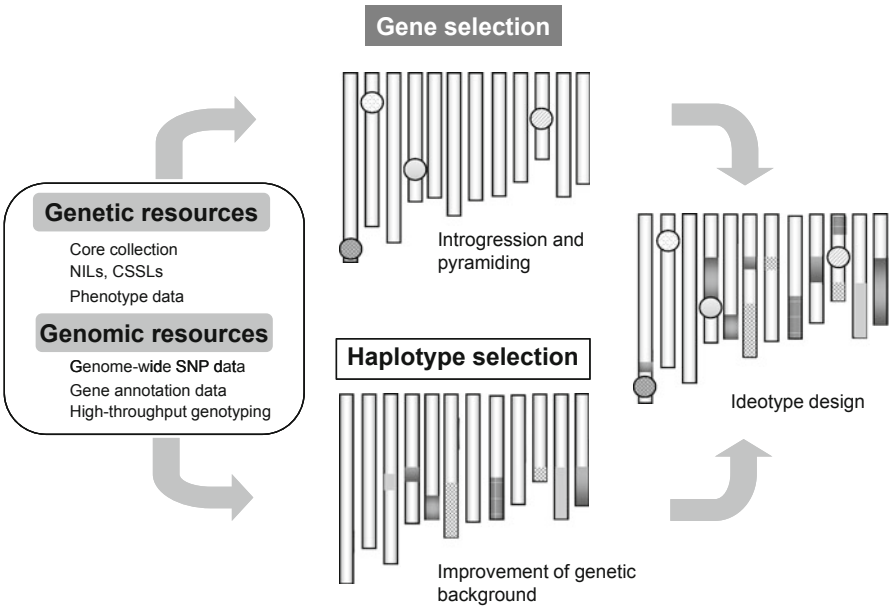

Fig. 10.3 Genomics-assisted breeding: integration of gene selection and haplotype selection

way, researchers have usually developed "semi-universal" marker sets selected from large numbers of SSRs or InDel sites distributed in the rice genome, which show an intermediate level of polymorphism among rice ecotypes (Orjuela et al. 2010). However, recent progress in DNA sequencing and genotyping has been reducing this limitation in a large number of crops (Feltus et al. 2004; Feuillet et al. 2010). Also, genome-wide resequencing of any of rice accession of interest to the researcher or breeder is now possible (Huang et al. 2009b). With concurrent innovations in genotyping technology (Gupta et al. 2008), it is possible to monitor the genomic constitution of breeding materials even with relatively similar genetic backgrounds (McNally et al. 2009; Yamamoto et al. 2010; Arai-Kichise et al. 2011).

The initial phase of MAS in rice was direct or indirect selection of useful gene loci from a donor variety and transfer into modern elite varieties. But the current lineup of such useful genes is limited to QTLs with major effects and mutated alleles whose phenotypic effects are easy to recognize. Assuming that slight modification of phenotype may be possible by introducing QTLs with minor effects or epistatic interactions, both of which are difficult to handle by MAS, we should be able to establish the relationship between each haplotype and specific phenotype. Genomics-assisted breeding will make it possible not only to select a useful allele at a specific locus but also to improve the total genetic background through haplotype selection (reviewed by Fukuoka et al. 2010b; Fig. 10.3). Plant breeders are always conscious of the extent of linkage blocks (haplotypes) being transferred during the breeding process. A comprehensive survey of modern rice varieties by genome-wide SNP analysis clarified the combinations of haplotype blocks within varieties that had been

derived from intercrossing of particular founder varieties (Yamamoto et al. 2010, Yonemaru et al. 2012). When we can annotate a haplotype in the pedigree record with a specific phenotypic trait, it will be possible to select the best cross combination for a certain breeding objective and to predict the level of improvement.

10.5 Conclusions and Future Prospects

Rice breeding can be described as a three-step process: (1) identify a favorable phenotype from among diverse genetic resources, (2) produce new gene combinations by crossing, and (3) select favorable segregants from the progeny of these crosses and make them genetically uniform. Allele mining is part of the first step of the above process and one of the major determinants for successful crop improvement. We have demonstrated the feasibility of genomics-assisted allele mining by developing CSSLs covering the natural variation of rice and using them to improve traits such as heading date. The technique can also be applied to agriculturally important traits, such as blast resistance, root morphology, or yield-related traits. We have also emphasized the importance of haplotype selection by genome-wide SNPs as a complementary approach to major-gene selection by MAS.

In addition to our approach described above, recent progress in genomics might further improve current existing strategy for allele mining. The accuracy of genome-wide association might be improved by the availability of new high-density genome-wide SNPs (Huang et al. 2010; Jahn et al. 2011), but these statistically predicted results need to be confirmed by observations of actual phenotypic differences. With respect to experimental materials, the utility of nested association mapping populations (McMullen et al. 2009) and multi-parent advanced-generation intercross populations (Cavanagh et al. 2008) has been validated in maize and *Arabidopsis*, respectively. These approaches combine diverse donors into one experimental population in return for keeping genetic uniformity, which is a major advantage of the CSSLs. These materials are expected to complement each other, and both types need to be developed in rice. Also, genomic selection (Hayes et al. 2009; Heffner et al. 2009) is now being applied in some crops (Zhong et al. 2009). To increase the number of new allele combinations, recurrent selection, which is commonly adopted in allogamous crops, should be considered in rice. Genome-wide SNPs could be helpful in verifying the effectiveness of recurrent selection by enabling surveys of changes in genome composition after each recurrent selection step.

With the aid of technological innovations resulting in automated and low-cost procedures, genotyping is no longer a limiting factor for MAS. On the other hand, the genetic basis of many economically important traits, most of which are complex traits, still remains to be elucidated. Therefore, both basic researchers and plant breeders should focus their efforts on improving methods for accurately evaluating small phenotypic differences to enable effective allele mining. The combination of improved evaluation tools and well-designed experimental materials such as CSSLs will enhance genomics-assisted breeding.

Acknowledgements We thank the Technical Support Section of NIAS for the management of the rice field. This work was supported by grants from the Ministry of Agriculture, Forestry and Fisheries of Japan (Integrated Research Project for Plant, Insect and Animal using Genome Technology QT-1005 and Genomics for Agricultural Innovation NVR-0001 to M. Yano, Genomics for Agricultural Innovation QTL-1002 to T. Yamamoto, and Genomics for Agricultural Innovation QTL-4003 to Y. Uga).

References

Adachi S, Tsuru Y, Kondo M et al (2010) Characterization of a rice variety with high hydraulic conductance and identification of the chromosome region responsible using chromosome segment substitution lines. Ann Bot 106:803–811

Adachi S, Nito N, Kondo M et al (2011a) Identification of chromosomal regions controlling the leaf photosynthetic rate in rice by using a progeny from *japonica* and high-yielding *indica* varieties. Plant Prod Sci 14:118–127

Adachi S, Tsuru Y, Nito N et al (2011b) Identification and characterization of genomic regions on chromosomes 4 and 8 that control the rate of photosynthesis in rice leaves. J Exp Bot 62:1927–1938

Ali ML, Pathan MS, Zhang J et al (2000) Mapping QTLs for root traits in a recombinant inbred population from two *indica* ecotypes in rice. Theor Appl Genet 101:756–766

Arai-Kichise Y, Shiwa Y, Nagasaki H et al (2011) Discovery of genome-wide DNA polymorphisms in a landrace cultivar of japonica rice by whole-genome sequencing. Plant Cell Physiol 52: 274–282

Ashikari M, Sakakibara H, Lin S et al (2005) Cytokinin oxidase regulates rice grain production. Science 309:741–745

Bernardo R, Yu J (2007) Prospects for genome-wide selection for quantitative traits in maize. Crop Sci 47:1082–1090

Cavanagh C, Morell M, Mackay I, Powell W (2008) From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. Curr Opin Plant Biol 11:215–221

Champoux MC, Wang G, Sarkarung S et al (1995) Locating genes associated with root morphology and drought avoidance in rice via linkage to molecular markers. Theor Appl Genet 90:969–981

Courtois B, Shen L, Petalcorin W et al (2003) Locating QTLs controlling constitutive root traits in the rice population IAC 165 × Co39. Euphytica 134:335–345

Courtois B, Ahmadi N, Khowaja F et al (2009) Rice root genetic architecture: meta-analysis from a drought QTL database. Rice 2:115–128

Ding X, Li X, Xiong L (2011) Evaluation of near-isogenic lines for drought resistance QTL and fine mapping of a locus affecting flag leaf width, spikelet number, and root volume in rice. Theor Appl Genet 123:815–826

Ebana K, Yonemaru J, Fukuoka S et al (2010) Genetic structure revealed by a whole-genome single-nucleotide polymorphism survey of diverse accessions of cultivated Asian rice (*Oryza sativa* L.). Breed Sci 60:390–397

Ebana K, Shibaya T, Wu J et al (2011) Uncovering of major genetic factors generating naturally occurring variation in heading date among Asian rice cultivars. Theor Appl Genet 122: 1199–1210

Eshed Y, Zamir D (1995) An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield associated QTL. Genetics 141: 1147–1162

Feltus FA, Wan J, Schulze SR et al (2004) An SNP resource for rice genetics and breeding based on subspecies *indica* and *japonica* genome alignments. Genome Res 14:1812–1819

Feuillet C, Leach JE, Rogers J et al (2010) Crop genome sequencing: lessons and rationales. Trends in Plant Science 16:77–88

Fukai S, Cooper M (1995) Development of drought-resistant cultivars using physio-morphological traits in rice. Field Crops Res 40:67–86
Fukuoka S, Okuno K (2001) QTL analysis and mapping of *pi21*, a recessive gene for field resistance to rice blast in Japanese upland rice. Theor Appl Genet 103:185–190
Fukuoka S, Saka N, Koga H et al (2009) Loss of function of a proline-containing protein confers durable disease resistance in rice. Science 325:998–1001
Fukuoka S, Nonoue Y, Yano M (2010a) Germplasm enhancement by developing advanced plant materials from diverse rice accessions. Breed Sci 60:509–517
Fukuoka S, Ebana K, Yamamoto T, Yano M (2010b) Integration of genomics into rice breeding. Rice 3:131–137
Gu J, Yin X, Struik PC et al (2011) Using chromosome introgression lines to map quantitative trait loci for photosynthesis parameters in rice (*Oryza sativa* L.) leaves under drought and well-watered field conditions. J Exp Bot 63:455–469
Gupta PK, Rustgi S, Mir RR (2008) Array-based high-throughput DNA markers for crop improvement. Heredity 101:5–18
Hayashi N, Inoue H, Kato T et al (2010) Durable panicle blast-resistance gene *Pb1* encodes an atypical CC-NBS-LRR protein and was generated by acquiring a promoter through local genome duplication. Plant J 64:498–510
Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME (2009) Genomic selection in dairy cattle: progress and challenges. J Dairy Sci 92:433–443
Heffner EL, Sorrels MR, Jannink JL (2009) Genomic selection for crop improvement. Crop Sci 49:1–12
Huang X, Qian Q, Liu Z et al (2009a) Natural variation at the *DEP1* locus enhances grain yield in rice. Nature Genet 41:494–497
Huang X, Feng Q, Qian Q et al (2009b) High-throughput genotyping by whole-genome resequencing. Genome Res 19:1068–1076
Huang X, Wei X, Sang T et al (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. Nature Genet 42:961–967
International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. Nature 436:793–800
Izawa T (2007) Daylength measurements by rice plants in photoperiodic short-day flowering. Inter Rev Cytol 256:191–222
Jahn CE, Mckay JL, Mauleon R et al (2011) Genetic variation in biomass traits among 20 diverse rice varieties. Plant Physiol 155:157–168
Kamoshita A, Zhang J, Siopongco J et al (2002a) Effects of phenotyping environment on identification of quantitative trait loci for rice root morphology under anaerobic condition. Crop Sci 42:255–265
Kamoshita A, Wade LJ, Ali ML et al (2002b) Mapping QTLs for root morphology of a rice population adapted to rained lowland conditions. Theor Appl Genet 104:880–893
Kasajima I, Ebana K, Yamamoto T et al (2011) Molecular distinction in genetic regulation of nonphotochemical quenching in rice. Proc Natl Acad Sci USA 108:13835–13840
Kojima Y, Ebana K, Fukuoka S et al (2005) Development of an RFLP-based rice diversity research set of germplasm. Breed Sci 55:431–440
Lafitte HR, Champoux MC, McLaren G, O'Toole JC (2001) Rice root morphological traits are related to isozyme group and adaptation. Field Crops Res 71:57–70
Li Z, Mu P, Li C et al (2005) QTL mapping of root traits in a doubled haploid population from a cross between upland and lowland *japonica* rice in three environments. Theor Appl Genet 110:1244–1252
McMullen MD, Kresovich S, Villeda HS et al (2009) Genetic properties of the maize nested association mapping population. Science 325:737–740
McNally KL, Childs KL, Bohnert R et al (2009) Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. Proc Natl Acad Sci USA 106:12273–12278
Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829

Miura K, Ikeda M, Matsubara A et al (2010) *OsSPL14* promotes panicle branching and higher grain productivity in rice. Nature Genet 42:545–549

Nagasaki H, Ebana K, Shibaya T et al (2010) Core SNPs set for the genetic analysis tool in Japanese rice population. Breed Sci 60:648–655

Nordborg M, Weigel D (2008) Next-generation genetics in plants. Nature 7223:720–723

Obara M, Tamura W, Ebitani T et al (2010) Fine-mapping of *qRL6.1*, a major QTL for root length of rice seedlings grown under a wide range of NH_4^+ concentrations in hydroponic conditions. Theor Appl Genet 121:535–547

Ohsumi A, Takai T, Ida M et al (2010) Evaluation of yield performance in rice near-isogenic lines with increased spikelet number. Field Crops Res 120:68–75

Ookawa T, Hobo T, Yano M et al (2010) New approach for rice improvement using a pleiotropic QTL gene for lodging resistance and yield. Nature Commun 1:132

Orjuela J, Garavito A, Bouniol M et al (2010) A universal core genetic map for rice. Theor Appl Genet 120:563–572

O'Toole JC, Bland WL (1987) Genotypic variation in crop plant root systems. Adv Agron 41:91–143

Price AH, Tomos AD (1997) Genetic dissection of root growth in rice (*Oryza sativa* L.). II: mapping quantitative trait loci using molecular markers. Theor Appl Genet 95:143–152

Price AH, Steele KA, Moore BJ et al (2000) A combined RFLP and AFLP linkage map of upland rice (*Oryza sativa* L.) used to identify QTLs for root-penetration ability. Theor Appl Genet 100:49–56

Price AH, Steele KA, Moore BJ, Jones RGW (2002) Upland rice grown in soil-filled chambers and exposed to contrasting water-deficit regimes: II. Mapping quantitative trait loci for root morphology and distribution. Field Crops Res 76:25–43

Sasaki A, Ashikari M, Ueguchi-Tanaka M et al (2002) Green revolution: a mutant gibberellin-synthesis gene in rice. Nature 416:701–702

Shen L, Courtois B, McNally KL et al (2001) Evaluation of near-isogenic lines of rice introgressed with QTLs for root depth through marker-aided selection. Theor Appl Genet 103:75–83

Shibaya T, Nonoue Y, Ono N et al (2011) Genetic interaction is involved in inhibition of heading by *Heading date 2* in rice under long day conditions. Theor Appl Genet 123:1133–1143

Shomura A, Izawa T, Ebana K et al (2008) Deletion in a gene associated with grain size increased yields during rice domestication. Nature Genet 40:1023–1028

Song XJ, Huang W, Shi M et al (2007) A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. Nature Genet 39:623–630

Steele KA, Price AH, Shashidhar HE, Witcombe JR (2006) Marker-assisted selection to introgress rice QTLs controlling root traits into an Indian upland rice variety. Theor Appl Genet 112: 208–221

Takai T, Ohsumi A, San-oh Y et al (2009) Detection of a quantitative trait locus controlling carbon isotope discrimination and its contribution to stomatal conductance in japonica rice. Theor Appl Genet 118:1401–1410

Takai T, Kondo M, Yano M, Yamamoto T (2010a) A quantitative trait locus for chlorophyll content and its association with leaf photosynthesis in rice. Rice 3:172–180

Takai T, Yano M, Yamamoto T (2010b) Canopy temperature on clear and cloudy days can be used to estimate varietal differences in stomatal conductance in rice. Field Crops Res 115:165–170

Takai T, Adachi S, Taguchi-Shiobara F, Sanoh-Arai Y, lwasawa N, Yoshinaga S, Hirose S, Taniguchi Y, Yamanouchi U, Wu J, Matsumoto T, Sugimoto K, Kondo K, lkka T, Ando T, Kono K, Ito S, Shomura A, Ookawa T, Hirasawa T, Yano M, Kondo M, Yamamoto T (2013) A natural variant of NAL1. selected in high-yield rice breeding programs, pleiotropically increases photosynthesis rate. Scientific Reports 3:2149

Tsuji H, Tamaki S, Komiya R, Shimamoto K (2008) Florigen and the photoperiodic control of flowering in rice. Rice 1:25–35

Uga Y, Okuno K, Yano M (2008) QTL underlying natural variation in stele and xylem structures of rice root. Breed Sci 58:7–14

Uga Y, Ebana K, Abe J et al (2009) Variation in root morphology and anatomy among accessions of cultivated rice (*Oryza sativa* L.) with different genetic backgrounds. Breed Sci 59:87–93

Uga Y, Okuno K, Yano M (2010) Fine mapping of *Sta1*, a quantitative trait locus determining stele transversal area, on rice chromosome 9. Mol Breed 26:533–538

Uga Y, Okuno K, Yano M (2011) *Dro1*, a major QTL involved in deep rooting of rice under upland field conditions. J Exp Bot 62:2485–2494

Uga Y, Hanzawa E, Nagai S et al (2012) Identification of *qSOR1*, a major rice QTL involved in soil-surface rooting in paddy fields. Theor Appl Genet 124:75–86

Uga Y, Sugimoto K, Ogawa S et al (2013) Control of root system architecture by *DEEPER ROOTING 1* increases rice yield under drought conditions. Nature Genet 45:1097–1102

Venuprasad R, Shashidhar HE, Hittalmani S, Hemamalini GS (2002) Tagging quantitative trait loci associated with grain yield and root morphological traits in rice (*Oryza sativa* L.) under contrasting moisture regimes. Euphytica 128:293–300

Wang E, Wang J, Zhu X et al (2008) Control of rice grain-filling and yield by a gene with a potential signature of domestication. Nature Genet 40:1370–1374

Wei X, Xu J, Guo H et al (2010) *DTH8* suppresses flowering in rice, influencing plant height and yield potential simultaneously. Plant Physiol 153:1747–1758

Xu YB, This D, Pausch RC et al (2009) Leaf-level water use efficiency determined by carbon isotope discrimination in rice seedlings: genetic variation associated with population structure and QTL mapping. Theor Appl Genet 118:1065–1081

Xue W, Xing Y, Weng X et al (2008) Natural variation in *Ghd7* is an important regulator of heading date and yield potential in rice. Nature Genet 40:761–767

Yadav R, Courtois B, Huang N, McLaren G (1997) Mapping genes controlling root morphology and root distribution in a doubled-haploid population of rice. Theor Appl Genet 95:619–632

Yamamoto T, Yonemaru J, Yano M (2009) Towards the understanding of complex traits in rice: substantially or superficially? DNA Res 16:141–154

Yamamoto T, Nagasaki H, Yonemaru J et al (2010) Fine definition of the pedigree haplotypes of closely related rice cultivars by means of genome-wide discovery of single-nucleotide polymorphisms. BMC Genomics 11:267

Yan WH, Wang P, Chen HX et al (2011) A major QTL, *Ghd8*, plays pleiotropic roles in regulating grain productivity, plant height, and heading date in rice. Mol Plant 4:319–330

Yano M, Tuberosa R (2009) Genome studies and molecular genetics—From sequence to crops: genomics comes of age. Curr Opin Plant Biol 12:103–106

Yonemaru J, Yamamoto T, Fukuoka S et al (2010) Q-TARO: QTL annotation rice online database. Rice 3:194–203

Yonemaru J, Yamamoto T, Ebana K et al (2012) Genome-wide haplotype changes produced by artificial selection during modern rice breeding in Japan. PLoS ONE 7:e32982

Yoshida S, Hasegawa S (1982) The rice root system: its development and function. In: Drought resistance in crops with emphasis on rice. International Rice Research Institute, Los Baños, Laguna, Philippines, pp 97–114

Yue B, Xiong L, Xue W et al (2005) Genetic analysis for drought resistance of rice at reproductive stage in field with different types of soil. Theor Appl Genet 111:1127–1136

Yue B, Xue W, Xiong L et al (2006) Genetic basis of drought resistance at reproductive stage in rice: separation of drought tolerance from drought avoidance. Genetics 172:1213–1228

Zhang J, Zheng HG, Aarti A et al (2001) Locating genomic regions associated with components of drought resistance in rice: comparative mapping within and across species. Theor Appl Genet 103:19–29

Zhao K, Tung CW, Eizenga GC et al (2011) Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. Nature Commun 2:467

Zheng H, Babu RC, Pathan MS et al (2000) Quantitative trait loci for root-penetration ability and root thickness in rice: comparison of genetic backgrounds. Genome 43:53–61

Zheng BS, Yang L, Zhang WP et al (2003) Mapping QTLs and candidate genes for rice root traits under different water-supply conditions and comparative analysis across three populations. Theor Appl Genet 107:1505–1515

Zhong S, Dekkers JC, Fernando RL, Jannink JL (2009) Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a barley case study. Genetics 182:355–364

Chapter 11
New Insights Arising from Genomics for Enhancing Rice Resistance Against the Blast Fungus

Elsa Ballini and Jean-Benoit Morel

Abstract The development of genomics has boosted research and our understanding of disease resistance in the model crop rice. While mapping and fine-mapping have been empowered by the release of thousands of new markers, this field of research is now entering a new era with the first examples of cloning genes by sequencing mutants. The few examples of genome wide association studies also indicate that identifying genes by this method will be challenging for minor effect loci. In all cases, phenotyping will be the key limitation to by-pass. The availability of large sets of genomic data also revealed several features of the defense system. While the polymorphism of genes upstream and downstream of the disease resistance pathway is mostly based on presence/absence, the polymorphism of genes in-between seems to rely more on expression level. Moreover, the different categories of genes involved in disease resistance are not evenly distributed in the genome. Altogether, the way this knowledge should modify our methods for breeding for disease resistance is discussed.

11.1 Introduction

Rice was the first crop of economical importance to be fully sequenced in the early times of the genomic era. Since the founding of the International Rice Genome Sequencing Project (IRGSP) in September 1997, the first releases of genomic information in the early 2000s[1] and the publication of two full genome sequences (*indica* genome 93-11 by Yu et al. (2002); *japonica* genome Nipponbare by Goff et al. (2002) and Genome (2005)), huge sets of data have been produced. It is no longer a matter of debate whether these large efforts to sequence the rice genome (∼ 370 Mb, ∼ 41,000 genes) have empowered gene discovery and our understanding

[1] http://www.sciencedirect.com/science/article/pii/S1369526699800181.

J.-B. Morel (✉) · E. Ballini
UMR BGPI INRA/CIRAD/SupAgro, Campus International de Baillarguet,
TA A 54/K, 34398 Montpellier, France
e-mail: benoit.morel@cirad.fr

R. Tuberosa et al. (eds.), *Genomics of Plant Genetic Resources,*
DOI 10.1007/978-94-007-7575-6_11,

of several biological processes. Here, we review the effects of the availability of the rice genome sequence on our understanding of the mechanisms behind disease resistance. In particular this review focuses on the interaction with *Magnaporthe oryzae* (causal agent of rice blast) which alone causes the equivalent of 1 Mha loss yearly in China (Ribot et al. 2008) and together with *Xanthomonas oryzae* pv *oryzae* (causal agent of bacterial blight) is estimated to reduce up to 21 % the global annual rice production (Waddington et al. 2010).

Plant disease resistance mechanisms can be schematically divided into three steps: pathogen recognition, signal transduction pathways and finally activation of defense, mostly relying on transcriptional regulation (Eulgem et al. 2007). For the purpose of this review, we define the entirety of the genes belonging to these three steps as the rice arsenal. Genes involved in recognition are mostly coding for proteins of the NBS-LRR category (Meyers et al. 2005) and often correspond to *R* (Resistance)-gene. Following recognition, a complex series of molecular events relying on a very diverse set of proteins like MAP kinases (Hamel et al. 2006) allow appropriate signal transduction that regulates transcriptional responses leading to the build-up of resistance. We collectively will call these gene regulators of disease resistance *REG* genes. Transcription factors regulate downstream genes in the process, many of which are classical *Pathogenesis-Related* genes (*PR*; van Loon et al. 2006) possessing antimicrobial activities. Besides *PR*-genes, many other genes are differentially expressed during infection, some of which belong to primary metabolic pathways like photosynthesis. These downstream genes will be identified as defense-related genes (*DEF*) as they merely correspond to indirect response of the plant to pathogen invasion.

Publications reporting the identification of *REG* genes in rice involved in disease resistance to either rice blast and/or bacterial blight, starting from the first one in 1999 (Kawasaki et al. 1999), clearly shows a direct link between the availability of the rice genome sequence and a steep increase in *REG* genes discovery rate (Fig. 11.1). This review focuses on the structural and functional aspects of the consequences of the genome publication on our understanding of rice blast disease resistance, taking into account also the information obtained through transcriptomics and other omics approaches.

11.2 Insights in Rice Blast Resistance Gained from the Genome Sequence

11.2.1 Mapping R-genes and QTLs

Since the first report of the cartography of a resistance gene to *M. oryzae* in 1966 (Yamasaki et al. 1966) more than 100 *R*genes have been identified and mapped (Ballini et al. 2008; Miah et al. 2013; Sharma et al. 2012). The access to the genome sequence and the common usage of basic molecular genetics tools has boosted and

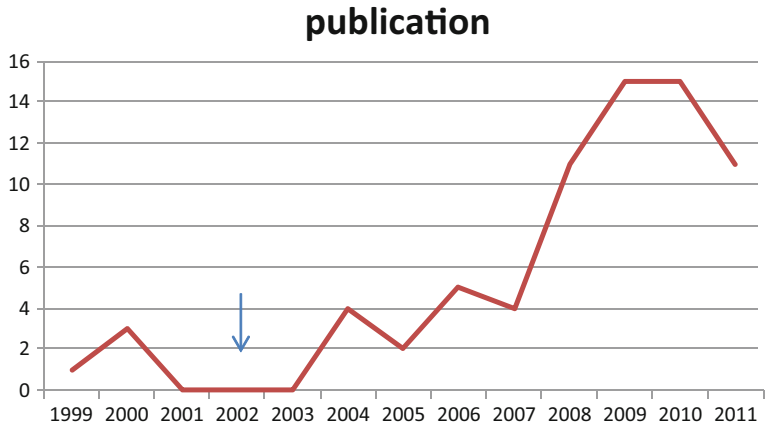

Fig. 11.1 Publications of identified rice genes required for disease resistance in rice. The genes selected here are those for which mutant lines (loss-of-function and/or gain-of-function) were produced to demonstrate their involvement in disease resistance (rice blast and/or bacterial blight). This represents a total of 89 genes in 2011, 57 of which are involved in rice blast resistance. The arrow indicates the release of the rice 93-11 indica genome (Goff et al. 2002)

refined the cartography of resistance gene in the last 10 years. Indeed 50 % of the *R* genes have been mapped since 2000 and the average mapping interval was refined from 6.5 to 1.8 Mbp (based on Archipelago[2]). Similarly, the QTL-mapping approach allowed to map more than 350 QTLs of resistance against *M. oryzae* (Ballini et al. 2008; Kou et al. 2012). Most of these data have been referenced in databases like Gramene or Archipelago that allow the physical mapping of these genetic data on the genome of the reference sequence Nipponbare (Vergne et al. 2008; Ni et al. 2009). Before 2000, molecular map-based approaches required time and money-consuming strategies: high-density linkage map construction, BAC-based physical map construction, sequencing of BAC clones. The release of genomic data has accelerated the generation of large segregating populations by Marker-Assisted Selection and allowed to skip the physical mapping by using an *in silico* physical map. The availability of molecular markers directly derived from the rice genomic sequence not only facilitates the mapping of resistance genes but also their transfer by breeders who can easily access this information (Hayashi et al. 2006; Ballini et al. 2009; Koide et al. 2009; Miah et al. 2013).

Additionally, a new era for mapping of resistance loci has begun with the recent deployment of association genetics strategies (Mackay et al. 2007; Nordborg et al. 2008; Ingvarsson et al. 2011). Indeed, besides the "classical" approach using a bi-parental population, a more recent approach is to use linkage disequilibrium in natural populations combined with the possibility to genotype the population at the whole genome level via genome wide association (GWAS). Several studies based on GWAS have been published recently in rice for agronomic traits (Huang et al. 2010, 2012; Zhao et al. 2011) and have taught us several points to be considered for

[2] http://orygenesdb.cirad.fr/.

applying this technique to map resistance loci against *M. oryzae*. First, the resolution of GWAS will not lead us to single genes due to the low rate of linkage disequilibrium decay in rice. Second, the population structure of *Oryza sativa* in five subspecies (including the two major *indica* and *japonica* sub-groups) and the genetic relatedness between cultivars has a strong impact on the trait association mapping (Zhao et al. 2011). Several analysis strategies can be proposed, incorporating or not the genetic relatedness but no consensus method can be derived (Myles et al. 2009). A mixed-model approach (Henderson 1975) allows to map resistance that segregates in several populations. For example, the *Pi-ta* resistance gene was mapped with a good confidence by this approach in Zhao et al. (2011). In the same study, using an approach not taking account population structure, the authors were able to detect a QTL close to *Pi37*. Based on their phenotyping data, there is a strong segregation for rice blast resistance in their population, with more resistant phenotypes detected in the *indica* and tropical *japonica* subpopulations. Third, these studies shed a new light on the debate concerning the contribution to phenotypic variance explained by loci identified by association genetics. In Zhao et al. (2011), application of association genetics to rice blast resistance shows relatively good results compared to other traits such as seed number per panicle. The maximum effect locus (*Pi-ta*) explained 8 % of the variance and altogether the significant loci explain 25 % of the variance. However, if a major, highly heritable *R* gene is detected with only 8 % of the variance explained, one can wonder whether this strategy will be well-suited to map partial resistance loci that have minor effects. Fourth, GWAS may be inappropriate to map partial resistance loci if they are rare and with a small phenotypic effect. In that scenario, it could be appropriate to combine mapping by association genetics with classical QTL mapping approaches (Myles et al. 2009). For this purpose the generation of a nested association mapping population could be highly beneficial to the rice community (Bandillo et al. 2010).

An intermediate strategy is to use a targeted approach using association genetics with already known candidate genes. Such method was successfully used for salt tolerance on the European rice breeding population (Ahmadi et al. 2011). It allowed to validate candidate QTLs with a strong phenotypic impact and to identify suitable European germplasm for breeding. As a first step to this approach, we found in the Oryza SNP database (McNally et al. 2009) highly reliable SNP in 24 of the 57 known rice *REG* genes. It could be possible to investigate the contribution to phenotypic variance of these loci in a subpopulation of rice.

11.2.2 Forward-Genetics Approaches Applied to Arsenal Genes in the Genomic Era

While the availability of the genome has considerably improved the resolution of the genetic analyses for rice blast resistance, it also led to a burst in gene cloning by forward-genetic approaches. Since the first cloned *R*-gene *Pib* (Wang et al. 1999), 20 additional *R*-genes have been cloned for rice blast resistance (Das et al. 2012; Hua et al. 2012; Sharma et al. 2012). Most of the *R*-genes cloned so far code for

NBS-LRR protein, the most common class of *R*-genes known in plants. Five of these genes have been cloned recently with a map-based *in silico* approach or multifaceted genomic approach (Sharma et al. 2012). Similarly, to clone *Pia*, a candidate gene approach was used based on the genome comparison between Nipponbare and 93-11 (Okuyama et al. 2011). *Pia* has been subsequently identified among seven candidates using genotype/phenotype association in seven varieties, EMS mutant analysis and complementation.

Identifying the genes that contribute to QTLs is a great challenge made possible only recently by the development of dedicated genomic tools (Kou and Wang 2012). Two partial resistance QTLs have been cloned by a map-based approach: *pi21* (Fukuoka et al. 2009) and *Pb1* (Hayashi et al. 2010). In both instances, the Nipponbare genome was used as a reference to find candidate genes that were then validated by complementation or mutagenesis. This method is well-suited for loci with a major effect but not for minor effect QTL. A functional analysis approach was used to characterize minor QTLs (Hu et al. 2008). In this case, candidate genes are identified based on the genome sequence and a set of a candidate genes are chosen based on transcriptomic or sequence polymorphism data and validated using complementation or mutant analysis. Nine genes or gene families have been characterized this way (Hu et al. 2008; Manosalva et al. 2009; Kou and Wang 2012). However, this type of approach is based on assumptions about the biological function of the genes underlying the QTL, about correct genome annotation and about the existence of molecular polymorphism or changes in gene expression correlated with the phenotype. Although likely valid in certain cases, such assumptions have been disproved as generally valid at the genomic level (Ballini et al. 2008, 2009). Moreover, in each case, although a phenotypic effect was observed to provide arguments in favor of the candidate gene tested, no evidence was provided to rule out the existence of additional genes significantly contributing to the QTL. Even if genetic co-segregation was found, a validation step is needed to confirm the actual involvement of the gene in the phenotypic variation (Pflieger et al. 2001). Thus one may conclude from this analysis that one major drawback deriving from having the genome sequence is the tendency to bias sequence-oriented final steps of forward genetics, an approach that has proven to be powerful in discovering genes when applied without *a priori*.

Quite surprisingly, the arising of the genomic era in rice did not boost cloning of genes identified by genuine forward approaches, based on identification of mutant phenotypes and gene complementation. This approach was indeed very productive in *Arabidopsis* and led to the identification of many key regulators. Besides classical map-based cloning, another genome-based approach that has proven rather successful in rice is microarray-based cloning. In the case of the *SL1* gene, the authors used a microarray-based approach to identify deleted genomic regions in rice mutants selected from a large collection generated by gamma ray or fast neutron treatment (Bruce et al. 2009). A similar approach has been used to identify a gene required for *X. oryzae* pv *oryzae* resistance (*SNL6* gene; (Bart et al. 2010)). Besides these two examples, only four out of the 57 genes known so far to be required for blast resistance were identified using such an approach (Fig. 11.1). It is noteworthy that these four genes all correspond to spontaneous lesion phenotypes that are easy to screen

for in recombinant populations. Thus phenotyping disease resistance seems to be the real limit that rice researchers will have to by-pass in the future for forward-genetics approaches. Most recently, full genome sequencing was recently demonstrated to be amenable for cloning mutants in rice (Abe et al. 2012). This new approach mostly requires segregating populations derived from backcross and opens almost infinite perspectives for cloning genes in rice (see below).

Finally, a very direct exploitation of the availability of the rice genome was the identification and validation in rice of gene orthologs that were shown to be required for disease resistance in other plants. Several genes were identified using this approach like the rice *OsCERK1* and *OsLSD1* genes which are the functional orthologs of the Arabidopsis *CERK1* (Shimizu et al. 2010) and *LSD1* genes (Wang et al. 2005), respectively. This has been empowered by the development of systematic orthologous relationships between rice and Arabidopsis (Greenphyl; Rouard et al. 2011)).

11.2.3 Insights in Rice Blast Resistance Gained from Functional Genomics

All omics approaches have been applied to the analysis of rice blast resistance. Besides metabolomics (Parker et al. 2009) and proteomics (Li et al. 2012b), the most frequently used technique is transcriptomics. Recently the RNA-seq technology has been used to analyzed the transcriptome of rice and blast fungus (Kawahara et al. 2012). Since our last review on gene expression studies for rice blast referencing 31 publications (Vergne et al. 2008), there has been an increasing number of new reports, most of which are referenced in the Orygenes DB database (Droc et al. 2009) and the OGRO database (Yamamoto et al. 2012). Our last update in 2010 reports 3,102 genes to be differentially expressed upon infection by rice blast out of almost 6,000 genes found to be deregulated during pathogen challenge. There is yet no integrated/comprehensive analysis of this huge amount of data.

This large set of data was the basis for the identification of most disease resistance regulators (*REGs*) by reverse-genetics approach. This was made possible by the existence of two key tools for rice: transformation (RNAi or over-expression) and large insertion mutant collections $\sim 61\,\%$ of the genes tagged (Jiang et al. 2011). To our knowledge, 88 rice genes have been reported to be required for either or both rice blast and bacterial blight resistance[3], 57 of which are involved in rice blast resistance (Table 11.1). Several points can be learned from this list; (i) there has been a particular focus on transcription regulators (15 genes/57), (ii) in most cases, regulators of rice blast resistance also affect bacterial blight in a similar manner (only one exception with *OsDR8*; (Wang et al. 2006)) and (iii) among the identified genes negative regulators are now almost as numerous (20) as positive regulators (37). This category of genes coding for negative regulators opens new possibilities for breeding (see below).

[3] The complete list of genes is available at http://orygenesdb.cirad.fr/.

Table 11.1 Disease regulators required for rice blast resistance. References are available at http://orygenesdb.cirad.fr/

Accession	Alias	Annotation
Positive regulators of rice blast and bacterial blight resistance		
Os01g09800	NH1	Regulatory protein
Os01g12900	OsRac1	Rac-like GTP-binding protein 2
Os01g43540	OsSGT1	Co-chaperone
Os01g54600	OsWRKY13	WRKY TF protein
Os01g68770	OsSBP	Selenium-binding protein homologue
Os02g33180	OsRAR1	Integrin beta-1-binding protein 2
Os05g25770	OsWRKY45-1	WRKY TF protein
Os07g40290	OsGH3.8	Indole-3-acetic acid-amido synthetase GH3.8
Os09g36320	BSR1	Tyrosine protein kinase
Os10g11980	OsAT1 (spl18)	Transferase family protein
Positive regulators of rice blast		
Os01g03360	RBBI2-3	Serine proteinase inhibitor
Os01g08330	OsCDR1	Aspartic proteinase
Os01g49290	Rack1	Guanine nucleotide-binding protein subunit beta-like protein
Os01g57610	OsGH3.1	Indole-3-acetic acid-amido synthetase
Os01g59440	OsLRR1	eLRR-RLP
Os01g60020	OsNAC4	Transcription factor
Os01g60490	OsWRKY22	OsWRKY22 - Superfamily of TFs having WRKY and zinc finger domains, expressed
Os01g66120	OsNAC6	NAC domain-containing protein 48
Os03g04110	CEBiP	Chitin oligosaccharide elicitor-binding protein
Os03g12500	OsAOS2	Allene oxide synthase (P450 74A2)
Os03g12730	OsBRR1	LRR-RLK
Os03g55320	OsBIPP2C2a	PP2C
Os04g41160	OsOxi1	AGC kinase
Os05g27730	OsWRKY53	WRKY TF protein
Os05g31140	GNS1	1,3;1,4-beta glucanase
Os06g03580	OsBBI1	E3 ligase
Os06g30860	OsWRKY31	WRKY TF protein
Os06g51050	chi11	Chitinase
Os08g42580	OsCERK1	Protein kinase domain containing protein, expressed
Os08g44640	OsBISCPL1	Serine Carboxypeptidase
Os09g36420	HSP90	Heat shock protein
Os10g41999	Rir1b	PR protein
Os11g02240	OsCIPK15	CBL-Interacting protein kinase
Os11g46900	OsWAK1	Protein kinase
Os12g02200	OsCIPK14	CBL-Interacting protein kinase
Os12g02440	OsWRKY89	WRKY TF protein
Os12g16720	SL1	cytP450
Negative regulators of rice blast and bacterial blight		
Os07g34570	OsDR8	Thiazole biosynthetic enzyme 1-1
Os01g50770	SPL28	Adaptor complexes medium subunit
Os01g69080	OsSSI2	Steroyl Acyl Carrier FA Desaturase
Os05g04520	OsPti1 (ttm 1)	Cytoplasmic kinase
Os07g48820	rTGA2.1	Transcription factor HBP-1b
Os10g38060	OsPLDl31	Phospholipase D
Os10g38950	OsMAPK6	MAP kinase
Os12g38210	OsSpl11	U-Box/Armadillo Repeat Protein Endowed with E3 Ubiquitin Ligase activity

Table 11.1 (continued)

Accession	Alias	Annotation
Negative regulators of rice blast		
Os01g54890	OsERF922	ERF transcription factor
Os03g06410	OsACDR1	MAPKKK/protein kinase (pas sur, resultat blast)
Os03g17700	OsMAPK5a	MAP kinase
Os03g18070	OsFAD7/FAD8	Fatty acid desaturase
Os04g32850	pi21	Proline-rich protein
Os04g47300	OsCDPK12	CDPK
Os05g02070	OsMT2b	Metallothionein
Os05g45410	OsSpl7	Heat stress transcription factor protein
Os06g44010	OsWRKY28	WRKY TF protein
Os07g30774	OsCBT	Calmodulin (CaM)-binding protein
Os08g06280	OsLSD1	Zinc finger
Os09g25070	OsWRKY62	WRKY TF protein

11.3 Structure of Arsenal

A few years ago, we proposed the idea that the rice arsenal has an archipelago-like structure (Vergne et al. 2008). This was manifested by the well-known genomic clustering of *R* genes but also by the apparent physical clustering of defense-related genes, in particular *PR* genes (Fig. 11.2). The present update further confirms that the rice arsenal is not homogeneously distributed in the genome.

11.3.1 A Few Surprises for R and PR Genes

With the completion of several plant genomes, it has been possible to investigate the organization of plant resistance genes at the genome scale (Meyers et al. 2005; Li et al. 2010). Around 60 % of *R* genes, including those from rice, are organized in clusters with a copy number variable between rice varieties (Yang et al. 2006; Zhang et al. 2009). This structure has an implication on *R* gene diversity because a richer genetic variation is maintained in larger gene families (Yang et al. 2006, 2008). Finally, based on the integrated map of cereal *R* loci, it appears that there is little synteny between grass genomes with high presence/absence polymorphism and only 16 of the 495 *R* gene analogs conserved between the five currently available cereal genomes (Luo et al. 2012). Unfortunately, this suggests that *in silico* mapping will be difficult between grass species and raises the question of what do these 16 loci recognize and why they are so conserved.

To our big surprise, the *PR* genes also appeared to be clustered in the genome (Vergne et al. 2008; Fig. 11.2). Besides this finding in rice, it has been shown that this was a common phenomenon in plants since part of the Arabidopsis arsenal also displays an archipelago-like structure (Rizzon et al. 2006). It is still unclear whether this physical grouping of *PR* genes implies transcriptional co-regulation. In

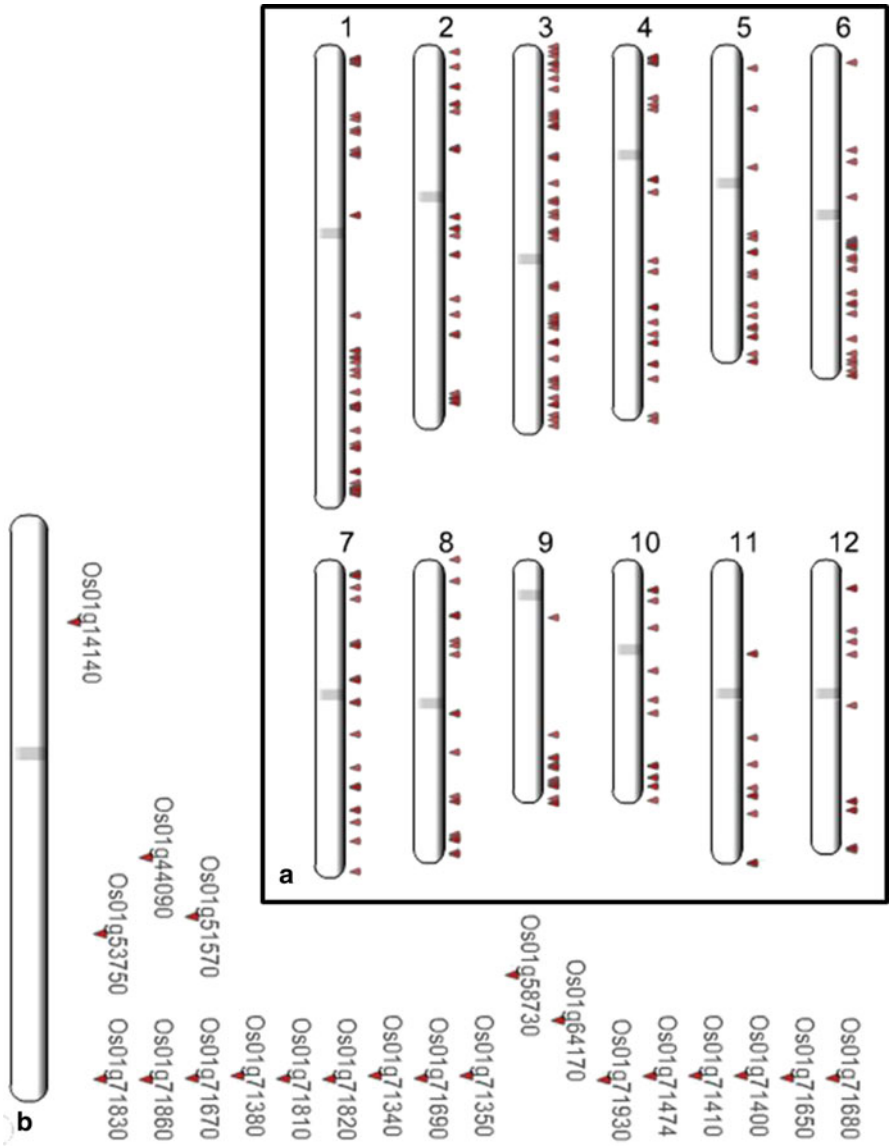

Fig. 11.2 Genomic localization PR genes in the rice genome. **a** The 378 genes annotated as *PRs* (Vergne et al. 2008) were vizualized on the genome. **b** An example of genomic cluster is shown for the glucanase genes on chromosome 1

some cases, a single transcription factor was shown to be sufficient to co-regulate transcription of such *PR* clusters (Grand et al. 2012).

Like in other cereals, some biochemical pathways seem to be physically clustered in the rice genome. For instance, the oat avenacin biosynthesis pathway is physically

clustered in a small interval (Qi et al. 2004). Similarly, the key enzymes for the biosynthesis of the antimicrobial Momilactones in rice are found within 17 kb on chromosome 4 (Shimura et al. 2007). This merely reflects the need for this pathway to be entirely inherited in order to work properly.

11.3.2 Some Chromosomes are Richer than Others in Key Regulators

Mapping the 57 known rice blast regulators (*REG* genes) on the genome shows that they are not uniformly distributed (Fig. 11.3). In particular, the long arm of chromosome 1 contains 12 of these 57 *REG* genes. On the other hand this is not related to a higher concentration of rice blast resistance QTLs on chromosome 1 in comparison to other chromosomes (Ballini et al. 2008). This suggests that for chromosome 1, *REG* genes may not be, for a large part, responsible for the mapped resistance QTLs. However, a larger examination of the co-localization of *REG* genes and mapped QTLs indicates that *REG* genes are often found, in a statistically significant manner, in regions where resistance QTLs have been mapped (Ballini et al. 2009). Thus, whether *REG*-like genes are good candidates for QTL or not is still an open question. It is interesting to see that chromosomes 11 and 12, which are enriched in *NBS-LRR* sequences and on which many *R* genes (36/88; Ballini et al. 2008) have been mapped, are poor in *REG* genes. This may have favorable consequences for breeding since it opens the possibility to independently select for some parental alleles on chromosomes 11 and 12 (*R* genes) and other alleles for chromosome 1 (*REG* genes). Thus, the fact that the rice arsenal does not seem uniformly distributed across the genome is a quite favorable feature for whole-genome selection approaches.

11.4 Structural and Functional Diversity of the Arsenal

11.4.1 Presence/Absence as a Source of Polymorphism

Based on the structure and diversity analysis, several types of *R* genes have been identified that seem to undergo different selection pressure and thus may offer different durability[4]. Indeed, it is possible to identify singleton genes with low diversity, *R* gene in clusters with high diversity and presence/absence loci (Yang et al. 2006, 2008).

Moreover, comparative genomic between two cultivated subspecies and wild species allowed to understand the effect of domestication on *R* genes (Yang et al. 2008; Zhang et al. 2009). Surprisingly, the effect of artificial selection is low and

[4] Durability is defined here as the property of a source of resistance to last over a long period of time once it has been deployed on a large surface.

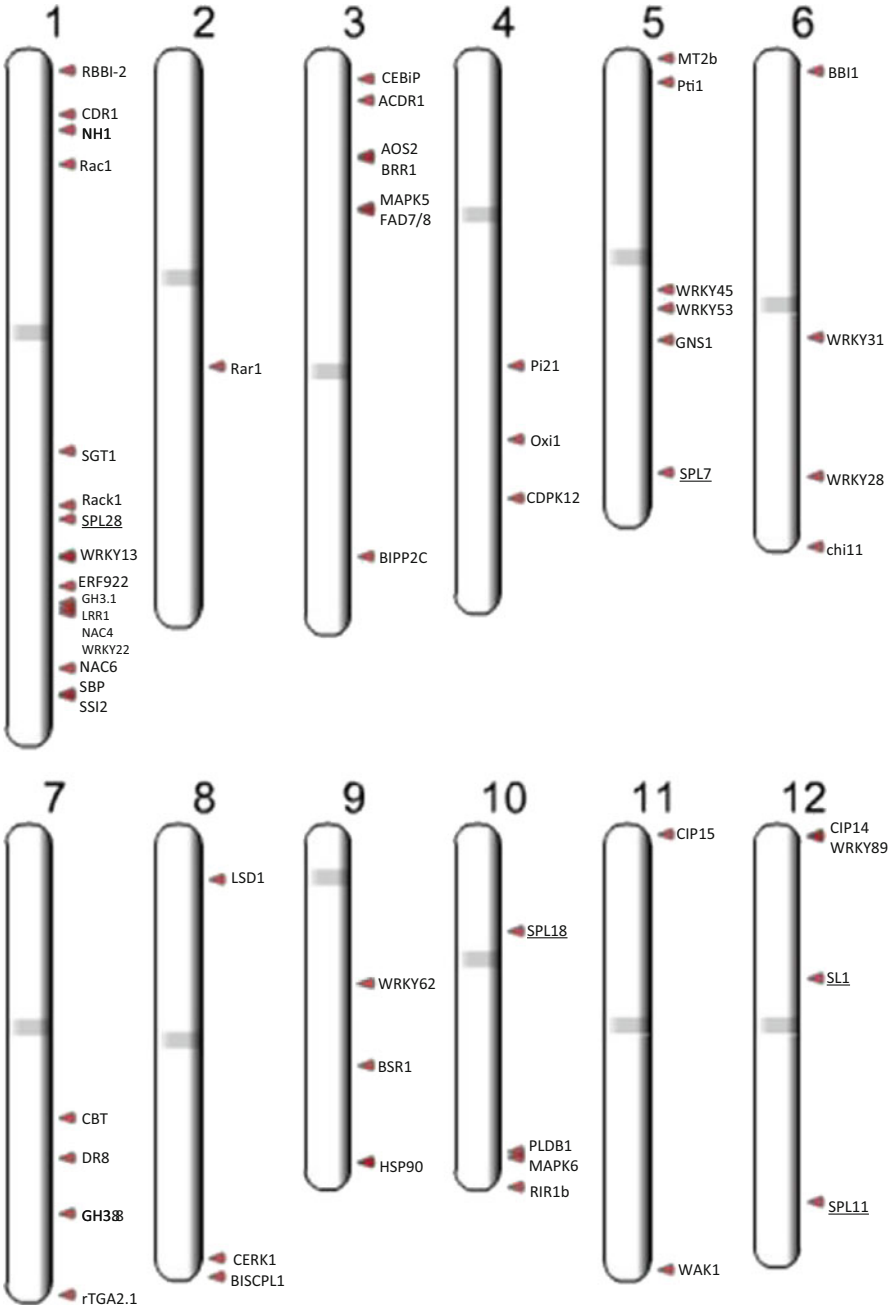

Fig. 11.3 Genomic localization of disease regulators involved in rice blast resistance. The 57 genes (*arrows*) were shown to be required for rice blast resistance. The genes underlined are those which were cloned using map-based cloning. The other genes were identified using reverse genetics approaches. Chromosome 1 shows a high density of regulators whereas there are a few on chromosomes 11 and 12, which contain a lot of R genes. Additional information is available at http://orygenesdb.cirad.fr/

a low differentiation was detected between wild and cultivated *R* gene alleles or between alleles of both subspecies. Comparative genomic between rice subspecies showed that 86 % of *R* genes are conserved but with in average 1.8 % sequence polymorphism (Yang et al. 2006, 2008). On the other hand, the average diversity of *R* genes is still around three times higher than the average total genome diversity (Yang et al. 2006; Zhang et al. 2009). The conserved polymorphism between wild species and cultivated species and the observed polymorphism pattern indicate that the majority of rice *R* genes undergo a strong balancing selection pressure instead of a diversifying selection. This also suggests that most *Oryza R* genes are ancient and that a similar selection pressure was present in both the *indica*and *japonica* environments. Alternatively, this could be explained by frequent introgressions of wild rice genome into cultivated rice after domestication.

One approach to extrapolate presence/absence of genes across rice diversity is to use presence/absence of significant signals in expression data. Through our analysis of constitutive expression of rice genes across rice diversity (including *japonica* and *indica*), we produced data that shed some light on the categories of genes that display high or low presence/absence polymorphism (Fig. 11.4a). This global analysis shows that *R* genes and *PR* genes display similar signatures; most genes in these categories (65 to almost 80 %) are often absent in at least one of the nine genotypes tested. This confirms that presence/absence is a common form of polymorphism for *R* genes as shown by enrichment analysis of NBS-LRR by array-based comparative genomic hybridization (Chen et al. 2011). In contrast, the vast majority (70–80 %) of *REG* genes or *DEF* genes were present in all tested genotypes. This suggests that these functional categories of genes are not dispensable, arguing in favor of universal regulation pathways for disease resistance.

11.4.2 Alternative Splicing of the Rice Arsenal

With the completion of the genome and the sequencing of several cDNA libraries, a genome-wide annotation of alternative splicing has been conducted in rice. The vast majority of *R* genes and *PR* genes are not alternatively spliced (Fig. 11.4b). Alternative splicing has been reported in rice NBS-LRR, in particular for the *Pi-ta R* gene where some alternative spliced versions encodes an additional C-terminal thioredoxin domain (Costanzo et al. 2009). Thus it has been suggested that alternative splicing may generate recognition diversity in NBS-LRR (Costanzo and Jia 2009). Another function of splice variants may be through nonsense-mediated decay that allows a fine regulation of R proteins accumulation and thus of cell death (Mastrangelo et al. 2012). However, *NBS-LRR* genes have a relatively low rate of alternative splicing (13.6 %; (Gu et al. 2007)). In contrast, *REG* genes seem to be more frequently alternatively spliced (Fig. 11.4b) and this may change the predicted protein structure (Campbell et al. 2006). Alternative splicing can be induced by both abiotic and biotic stresses (Mastrangelo et al. 2012). Indeed many stress-related genes present alternative splicing generating alternative variants that have in some cases a differential expression pattern during *M. oryzae* infection (Jung et al. 2009).

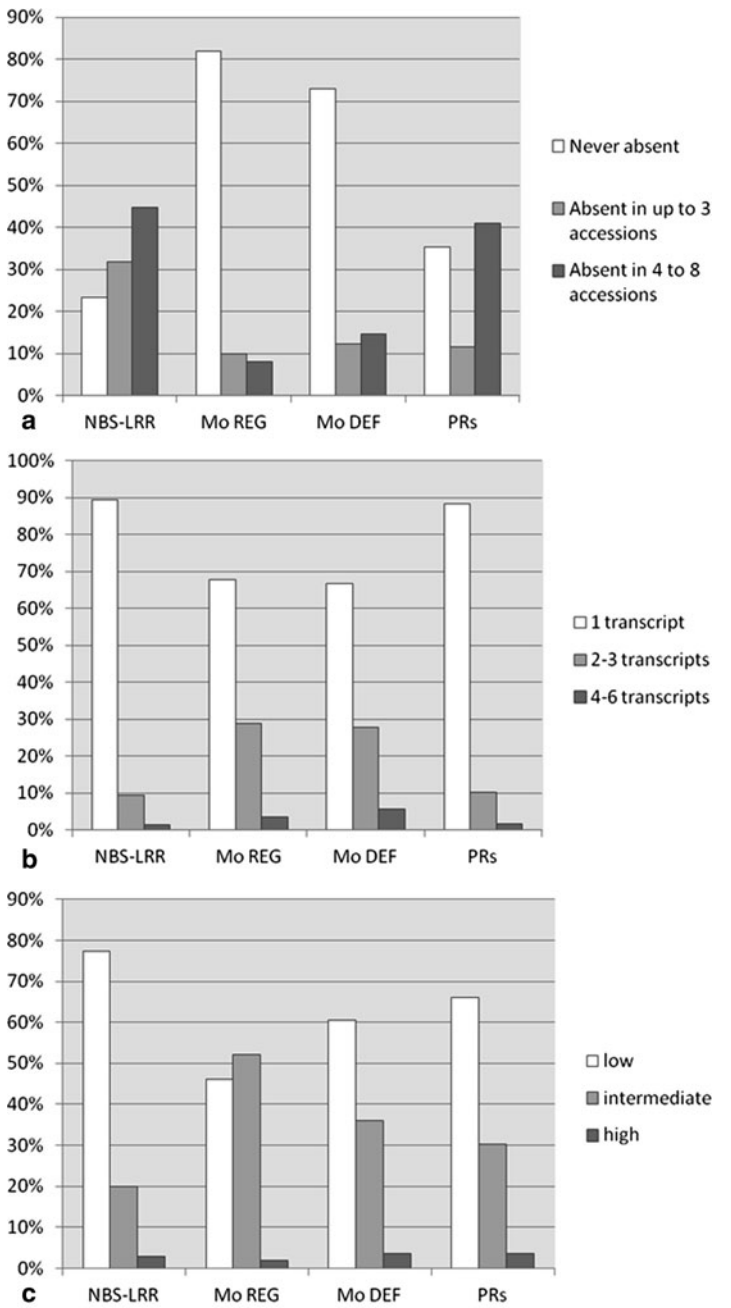

Fig. 11.4 Structural and expression polymorphisms of different components of the rice arsenal. The constitutive expression of the rice genes was measured in nine rice accessions (both from indica and japonica sub-groups) using Agilent chips (data from Grand et al. 2012). NBS-LRR represent *R*-genes of the *RPM1* type (145 genes), Mo REG represent 50 genes involved in resistance against blast (similar results were found with all known regulators), Mo DEF represents 2653 genes that

Thus, *R* genes and *PR* genes on one hand and *REG* genes on the other seem to display the same hallmarks in terms of presence/absence and of alternative splicing. Although it may be interesting to measure DNA polymorphism in the coding sequences of the arsenal genes, data from *Arabidopsis* suggest that there is low polymorphism at this level (Bakker et al. 2008). Another layer of diversity may arise from RNA editing, a phenomenon which has been associated with rice blast infection (Gowda et al. 2007).

11.4.3 Expression Diversity of the Arsenal

Besides alternative splicing, variability at the level of expression could also be a major way to generate functional diversity. This has been scarcely explored in plants. For instance in *Arabidopsis* it has been observed that the expression level polymorphism (ELP) of genes involved in response to biotic stimulus is high compared to other genes (Kliebenstein et al. 2006).

In their analysis across the *Oryza* gender including wild and cultivated species, Peng and colleagues (Peng et al. 2009) observed that genes involved in defense are not polymorphic with respect to constitutive expression in the panicle. They proposed that the evolutionary constraints on these genes were high and that domestication had little impact on them. In another study addressing ELP between *indica* 93-11 and *japonica* Nipponbare during the response to oxidative stress (Liu et al. 2010), the expression of genes involved in the phenylpropanoid pathway classically associated with disease response was found to be highly polymorphic.

The arsenal genes show elevated ELP as constitutive expression of some *PR* genes may vary more than 100-fold among cultivars (Vergne et al. 2010). When these constitutive expression levels and the known partial resistance levels were compared in a panel of cultivars, the constitutive expression of representative genes of the arsenal was found to be a hallmark of quantitative/partial resistance. This was exemplified by the observation that the more arsenal genes were expressed before infection in one given cultivar, the higher partial resistance this cultivar was going to display. Moreover, it was observed that there was more variability of expression of the arsenal before infection than after infection. This has implications for developing new approaches for breeding resistance (see below).

More recently, we measured by microarrays constitutive gene expression in rice leaves of nine *indica* and *japonica* cultivars (Grand et al. 2012). This set of data was used to examine ELP of the rice arsenal. This analysis reveals that *R* genes have relatively low ELPs and that *REG* genes have the highest ELPs (Fig. 11.4c). Thus, at least at the constitutive level, the arsenal genes show very contrasted ELPs.

---

are known to be differentially expressed upon rice blast infection and PR represents 191 *P*athogenesis-*r*elated genes for which data was available on the Agilent chips. **a** Presence/absence of significant signals was used as an indication of the presence of the corresponding gene. **b** Alternative splicing was evaluated using the MSU genome annotation. **c** The polymorphism of expression was estimated by measuring the CV for each gene across the nine genotypes

Thus, the rice arsenal is extremely polymorphic both at the DNA and RNA levels. This level of complexity has been poorly studied and will need additional research to really measure the impact of this diversity.

11.5 Breeding Perspectives

The identification of the component of the rice arsenal could be directly applied biotechnologically through GM approaches, whose potential against rice blast was recently reviewed (Delteil et al. 2010; Helliwell et al. 2013). However, up to now no rice plant genetically engineered for rice blast resistance has been deployed in the field. A relatively novel application of these findings is the use of negative regulators instead of the positive regulators to increase resistance. For instance, using presence/absence polymorphism for negative regulators, it should be possible to breed for improved resistance. Moreover, negative regulators can be used through non-GM approaches like Tilling or TALEN technology (Li et al. 2012a) where loss-of-function mutants are searched to gain a favorable feature. Mutation breeding has been used to improve rice for resistance to rice blast in several countries (Miah et al. 2013). The recently developed TALEN technology allows targeted recombination of DNA molecules with a very high efficiency. It was elegantly used for improving bacterial blight resistance by mutating the promoter of a gene required for susceptibility (Li et al. 2012a). One can imagine that exchanging a weak promoter of a central regulator with a strongly, pathogen-inducible, promoter should increase resistance. Recombinant technology, combined to the wealth of expression data available, opens many possibilities.

11.5.1 New Tools for Pre-Breeding

Another major outcome of this knowledge is the use of the genes discovered through classical breeding approaches. SNP markers can be more efficient than SSR markers for gene mapping or marker-assisted selection as SNP markers are more abundant, most of the time bi-allelic and finally, thanks to the new sequencing technologies, less expensive. They have been used to incorporate several R genes into elite breeding lines (Miah et al. 2013). Recently different rice SNP projects have allowed the discovery of several million SNPs in particular by the Rice SNP consortium (http://www.ricesnp.org/) (McNally et al. 2009; McCouch et al. 2010). Some of these SNPs have been validated with SNP chips containing a subset of SNPs that can be directly used by breeders (Chen et al. 2011; Thomson et al. 2012). For the moment, these chips have been designed to discriminate between the different rice subspecies. This kind of chips can be very interesting for rapidly genotyping a backcross family or a natural population (Yu et al. 2011). Until now, there is no example of QTL mapping using these chips for rice blast resistance. The next step may be to design dedicated SNP chips containing functional disease-related SNPs which could

be used to identify the alleles of most of the *R* genes, the principal *REG* genes and more generally the arsenal within the selected breeding population. This information could be used by breeders for a pre-breeding selection of their plant material. For example, the presence of the resistance gene *Pib* was recently screened in a breeding collection (Roychowdhury et al. 2012). Designing an "arsenal chip" will allow one to have similar information for a larger number of genes.

11.5.2 RNA-Based Breeding

As shown above, expression polymorphism is quite high for the arsenal genes. Some pieces of data even suggest that measuring constitutive expression of selected arsenal genes could be an efficient way of breeding for resistance (Vergne et al. 2010). This enhanced constitutive expression of the arsenal could merely reflect spurious activation of signaling pathways in some cultivars. In other words, measuring constitutive expression of the arsenal may be a way to detect the potential of inducible defense systems; the higher this potential is, the more constitutive expression can be detected. With the availability of medium throughput technology like quantitative PCR or nanostrings (Geiss et al. 2008), it is now conceivable to begin breeding using constitutive gene expression as a phenotype. Nanostrings are particularly adapted to this type of approach as they allow measuring the expression of hundreds of targeted genes (arsenal genes) in one step. It was successfully used in the case of the analysis of the cytokinin pathway in rice (Tsai et al. 2012). This RNA-based approach presents the advantage that it does not involve pathogen infection to establish phenotypic data that is related to final disease resistance. However, the proof-of-concept that this approach will be efficient remains to be provided.

11.6 Research Perspectives

As for other biological questions, new technologies are going to change our approaches to rice blast understanding. In particular, it is striking that the vast majority (93 %) of the discovered *REG* genes were identified using reverse-genetics approaches (53/57 genes). This is due to the fact that phenotyping is often limiting when doing classical map-based cloning with a limited number of markers. The situation has now drastically changed with the possibility of cloning by whole-genome sequencing. This was recently and elegantly shown to work in rice for two leaf mutants produced by EMS mutagenesis (MutMat approach; (Abe et al. 2012)). Quite amazingly, the identification of the genes responsible for the mutations was obtained with as little as 20 recombinant plants in clear contrast with the usual thousands of F_2 recombinants required for map-based cloning. This implies that the phenotyping of 20 recombinant families will be sufficient to clone a gene in the future as long as two conditions are fulfilled: the mutant phenotype must be the result of point or

small deletion mutations and the whole genomes of both the mutant and its original wild-type should be sequenced and compared. Thus, researchers should go back to their good-old collections in which rice blast resistance mutants have been found (Wu et al. 2005; Lorieux et al. 2012) and do some next-generation genetics using the MutMat approach.

Should this MutMat approach be undertaken, hunting for recessively-inherited gain-of-resistance is probably the next challenge. Indeed, the few examples of this kind available suggest a good potential for durability. The best examples are the *pi21* gene in rice (Fukuoka et al. 2009) and the *mlo* gene in barley (Peterhansel et al. 2005) which confer durable and recessive resistance to rice blast fungus and *Blumeria,*respectively. Similarly to *mlo*, the *pi21* gene seems to act as a negative regulator of disease resistance. This category of regulator may thus appear as good targets for breeding.

Reverse-genetics is also well-suited for identifying negative regulators. Thus far, researchers mostly focused on genes that were induced during infection. However, negative regulators like *pi21* are repressed during infection in order to further activate the disease resistance-signaling pathway (Delteil et al. 2012). Thus, it may be time to go back to the huge sets of transcriptomic data available and identify genes with repressed expression patterns. Another way of identifying *REG* genes with reverse genetics is to use constitutive expression patterns. We have shown that using this approach allows the identification of both positive and negative regulator genes in a very efficient manner (Grand et al. 2012).

Similarly to what was done in *Arabidopsis* (Hammond-Kosack et al. 2003), it will be important in the future to position the major components of rice blast resistance with respect to each other. For example, the requirement of central regulators like *NH1* or *MAPK* for different signaling pathways (triggered by *R* genes or by general receptors) will need to be addressed. This will be facilitated by the production of mutant collections in a limited number of genetic backgrounds (Delteil et al. 2012). However, the task of positioning each gene with respect to the others is probably ambitious and other approaches will be needed to establish regulatory networks. Using a guilt-by-association principle and yeast-two-hybrid technique, a first network connecting a large set of data has been recently produced (Seo et al. 2011). This approach postulates that proteins that are physically connected are often transcriptionaly co-regulated, a phenomenon that has been extensively observed in several biological systems (Wang et al. 2010). Other analyses specifically targeting response to bacterial blight showed the power of this type of approach (RiceNet: (Lee et al. 2011)) in discovering new regulators and networks. This remains to be done for rice blast resistance.

Despite its agronomical importance, to our knowledge there is no report yet on gene expression upon infection in the field. This remains a major challenge for disease resistance research as it is crucial to establish to which extent the findings from our laboratories translate to the field. Similarly, there are very few reports on testing the behavior of GM plants with improved resistance in the field (Delteil et al. 2010). In particular, this would address the question of the efficiency of the discovered disease resistance pathways in an agronomic environment where additional sources

of stress, such as drought stress, coexist. This is only starting to be done in *Arabidopsis* (Richards et al. 2012) and would shed new light on our understanding of transcriptional regulation during disease resistance in nature.

11.7 Concluding Remarks: Insights from the Enemy

In a similar way that the rice genomic sequence has accelerated *R* gene mapping and cloning, the completion of *M. oryzae* genome sequence has accelerated mapping and cloning of rice blast effectors and in particular of *Avirulence* genes (*Avr*) that are recognized by *R* genes (Dean et al. 2005; Valent et al. 2010; Xue et al. 2012). Sequence analysis and predicted proteome analysis allowed the identification of small secreted proteins that are good candidate effectors (Yoshida et al. 2009; Choi et al. 2010; Jung et al. 2012; Kim et al. 2013). The transcriptomic approach was also used combined with cellular techniques to extract fungal RNA inside plant tissue to identify new effector candidates (Oh et al. 2008; Mosquera et al. 2009; Soanes et al. 2012). Reverse-genetics was used to create *M. oryzae* mutants by large-scale targeted disruption of candidate effector genes and allowed the identification of genes necessary for pathogenicity (Saitoh et al. 2012). Association mapping was recently used to identify three new *Avr* genes: *AVR-Pia*, *AVR-Pii*, and *AVR-Pik/km/kp* with an EcoTilling approach (Yoshida et al. 2009). Comparative genomics approaches will allow us to identify core-effectors that are hardly dispensable. The matching *R* genes could then be identified and used as a promising source of durable resistance (Terauchi et al. 2010). Another perspective of the effector "omics" will be to identify the plant effector targets (besides *R* genes) by large-scale analysis through system biology approaches (Pritchard et al. 2011). These rice genes likely represent new targets for breeding durable resistance.

Acknowledgements We thank Enrico Gobbato for proof-reading this manuscript. This work is partly funded by the Cerealdefense project (ANR-09-GENM-013) and BAYER-CIFRE project.

References

Abe A, Kosugi S, Yoshida K et al (2012) Genome sequencing reveals agronomically important loci in rice using MutMap. Nat Biotechnol 30:174–178

Ahmadi N, Negrao S, Katsantonis D et al (2011) Targeted association analysis identified japonica rice varieties achieving Na+/K+ homeostasis without the allelic make-up of the salt tolerant indica variety Nona Bokra. Theor Appl Genet 123:881–895

Bakker EG, Traw MB, Toomajian C et al (2008) Low levels of polymorphism in genes that control the activation of defense response in *Arabidopsis thaliana*. Genetics 178:2031–2043

Ballini E, Morel J-B, Droc G et al (2008) A genome-wide meta-analysis of rice blast resistance genes and quantitative trait loci provides new insights into partial and complete resistance. Mol Plant Microbe Interact 21:859–868

Ballini E, Vergne E, Tharreau D et al (2009) ARCHIPELAGO: towards bridging the gap between molecular and genetic information in rice blast disease resistance. Springer, Dordrecht

Bandillo N, Muyco PA, Caspillo C et al (2010) Development of multiparent advanced generation intercross (magic) populations for gene discovery in rice (*Oryza sativa* L.). In: International Rice Research Institute LB, Laguna (Philippines) (ed) CSSP (Crop Science Society of the Philippines) Scientific Conference and Anniversary. Philipp J Crop Sci, Davao city, Philippines, p 96

Bart RS, Chern M, Vega-Sánchez ME et al (2010) Rice *Snl6* a cinnamoyl-coA reductase-like gene family member, is required for NH1-mediated immunity to *Xanthomonas oryzae* pv. oryzae. PLoS Genet 6:e1001123

Bruce M, Hess A, Bai J et al (2009) Detection of genomic deletions in rice using oligonucleotide microarrays. BMC Genomics 10:129

Campbell MA, Haas BJ, Hamilton JP et al (2006) Comprehensive analysis of alternative splicing in rice and comparative analyses with Arabidopsis. BMC Genomics 7:327

Chen HD, He H, Zou YJ et al (2011) Development and application of a set of breeder-friendly SNP markers for genetic analyses and molecular breeding of rice (*Oryza sativa* L.). Theor Appl Genet 123:869–879

Choi J, Park J, Kim D et al (2010) Fungal secretome database: integrated platform for annotation of fungal secretomes. BMC Genomics 11:105

Costanzo S, Jia YL (2009) Alternatively spliced transcripts of *Pi-ta* blast resistance gene in *Oryza sativa*. Plant Sci 177:468–478

Das A, Soubam D, Singh PK et al (2012) A novel blast resistance gene, *Pi54rh* cloned from wild species of rice, *Oryza rhizomatis* confers broad spectrum resistance to *Magnaporthe oryzae*. Funct Integr Genomics 12:215–228

Dean RA, Talbot NJ, Ebbole DJ et al (2005) The genome sequence of the rice blast fungus *Magnaporthe grisea*. Nature 434:980–986

Delteil A, Zhang J, Lessard P et al (2010) Potential candidate genes for improving rice disease resistance. Rice 3:56–71

Delteil A, Blein M, Faivre-Rampant O et al (2012) Building a mutant resource for the study of disease resistance in rice reveals the pivotal role of several genes involved in defence. Mol Plant Pathol 13:72–82

Droc G, Perin C, Fromentin S et al (2009) OryGenesDB 2008 update: database interoperability for functional genomics of rice. Nucleic Acids Res 37:D992–D995

Eulgem T, Somssich IE (2007) Networks of WRKY transcription factors in defense signaling. Curr Opin Plant Biol 10:366–371

Fukuoka S, Saka N, Koga H et al (2009) Loss of function of a proline-containing protein confers durable disease resistance in rice. Science 325:998–1001

Geiss GK, Bumgarner RE, Birditt B et al (2008) Direct multiplexed measurement of gene expression with color-coded probe pairs. Nat Biotechnol 26:317–325

Genome SPIR (2005) The map-based sequence of the rice genome. Nature 436:793–800

Goff SA, Ricke D, Lan T-H et al (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. japonica). Science 296:92–100

Gowda M, Venu RC, Li H et al (2007) *Magnaporthe grisea* infection triggers RNA variation and antisense transcript expression in rice. Plant Physiol 144:524–533

Grand X, Espinoza R, Michel C et al (2012) Identification of positive and negative regulators of disease resistance to rice blast fungus using constitutive gene expression patterns. Plant Biotechnol J 10:840–850

Gu L, Guo R (2007) Genome-wide detection and analysis of alternative splicing for nucleotide binding site-leucine-rich repeats sequences in rice. J Genet Genomics 34:247–257

Hamel LP, Nicole MC, Sritubtim S et al (2006) Ancient signals: comparative genomics of plant MAPK and MAPKK gene families. Trends Plant Sci 11:192–198

Hammond-Kosack KE, Parker JE (2003) Deciphering plant-pathogen communication: fresh perspectives for molecular resistance breeding. Curr Opin Biotechnol 14:177–193

Hayashi K, Yoshida H, Ashikawa I (2006) Development of PCR-based allele-specific and InDel marker sets for nine rice blast resistance genes. Theor Appl Genet 113:251–260

Hayashi N, Inoue H, Kato T et al (2010) Durable panicle blast-resistance gene *Pb1* encodes an atypical CC-NBS-LRR protein and was generated by acquiring a promoter through local genome duplication. Plant J 64:498–510

Helliwell E, Yang Y (2013) Molecular strategies to improve rice disease resistance. Methods Mol Biol 956:285–309

Henderson CR (1975) Best linear unbiased estimation and prediction under a selection model. Biometrics 31:423–447

Hu K-M, Qiu D-Y, Shen X-L et al (2008) Isolation and manipulation of Quantitative Trait Loci for disease resistance in rice using a candidate gene approach. Mol Plant 1:786–793

Hua L, Wu J, Chen C et al (2012) The isolation of *Pi1*, an allele at the *Pik* locus which confers broad spectrum resistance to rice blast. Theor Appl Genet 125:1047–1055

Huang X, Wei X, Sang T et al (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. Nat Genet 42:961–967

Huang X, Zhao Y, Wei X et al (2012) Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. Nat Genet 44:32–39

Ingvarsson PK, Street NR (2011) Association genetics of complex traits in plants. New Phytol 189:909–922

Jiang Y, Cai Z, Xie W et al (2011) Rice functional genomics research: Progress and implications for crop genetic improvement. Biotechnol Adv 30:1059–1070

Jung KH, Bartley LE, Cao PJ et al (2009) Analysis of alternatively spliced rice transcripts using microarray data. Rice 2:44–55

Jung Y-H, Jeong S-H, Kim SH et al (2012) Secretome analysis of *Magnaporthe oryzae* using in vitro systems. Proteomics 12:878–900

Kawasaki T, Henmi K, Ono E et al (1999) The small GTP-binding protein Rac is a regulator of cell death in plants. Proc Natl Acad Sci U S A 96:10922–10926

Kawahara Y, Oono Y, Kanamori H et al (2012) Simultaneous RNA-seq analysis of a mixed transcriptome of rice and blast fungus interaction. PLoS One 7:e49423

Kim SG, Wang Y, Lee KH et al (2013) In-depth insight into in vivo apoplastic secretome of rice-*Magnaporthe oryzae* interaction. J Proteomics 78:58–71

Kliebenstein DJ, West MAL, van Leeuwen H et al (2006) Genomic survey of gene expression diversity in *Arabidopsis thaliana*. Genetics 172:1179–1189

Koide Y, Kobayashi N, Xu DH et al (2009) Resistance genes and selection DNA markers for blast disease in rice (*Oryza sativa* L.). JARQ-Jpn Agr Res Q 43:255–280

Kou Y, Wang S (2012) Toward an understanding of the molecular basis of quantitative disease resistance in rice. J Biotechnol 159:283–290

Lee I, Seo YS, Coltrane D et al (2011) Genetic dissection of the biotic stress response using a genome-scale gene network for rice. Proc Natl Acad Sci U S A 108:18548–18553

Li J, Ding J, Zhang W et al (2010) Unique evolutionary pattern of numbers of gramineous NBS–LRR genes. Mol Gen Genomics 283:427–438

Li T, Liu B, Spalding MH et al (2012a) High-efficiency TALEN-based gene editing produces disease-resistant rice. Nat Biotechnol 30:390–392

Li Y, Zhang Z, Nie Y et al (2012b) Proteomic analysis of salicylic acid induced resistance to *Magnaporthe oryzae* in susceptible and resistant rice. PROTEOMICS:n/a-n/a

Liu F, Xu W, Wei Q et al (2010) Gene expression profiles deciphering rice phenotypic variation between Nipponbare (Japonica) and 93-11 (Indica) during oxidative stress. PLoS One 5:e8632

Lorieux M, Blein M, Lozano J et al (2012) In-depth molecular and phenotypic characterization in a rice insertion line library facilitates gene identification through reverse and forward genetics approaches. Plant Biotechnol J 10:555–568

Luo S, Zhang Y, Hu Q et al (2012) Dynamic nucleotide-binding site and leucine-rich repeat-encoding genes in the grass family. Plant Physiol 159:197–210

Mackay I, Powell W (2007) Methods for linkage disequilibrium mapping in crops. Trends Plant Sci 12:57–63

Manosalva PM, Davidson RM, Liu B et al (2009) A germin-like protein gene family functions as a complex quantitative trait locus conferring broad-spectrum disease resistance in rice. Plant Physiol 149:286–296
Mastrangelo AM, Marone D, Laido G et al (2012) Alternative splicing: enhancing ability to cope with stress via transcriptome plasticity. Plant Sci 185:40–49
McCouch SR, Zhao K, Wright M et al (2010) Development of genome-wide SNP assays for rice. Breed Sci 60:524–535
McNally KL, Childs KL, Bohnert R et al (2009) Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. Proc Natl Acad Sci USA 106:12273–12278
Meyers BC, Kaushik S, Nandety RS (2005) Evolving disease resistance genes. Curr Opin Plant Biol 8:129–134
Miah G, Rafii MY et al (2013) Blast resistance in rice: a review of conventional breeding to molecular approaches. Mol Biol Rep 40(3):2369–2388
Mosquera G, Giraldo MC, Khang CH et al (2009) Interaction transcriptome analysis identifies *Magnaporthe oryzae* BAS1-4 as biotrophy-associated secreted proteins in rice blast disease. Plant Cell 21:1273–1290
Myles S, Peiffer J, Brown PJ et al (2009) Association mapping: critical considerations shift from genotyping to experimental design. Plant Cell 21:2194–2202
Ni JJ, Pujar A, Youens-Clark K et al (2009) Gramene QTL database: development, content and applications. Database-J Biol Databases Curation
Nordborg M, Weigel D (2008) Next-generation genetics in plants. Nature 456:720–723
Oh Y, Donofrio N, Pan H et al (2008) Transcriptome analysis reveals new insight into appressorium formation and function in the rice blast fungus Magnaporthe oryzae. Genome Biol 9:R85
Okuyama Y, Kanzaki H, Abe A et al (2011) A multifaceted genomics approach allows the isolation of the rice *Pia*-blast resistance gene consisting of two adjacent NBS-LRR protein genes. Plant J 66:467–479
Parker D, Beckmann M, Zubair H et al (2009) Metabolomic analysis reveals a common pattern of metabolic re-programming during invasion of three host plant species by *Magnaporthe grisea*. The Plant J 59:723–737
Peng Z-Y, Zhang H, Liu T et al (2009) Characterization of the genome expression trends in the heading-stage panicle of six rice lineages. Genomics 93:169–178
Peterhansel C, Lahaye T (2005) Be fruitful and multiply: gene amplification inducing pathogen resistance. Trends Plant Sci 10:257–260
Pflieger S, Lefebvre V, Causse M (2001) The candidate gene approach in plant genetics: a review. Mol Breeding 7:275–291
Pritchard L, Birch P (2011) A systems biology perspective on plant–microbe interactions: Biochemical and structural targets of pathogen effectors. Plant Sci 180:584–603
Qi X, Bakht S, Leggett M et al (2004) A gene cluster for secondary metabolism in oat: implications for the evolution of metabolic diversity in plants. Proc Natl Acad Sci USA 101:8233–8238
Ribot C, Hirsch J, Balzergue S et al (2008) Susceptibility of rice to the blast fungus, *Magnaporthe grisea*. J Plant Physiol 165:114–124
Richards CL, Rosas U, Banta J et al (2012) Genome-wide patterns of *Arabidopsis* gene expression in nature. PLoS Genet 8:e1002662
Rizzon C, Ponger L, Gaut BS (2006) Striking similarities in the genomic distribution of tandemly arrayed genes in Arabidopsis and rice. PLoS Comput Biol 2:e115
Rouard M, Guignon V, Aluome C et al (2011) GreenPhylDB v2.0: comparative and functional genomics in plants. Nucleic Acids Res 39:D1095–D1102
Roychowdhury M, Jia Y, Jia MH et al (2012) Identification of the rice blast resistance gene *Pib* in the national small grains collection. Phytopathology 102:700–706
Saitoh H, Fujisawa S, Mitsuoka C et al (2012) Large-scale gene disruption in *Magnaporthe oryzae* identifies MC69, a secreted protein required for infection by monocot and dicot fungal pathogens. PLoS Pathog 8:e1002711
Seo Y-S, Chern M, Bartley LE et al (2011) Towards establishment of a rice stress response interactome. PLoS Genet 7:e1002020

Sharma T, Rai A, Gupta S et al (2012) Rice blast management through host-plant resistance: retrospect and prospects. Agr Res 1:37–52
Shimizu T, Nakano T, Takamizawa D et al (2010) Two LysM receptor molecules, CEBiP and OsCERK1, cooperatively regulate chitin elicitor signaling in rice. Plant J 64:204–214
Shimura K, Okada A, Okada K et al (2007) Identification of a biosynthetic gene cluster in rice for momilactones. J Biol Chem 282:34013–34018
Soanes DM, Chakrabarti A, Paszkiewicz KH et al (2012) Genome-wide transcriptional profiling of appressorium development by the rice blast fungus *Magnaporthe oryzae*. PLoS Pathog 8:e1002514
Terauchi R, Yoshida K (2010) Towards population genomics of effector–effector target interactions. New Phytol 187:929–939
Thomson M, Zhao K, Wright M et al (2012) High-throughput single nucleotide polymorphism genotyping for breeding applications in rice using the BeadXpress platform. Mol Breed 29:875–886
Tsai Y-C, Weir N, Hill K et al (2012) Characterization of genes involved in cytokinin signaling and metabolism from rice. Plant Physiol 158:1666–1684
Valent B, Khang CH (2010) Recent advances in rice blast effector research. Curr Opin Plant Biol 13:434–441
van Loon LC, Rep M, Pieterse CMJ (2006) Significance of inducible defense-related proteins in infected plants. Annu Rev Phytopathol 135–162
Vergne E, Ballini E, Droc G et al (2008) ARCHIPELAGO: a dedicated resource for exploiting past, present, and future genomic data on disease resistance regulation in rice. Mol Plant-Microbe Interact 21:869–878
Vergne E, Grand X, Ballini E et al (2010) Preformed expression of defense is a hallmark of partial resistance to rice blast fungal pathogen *Magnaporthe oryzae*. BMC Plant Biol 10:206
Waddington SR, Li XY, Dixon J et al (2010) Getting the focus right: production constraints for six major food crops in Asian and African farming systems. Food Sec 2:27–48
Wang PI, Marcotte EM (2010) It's the machine that matters: predicting gene function and phenotype from protein networks. J Proteomics 73:2277–2289
Wang ZX, Yano M, Yamanouchi U et al (1999) The Pib gene for rice blast resistance belongs to the nucleotide binding and leucine-rich repeat class of plant disease resistance genes. Plant J 19:55–64
Wang L, Pei Z, Tian Y et al (2005) OsLSD1, a rice zinc finger protein, regulates programmed cell death and callus differentiation. Mol Plant-Microbe Interact: MPMI 18:375–384
Wang G, Ding X, Yuan M et al (2006) Dual function of rice *OsDR8* gene in disease resistance and thiamine accumulation. Plant Mol Biol 60:437–449
Wu JL, Wu C, Lei C et al (2005) Chemical- and irradiation-induced mutants of indica rice IR64 for forward and reverse genetics. Plant Mol Biol 59:85–97
Xue M, Yang J, Li Z et al (2012) Comparative analysis of the genomes of two field isolates of the rice blast fungus *Magnaporthe oryzae*. PLoS Genet 8:e1002869
Yamamoto E, Yonemaru J, Yamamoto T et al (2012) OGRO: the overview of functionally characterized genes in rice online database. Rice 5:26
Yamasaki Y, Kiyosawa S (1966) Studies on inheritance of resistance of rice varieties to blast. I. Inheritance of resistance of Japanese varieties to several strains of the fungus. Bull Nat Inst Agric Sci Series D 14:39–69
Yang S, Feng Z, Zhang X et al (2006) Genome-wide investigation on the genetic variations of rice disease resistance genes. Plant Mol Biol 62:181–193
Yang SH, Gu TT, Pan CY et al (2008) Genetic variation of NBS-LRR class resistance genes in rice lines. Theor Appl Genet 116:165–177
Yoshida K, Saitoh H, Fujisawa S et al (2009) Association genetics reveals three novel avirulence genes from the rice blast fungal pathogen *Magnaporthe oryzae*. Plant Cell 21:1573–1591
Yu J, Hu S, Wang J et al (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). Science 296:79–92

Yu H, Xie W, Wang J et al (2011) Gains in QTL detection using an ultra-high density SNP map based on population sequencing relative to traditional RFLP/SSR markers. PLoS ONE 6:e17595

Zhang Y, Wang J, Zhang X et al (2009) Genetic signature of rice domestication shown by a variety of genes. J Mol Evol 68:393–402

Zhao K, Tung C-W, Eizenga GC et al (2011) Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. Nat Commun 2:467

Chapter 12
Enhancing Abiotic Stress Tolerance in Plants by Modulating Properties of Stress Responsive Transcription Factors

Maria Hrmova and Sergiy Lopato

Abstract Drought, heat and other abiotic stresses negatively impact growth, development, yield and seed quality of plants. The perception of stress and later adaptation to it occurs *via* signal transduction pathways that regulate expression of stress-responsive genes. Products of these genes include proteins that are directly involved in plant protection and those that fulfill regulatory function. The latter group includes transcription factors (TFs) and other transcription-related proteins that are investigated using the tools of forward and reverse genetics. Genomics analyses also revealed the importance of other proteins such as protein kinases and phosphatases, enzymes involved in metabolism of phospholipids, signalling molecules, etc. Once the stress response pathways are described, the role of key players in these pathways can be optimised through allele mining, selection and genetic engineering. These approaches offer alternatives to classical breeding and marker-assisted selection.

During plant responses to drought, a set of basic leucine zipper (bZIP), homeodomain-leucine zipper (HD-Zip) and WRKY TFs are transcriptionally or post-translationally activated via abscisic acid (ABA)-dependent signal transduction pathways. Despite a surge of data on the significance of plant bZIP, HD-Zip and WRKY TFs in the regulation of drought responses, the three-dimensional (3D) structures of these classes of TFs have been poorly defined. This structural information can be used for rational design of variant TFs that can help in understanding their oligomerisation and post-translational modification patterns, as well as their abilities to recognise target DNA sequences. In turn, this knowledge would permit the commercial application of genetically engineered TFs in agricultural biotechnology, by expression of TF variants and using the wild-type or modified promoter regions of stress-responsive genes. To this end, the aim of this review is to discuss strategies for improving tolerance of cereals to drought and other environmental stresses using molecular variants of the abiotic stress responsive TFs.

M. Hrmova (✉) · S. Lopato
Australian Centre for Plant Functional Genomics,
University of Adelaide, PMB 1, 5064 Glen Osmond, SA, Australia
e-mail: maria.hrmova@adelaide.edu.au

R. Tuberosa et al. (eds.), *Genomics of Plant Genetic Resources,*
DOI 10.1007/978-94-007-7575-6_12,

12.1 Drought Definition in the Agricultural Context

There is no universal definition of drought, although in an agricultural context drought can be defined as a prolonged, abnormally dry period when water resources are not sufficient for agricultural needs (Australian Government and Bureau of Meteorology 2012). Drought is also thought of as a form of complex environmental stress, which impacts agriculture in a variety of ways (Langridge et al. 2006; Prasad et al. 2008). Drought disrupts cropping programs, reduces breeding stock and by permanent erosion threatens a resource base of farming initiatives, hence causing a decline in the agricultural productivity of rural enterprises and national economies. For these reasons, managing drought is about managing risks associated with dry-land agriculture and aims to reduce the impact of drought.

In 2002, The Food and Agriculture Organization of the United Nations (FAO UN) developed the term 'Food security and safety', referring to 'a situation when all people, at all times, have physical, social and economic access to sufficient, safe and nutritious food' (FAO UN 2002). Food security and safety are expected to be threatened by the impact of drought in agriculture. Further, it has been concluded by the Intergovernmental Panel on Climate Change (IPCC) (IPCC 2007) that drought and drought-associated stresses in agriculture are predicted to be worsened by climate change and global warming. The IPCC has pointed out that elevated greenhouse gas concentrations are likely to create widespread climatic water stress, driven by drought and heat stresses (IPCC 2007). A comprehensive understanding of the impact of drought and heat stress will be critical in the evaluation of the impact of climate change and climate variability on crop production (Prasad et al. 2008).

In this context, a variety of other abiotic stresses other than drought and heat stresses, such as high salt levels, mineral deficiency and toxicity, intense light, excessive ozone and CO_2 levels act in combination, so these various categories of abiotic stress factors collectively challenge the planet and consequently affect all forms of life (IPCC 2007). Crop tolerance and susceptibility to these abiotic stresses is highly complex, and thus it is not surprising that abiotic stresses are the primary cause of crop loss worldwide, causing average yield losses of 50 % or more for major crops such as wheat, rice and maize (Valliyodan and Nguen 2006; Langridge et al. 2006; Prasad et al. 2008).

Being aware of these challenges, the FAO UN is advocating integrated approaches to drought management. These approaches aim to raise agricultural productivity on a sustainable basis in drought-prone areas around the world. Amongst the most important of the listed approaches are the development of drought-tolerant crop species and varieties, and advancements in intensified and diversified crop production systems (FAO UN 2002).

12.1.1 Counter-Acting Drought and Associated Stresses Through Traditional Crop Breeding, Gene Discovery and Genetic Engineering

Both drought and heat stress influence an array of plant characteristics, including physiology and development, and consequently change yield and quality of plant components consumed by people and animals (Prasad et al. 2008). In order to adjust to environmental stresses, plants have evolved a series of morphological and physiological adjustments during their vegetative growth and development. Crops are generally more sensitive to drought and heat stresses during their reproductive stages, as these stresses influence fertilisation and early stages of seed development, leading to a decrease in seed number and size, and influence overall seed quality (Prasad et al. 2008). However, strategies that plants have developed during millions of years of evolution are not well developed in most agricultural crops (Langridge et al. 2006).

In classical breeding and marker-assisted selection technologies, crosses are made in a relatively uncontrolled manner (Wieczorek 2003). Although the breeder chooses the parents, the results of such crosses are often unpredictable. Significant abiotic stress tolerances have been identified in land races and wild grass relatives and these desired traits can be crossed into crop species. It has been estimated that only 10–20 % of the wild variation has been used in modern wheat varieties (Langridge et al. 2006).

The alternative to classical breeding is combining established and emerging technologies with forward genetics, which discover and characterise key loci controlling stress tolerance. Emerging genomics technologies use strategies such as bioinformatics analyses, investigation of molecular function, gene expression and silencing in transgenic plants, phenotype analyses, analysis of regulatory networks, etc. Advances in the field of molecular biology during the last few decades, have provided us with the ability to manipulate DNA. One approach that genomic approaches use is to modify or manipulate biological function, through genetic engineering. The benefits or successful application of genetic engineering have led to increased crop productivity, enhanced crop protection and improved nutritional value and environmental benefits of transgenic plants. Conversely, there are serious issues related to consumer acceptance of genetically engineered crops in some countries. Over the past decade, genetic engineering has been intensively evaluated as a new tool for enhancing abiotic stress tolerance in model and crop plants. Evidence has started to accumulate that genetic engineering can be successfully used to produce plants with improved stress tolerance mechanisms (Tuberosa and Salvi 2006; Umezawa et al. 2006; Wan et al. 2009; Jewell et al. 2010; Yang et al. 2010).

12.2 Stress-Responsive Mechanisms During Drought and Other Abiotic Stress Conditions

Two major groups of proteins functioning in abiotic stress tolerance such as drought were identified in *Arabidopsis* (Shinozaki and Yamaguchi-Shinozaki 2007). The first group of proteins functions directly in plant protection, while the second group operates in signalling cascades and during transcriptional control (Valliyodan and Nguen 2006).

The first group of stress-responsive proteins includes Late Embryogenesis Abundant (LEA) proteins, chaperones, osmotins, anti-freeze proteins, mRNA binding proteins, enzymes involved in osmolyte biosynthesis, water channel proteins, sugar and proline transport proteins, detoxification enzymes and a variety of proteases. From this first group, the LEA proteins are particularly interesting as they represent 'the continuing conundrum of function' (Tunnacliffe and Wise 2007). The LEA proteins attribute their function in part to their structural plasticity or flexibility, as they are largely lacking in secondary structure in the fully hydrated state, but can become more folded during water stress and/or through association with membrane surfaces. Useful examples of LEA proteins are the dehydrins, which prevent the coagulation and inactivation of proteins during water deficit through a protein anti-aggregation mechanism (Goyal et al. 2005). The function of a specific pea LEA protein from mitochondria (LEAM), which is intrinsically disordered and folds into an amphipathic α-helix upon desiccation, has been investigated at the molecular level in model membrane systems (Tolleter et al. 2010). It was found that the LEAM proteins interacted specifically with negatively charged phosphate groups in dry phospholipids and increased fatty acyl chain mobility. Thus, LEAM proteins, depending on a lipid composition, interacted with membranes and protected the mitochondrial membrane structure against desiccation. These investigations highlighted the fact that LEA proteins could directly be involved in plant protection against water deficit.

12.3 Transcriptional Regulation of Plant Responses to Drought and Other Stress Conditions

The second group of stress-responsive proteins comprises regulatory proteins, i. e. proteins involved during regulation of signal transduction and in stress-responsive gene expression (Tuberosa and Salvi 2006; Valliyodan and Nguen 2006; Jewell et al. 2010). Genomics analyses of these proteins revealed various transcription factors (TFs), protein kinases and phosphatases, enzymes involved in metabolism of phospholipids and other signalling molecules, such as calmodulin-binding proteins.

As drought is perceived by a plant, the expression of numerous regulatory genes is affected and a series of TFs and other proteins are synthesised *de novo* or their levels are changed (Yamaguchi-Shinozaki and Shinozaki 2006; Cutler et al. 2010; Jewell et. al. 2010; Huang et al. 2012). However, before TFs are up- or down-regulated by

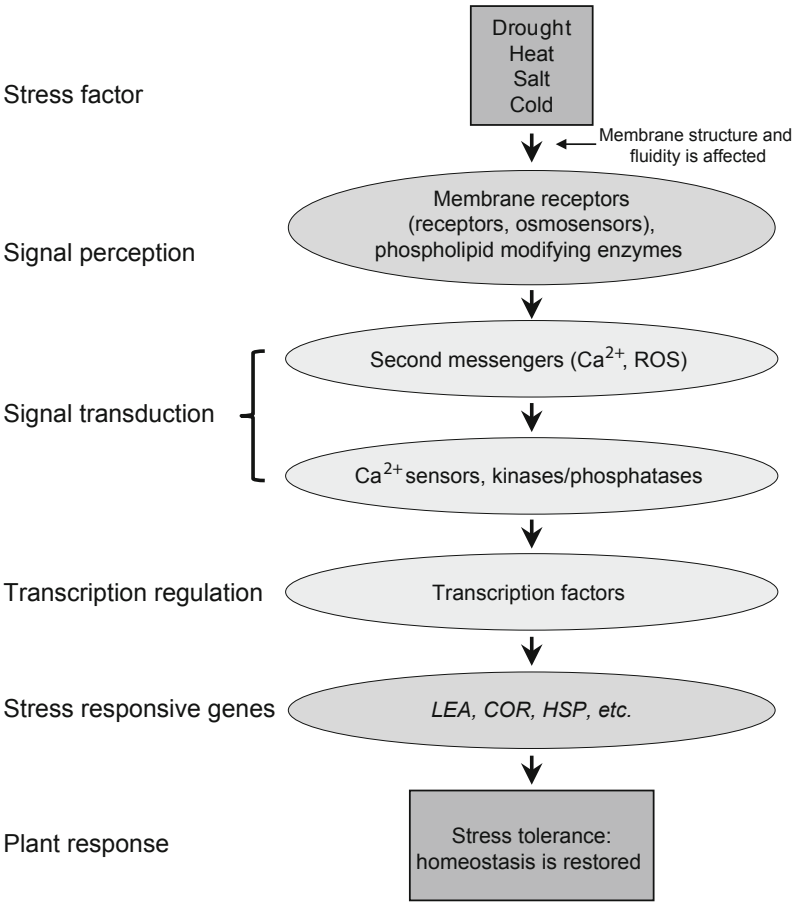

Fig. 12.1 Plants respond to abiotic stresses through a signal transduction pathway. The extracellular signal is perceived via membrane receptors that activate this complex signal transduction pathway, or by other means. The signal cascade results in the expression of multiple stress responsive genes, which mediate stress tolerance and restore homeostasis. (Figure based on Huang et al. (2012))

drought and other abiotic stresses, activation of signal transduction pathways that transmit the environmental signals to a molecular machinery of TFs takes place.

Understanding how these signal transduction pathways are regulated, as well as how their individual components can be manipulated, has fundamental importance for the development of drought tolerant crops, and by extension for the improvement of crop production. An example of the possible signal transduction pathway that is activated under abiotic stresses, such as drought, heat, salt and cold, and as this pathway progresses from perception of an abiotic stress to a plant adaptive stress response and development of tolerance and resistance, is schematically illustrated in Fig. 12.1. The signal transduction pathway starts with signal perception or sensing that has to

be relatively fast. It has been suggested that, for example during salt stress, such signal perception takes effect within seconds through salt sensory devices, presumably located in plasma membranes (Lopez-Perez et al. 2009). On the other hand, less accurate information exists on signal perception of drought and heat stresses (Jewell et al. 2012). Although it is expected that sensing and transmission of these abiotic stresses occur through plasma membrane receptors or sensory devices which lack enzymatic activity and operate by changing their conformation in a calcium-dependent manner (Saidi et al. 2011; Huang et al. 2012). It is currently unknown if drought and heat stresses are sensed directly through these specific receptors or sensors, or if stresses can be perceived by other means such as through changes of physical properties of plasma membranes (Saidi et al. 2010; Huang et al. 2012). Here, the changes in plasma membrane fluidity and/or permeability would lead to modifications of composition of lipids and other components within plasma membranes. For example, more saturated lipids and thus more rigid membranes were detected in heat-affected cellular membranes of *Physcomitrella patens* (Saidi et al. 2010). Further, the role of phospholipase D, that regulates the production of phosphatidic acid, a key class of lipid mediators in plant response to environmental stresses has been implicated in plant responses (Hong et al. 2010). It is expected that studies on the upstream regulators that activate different phospholipases and their downstream effectors have the potential to unveil the linkage between stimulus perception at the cell membrane with intracellular responses to drought and other stresses. Finally, in a forward genetic screen using 12 day-old *Arabidopsis* plants grown at 12 °C and shifted to 27 °C, the nucleosomes containing the alternative histone H2A.Z were found to be essential for correctly perceiving the ambient temperature signal. The H2A.Z-containing nucleosomes provided thermosensory information for *Arabidopsis* that would coordinate the ambient temperature transcriptome (Kumar and Wigge 2010). Apparently, the same thermosensor system was found in budding yeast, indicating that this is an evolutionarily conserved mechanism (Kumar and Wigge 2010).

After plants are exposed to stress and the stress signal is perceived by plants, their cellular membranes could be de-stabilised and activity and function of membrane proteins disrupted (Lopez-Perez et al. 2009; Maali Amiri et al. 2011; Huang et al. 2012). After the stress factor is initially perceived at the plasma membrane level, the signal is propagated in a second stage through accumulation or release of calcium ions, and/or through *de novo* formation and synthesis of second messenger molecules such as reactive oxygen species (ROS) and inositol phosphates. These second messengers start modifying intracellular concentrations of cytosolic calcium ion levels that lead to production of calcium ion binding proteins. These calcium ion binding proteins or calcium ion sensors, in turn initiate phosphorylation/de-phosphorylation cascades that target stress-responsive regulatory genes including TFs. As a result, switching on the phosphorylation/de-phosphorylation cascades via presumably kinase and phosphatase enzymes leads to stress tolerance. During this stress tolerance phase, many stress-related genes are up- and down-regulated in vegetative tissues and various stress resistance proteins and other molecules are accumulated (Fig. 12.1). As a result of activation of suites of downstream genes, plants adapt, survive and surpass the unfavourable environmental stress conditions.

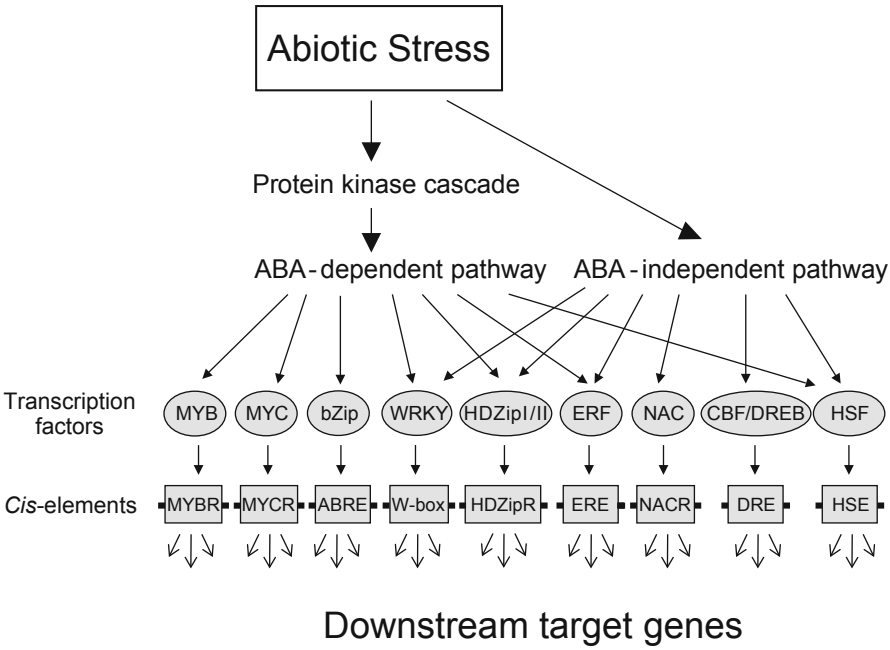

Fig. 12.2 Stress factors are perceived by plants through ABA-dependent and ABA-independent pathways or both. (Figure based on Hirayama and Shinozaki (2010))

12.3.1 Signal Transduction Pathways During Drought are Mediated Through Abscisic Acid-Dependent and Abscisic Acid-Independent Pathways

During vegetative growth the phyto-hormone abscisic acid (ABA) mediates the adaptive responses of plants to drought. A range of environmental stimuli and stresses are perceived through pathways that up- or down-regulate TFs (Fig. 12.2). In one example, over 200 TFs representing at least 20 protein families have been characterised by positive or negative regualtion by ABA at a single developmental stage, although the specific roles of individual TFs remain to be elucidated (Nemhauser et al. 2006). As illustrated in Fig. 12.2, the pathways leading to survival and adaptation of plants to abiotic stresses can be dichotomised into two major classes: (i) ABA-dependent pathways, and (ii) ABA-independent pathways (e.g. Shinozaki and Yamaguchi-Shinozaki 2007).

The adaptive response of plants to stress factors such as drought proceeds as follows. As water becomes depleted under water-deficit conditions, ABA is synthesised. The genes operating in plants during ABA biosynthesis and the sequence of the biosynthetic pathways have been elucidated (Nambara and Marion-Poll 2005). Additionally, ABA can also be produced by plant pathogenic fungi via a biosynthetic

route different from ABA biosynthesis in plants. Experiments with impermeable ABA analogues suggested that ABA may have both intracellular and extracellular sites of perception (Cutler et al. 2010). Proteins sensing ABA have been described and these proteins include ABA-binding proteins from barley aleurone, bean epidermal proteins and a variety of G-protein-coupled and other receptors isolated from a wide range of plants. When investigating ABA-binding Pyrabactin Resistance (PYR) receptor proteins from *Arabidopsis* through yeast two-hybrid systems, the C2 class of protein phosphatases PP2C that act as negative regulators of ABA signalling were discovered (Luan 2003). The mechanism of binding of ABA to the PYR receptors and the PP2C phosphatases that dock onto to the ABA-bound PYR receptors have been elucidated at atomic levels (Melcher et al. 2009). Here, many useful lessons have been learned through X-ray crystallographic investigations to understand the mechanisms of ABA signalling (Melcher et al. 2009).

During the signal transduction pathways regulated by ABA, plants respond with changes in gene expression. This may involve changes in rates of transcription, transcript processing and stability, and modifications of the conformational states of regulatory molecules that control RNA processing (Cutler et al. 2010). The transcriptional changes regulated by ABA have recently been investigated through transcriptome analyses, mainly using *Arabidopsis* (e. g. Nakashima et al. 2009). It has been found that approximately 5 to 10 % genes were affected, and more than 50 % of these changes were regulated by drought, salinity and ABA treatments. For example, the ABA-induced genes included a variety of TFs, protein kinases and phosphatases, dehydrins, transport proteins and ROS detoxifying enzymes. Some of the ABA-repressed gene products were linked to growth and included the ribosomal, plasma membrane, chloroplast and cell-wall proteins (Cutler et al. 2010). Also, a high number of ABA-regulated un-annotated transcriptional units was detected that may represent 'inter-genic' regions (Matsui et al. 2008; Zeller et al. 2009). The latter observation indicates that despite the significant progress in investigations of signal transduction during drought mediated by ABA-dependent pathways, a significant gap still exists in our understanding of these processes.

Analyses of the ABA-regulated signal transduction pathways have made it evident that certain loci identified in ABA-signalling act as the focal points that inter-connect plant responses to hormones, developmental signals, nutrients and environmental stimuli. For example, ABA-mediated plant responses to drought, heat, cold and light are cross-regulated by bZIP factors (Dekkers et al. 2008; Zhou et al. 2013).

To reiterate, during the stress response and adaptation of plants, many stress-related genes are up- and down-regulated and various stress resistance proteins accumulate. Amongst these genes, TFs play a fundamental regulatory role in modulating gene expression. TFs repress or activate suites of downstream genes that help maintain cell integrity and control plant growth under stress (Nakashima et al. 2009; Harris et al. 2011).

12.3.2 Transcription Factors as Key Regulatory Proteins Involved in Drought and Other Stress Responses

As mentioned above, the transcriptional regulation of gene expression is of fundamental importance during the development of plants under abiotic stresses and in ensuring that plant growth and development matches changes of environmental conditions. We know that development of each living organism has to proceed according to a detailed genetic plan regardless of the environmental conditions. However, this plan can be modified if stress conditions arise and therefore the roles of TFs in such modifications are of paramount importance. TFs have the potential to regulate gene transcription through recognition of specific DNA sequences in the promoter regions of their target genes. As TFs are essentially proteins that bind specific DNA sequences and thus control the flow of genetic information, it is not unexpected that very specific classes of TFs in appropriate conformational states need to be engaged under drought or other stress conditions. TFs often act as dimers or are associated with other proteins to form multi-subunit protein/DNA complexes. The total number of plant TFs identified in the fully sequenced genomes of *Arabidopsis*, rice, sorghum and maize genomes is about 1,550, 1,600, 2,450 and 3,300, respectively, or 3.5 to 7 % of all identified genes (Wang et al. 2011). Overall, the plant TFs are typically classified in up to 60 families, based on their primary sequence, 3D structures, DNA-binding motifs, oligomerisation patterns and post-translational modifications (Wang et al. 2011).

12.3.3 Changes of the Phosphorylation Status of Transcription Factors by Post-Translational Modifications and Influence of Such Changes on Their Functional Properties

It has been determined that a variety of kinases and phosphatases play important roles in ABA signalling (Cutler et al. 2010). Amongst kinases, both calcium ion-dependent (Coello et al. 2005; Kobayashi et al. 2005; D'Angelo et al. 2006; Bai et al. 2009; Zhang et al. 2011; Fujita et al. 2012) and calcium ion-regulated enzymes (Harmon 2003) surface as important candidates. These kinases have been found to operate as positive and/or negative regulators in ABA signalling.

Given that a multitude of kinases could be implicated in ABA signalling, ABA-sensitive phenotypes might be noteworthy for assessing the relative contribution of individual kinases. Recently, stable isotope labelling by amino acids in cell culture (SILAC) (Zhang et al. 2012a) and combinatorial peptide array screening methods have been developed to identify protein interactomes and substrates for phosphatases and kinases (Vlad et al. 2009). Previously, mass spectrometry has been used to find specific kinases involved in ABA signalling. To this end, mass spectrometry generated peptide sequence information that was used to clone the *Vicia faba* guard-cell specific ABA-regulated serine-threonine protein kinase (AAPK) (Li et al. 2000).

AAPK was reported to activate plasma membrane-localised anion channels and to be involved in ABA-induced stomatal closure. The latter event may have significant importance in agronomy as detailed information on the mechanism of action of AAPK could allow cell-specific, targeted biotechnological manipulation of crop water status (Li et al. 2000). Further, Fuji et al. (2007) reported that the Snf1-Related protein Kinase (SnRK2) from *Arabidopsis* directly phosphorylates members of the ABA-binding factor clade of bZIP TFs. The latter finding is of a great significance because at least one class of physiologically relevant targets of the SnRK2 kinases represent DNA-binding bZIP TFs that are engaged in gene activation in response to ABA (Kobayashi et al. 2005; Fuji et al. 2007; Cutler et al. 2010; Zhang et al. 2011; Kulik et al. 2011, Fujita et al. 2012). It is currently unknown if SnRK2 kinases themselves require activating kinases or whether they are prone to auto-phosphorylation and if this event itself is sufficient for activation of SnRK2 (Boudsocq et al. 2007). Nevertheless, these findings collectively indicate that bZIP TFs could serve as subjects for genetic engineering of drought tolerance and that approaches using genomics tools could be used to directly modify the cognate TFs that control signal transduction pathways during drought.

12.4 Molecular Structure, Post-Translational Modification and Transcriptional Regulation of Plant Stress Responses to Drought and Other Abiotic Stresses by ABA-Inducible Members of the bZIP, HD-Zip and WRKY Families of Transcription Factors

Transcriptional regulation of a variety of TFs in response to abiotic stress factors has been extensively reviewed (e. g. Jewell et al. 2010; Hung et al. 2012). For these reasons the remainder of this article will focus on the basic leucine zipper (bZIP), homeodomain-leucine zipper (HD-Zip) and WRKY families of TFs that are transcriptionally induced by ABA (Zhang et al. 2012b). We will also discuss how manipulation of molecular properties of these three classes of TFs could enhance drought tolerance in plants.

It appears that bZIP, HD-Zip and WRKY TFs are integrated within the ABA-regulated developmental networks, enabling environmental drought stress to influence the genetically pre-programmed developmental progression and provide plant protection from severe stresses (Nakashima et al. 2009; Harris et al. 2011; Rushton et al. 2012; Wei et al., 2012b). Despite a large number of genes that are down-regulated by ABA, the mechanisms of repression are less characterised than those of activation. A number of bZIP, HD-Zip and WRKY TFs are transcriptionally regulated by drought through the ABA-dependent signal transduction pathway (Fig. 12.2). The bZIP, HD-Zip and WRKY TFs are believed to act towards the end of the ABA

transduction pathways, whereby they switch on genes that are involved in the protection of cell integrity and regulate plant development during drought and salt stress (Figs. 12.1 and 12.2).

Nevertheless, the precise functional information on developmental and stress-responsive bZIP, HD-Zip and WRKY TF networks and cascades is incomplete. Limited data is available, mostly from *Arabidopsis* studies (Valdés et al. 2012), and it is not known if these pathways have already been optimised, in particular in crop plants (Qin et al. 2011). To address the latter concern, structural and functional relationships of ABA and drought-responsive bZIP, HD-Zip and WRKY factors, their interactions with other TFs, and with their modifying enzymes and target DNA within their cognate promoter regions should be investigated (Zhao et al. 2011). It is important to perform these investigations in crop plants, as the knowledge gained could be directly used for practical applications through genetic engineering.

To this end, it is important to discuss the roles of regulatory DNA elements (*cis*-elements) within the promoter regions of the downstream stress-related genes that are up- or down-regulated by TFs during abiotic stresses (Fig. 12.2). The *cis*-element ABREs (e. g. ACGTGGC), GCC-box (GCCGCC), CRT/DRE (A/GCCGAC), MYCR (CACATC) and MYBR (TGGTTAG) motifs have been studied intensily over the past ten years in connection with abiotic stress-related transduction pathways (e. g. Abe et al. 1997; Siberil et al. 2001). The W-box (T/CTGACT/G) *cis*-element was found to be recognised by WRKY TFs and in some cases, this *cis*-element was also demonstrated to be involved in the response to ABA and abiotic stresses (Duan et al. 2007). The ABA response generally depends on the presence of at least one ABA response element (ABRE) and either a coupling element or additional ABRE (Cutler et al. 2010). However, the promoters of some ABA-inducible genes do not contain ABREs. For example MYC and MYB TFs regulate the ABA-inducible *RD22* promoter which lacks ABREs but nevertheless can be strongly activated by ABA and abiotic stresses (Abe et al. 2003). Further, a variety of other regulatory *cis*-elements that mediate ABA-induced changes in gene expression have been identified by both molecular biology and genomics approaches (Nakashima et al. 2009; Cutler et al. 2010). The most intensely investigated regulatory sequences are the G-box ABREs with the ACGTG core element that are recognised by bZIP TFs (Yoshida et al. 2009). In addition to bZIP, HD-Zip and WRKY TFs various other factors such as AP2/ERF, MYB, NAC and HD-ZF have been reported to be engaged in ABA-mediated gene expression (Fig. 12.2) (Fujita et al. 2011).

12.4.1 Basic Leucine Zipper Transcription Factors

Stress-inducible bZIP TFs are one of the groups of TFs that dominate the adaptation of plants to stress conditions (Xiang et al. 2008; Cutler et al. 2010). For example, in *Arabidopsis* and rice a total of 29 (40 %) and 43 (49 %) *bZIP* genes, respectively, are regulated by abiotic stresses. Structural analyses of the bZIP proteins mainly from mammalian and insect sources indicate that bZIP factors are modular proteins. The

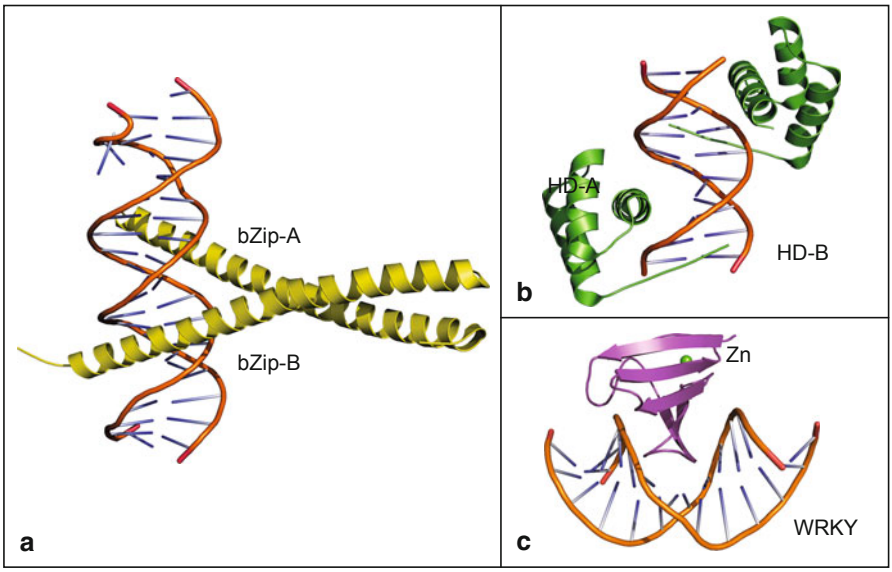

Fig. 12.3 3D structures of a bZIP dimer (A), two HD domains (B) and WRKY TFs interacting with DNA. **a**, A GCN4 bZIP dimer from yeast (bZIP-A and bZIP-B in *yellow*) illustrating the interaction of the two monomeric α-helices. The basic regions of the bZIP proteins bind AATGACTCAT/TACTGAGTA, centered on a GC base pair (in *bold*) (Ellenberger et al. 1992; PDB accession 1YSA). **b**, A tandem arrangement of two HDs (HD-A and HD-B in *green*) in the Eve HD-DNA complex from Drosophila, in which each HD represents a helix-loop-helix-turn-helix structure, and where both HDs bind AATTAAATTC (Hirsch and Aggarwal 1995; PDB accession 1JGG). Panels A and B were printed with permissions from New Phytologist 190:823–837 (Harris et al. 2012). **c**, A model of AtWRKY1-C (in *magenta*) illustrating how its antiparallel β-sheet interacts with DNA. Zinc ion is shown in green (Duan et al. 2007; PDB accession 1ODH). DNA is shown in a form of a cartoon in atomic colours in all three panels

proximal region of bZIP factors contains a canonical basic region involved in DNA binding and nuclear import, while the distal part of bZIP factors carries a leucine dimerisation motif which is known as a leucine zipper. Experimental evidence suggests that the basic regions of bZIP TFs (bZIP-bRs) are necessary and sufficient for DNA binding and specificity (Das et al. 2012). Further, it has been suggested through bioinformatic predictions and spectroscopic studies of bZIP-bRs that unbound monomeric bZIP-bRs were uniformly disordered as isolated domains (Das et al. 2012). The bZIP factors bind DNA as dimers, formed by the interaction of two α-helical stretches, consisting of seven amino acid residues *per* DNA turn (Fig. 12.3, left panel). It has also recently been suggested that beyond oligomerisation function, leucine zippers of certain bZIPs, for example GCN4 or c-Jun, catalyse degradation of RNA (Nikolaev and Pervushin 2012).

On the basis of biophysical characteristics of amino acid residues forming leucine zippers, the rules for predicting homo-, hetero- or quasi homo-dimerisation patterns

of the bZIP TFs have been outlined (Schutze et al. 2008). For example, the dimerization specificity of OsbZIP, AtbZIP and HsbZIP TFs was predicted (Deppmann et al. 2004; Nijhawan et al. 2008). It was observed that leucine zipper length was variable, ranging from two to nine heptads in the case of the OsbZIP proteins. Detailed analysis of the type of amino acid residues present in OsbZIP was carried out and a cross-comparison was made with those of *Arabidopsis* and human bZIP TFs. It was concluded that OsbZIP could behave as homo-dimerising leucine zipper (Acharya et al. 2002; Nijhawan et al. 2008). However, experimental evidence points towards a much higher complexity of bZIP dimerisation, including formation of unpredicted hetero-dimers. 3D structures on the plant bZIP TF proteins have not yet been elucidated, although the first attempts of crystallising OsAREB8 bZIP from rice have just appeared (Miyazono et al. 2012). In this context, the full description of the atomic structure of the yeast bZIP TF dimer is available (Ellenberger et al. 1992; Protein Data Bank accession 1YSA). Figure 12.3 (left panel) illustrates the 3D structure of two monomers bZIP-A and bZIP-B, which through the two monomeric α-helices interact with the pseudo-palindrome AATGACTCAT/TACTGAGTA sequence (Ellenberger et al. 1992).

The lack of adequate precision in predicting the dimerisation patterns might be due to the effects of post-translational modifications, or the presence of other interaction motifs that are located outside of the leucine zipper regions. As for post-translational modifications involved in drought responses, the most frequent one includes phosphorylation (Fujita et al. 2009; Kirchler et al. 2010; Sirichandra et al. 2010; Yang et al. 2010), although the farnesylation, SUMOylation and poly(ADP-ribosyl)ation events were also reported. In addition, synergistic interactions were detected between specific members of distinct TF families, for example those of bZIP and DREB (Drought-Response Elements Binding proteins) TFs (Lee et al. 2010), as well as between HD-ZF and NAC factors that are known to play roles in abiotic stress responses (Tran et al. 2007). It is important to point out that DREB TFs belong to the ABA-independent dehydration-responsive TFs and not to the ABA-dependent dehydration-responsive TFs (Fig. 12.2) (Yang et al. 2010), hence TFs from both pathways can interact with each other.

Although expression of bZIP TFs is often transcriptionally regulated, the main regulatory mechanism responsible for their activity, stability and nuclear translocation is believed to be mediated by protein post-translational modifications, in particular by reversible phosphorylation events (Schutze et al. 2008). Here, protein kinases and/or phosphatases have been implicated to function through the ABA-mediated signal transduction or stress-responsive pathways. For example, as alluded to above, bZIP factors are known to be the downstream substrates for the SnRK kinase family (Kline et al. 2010). Upon ABA application or an abiotic stress treatment, bZIP TFs change their status through serine, threonine or tyrosine phosphorylation. These events lead to the introduction of a negative charge on a target residue meaning that bZIP conformations rearrange. For example, the rice TRAB1 is rapidly phosphorylated at serine 102 and this phosphorylation is crucial to the ABA-induced transcriptional activity *in vivo* (Kagaya et al. 2002). On the other hand, de-phosphorylation of bZIP TFs occurs via phosphatases that could also modulate bZIP oligomerisation or DNA

binding (Fujita et al. 2009). Thus, the phosphorylation and/or dephosphorylation events could change the transcriptional potential of bZIP TFs by several discrete mechanisms, such as nucleo-cytoplasmic distribution, stability, activation and DNA binding capacity (Fujita et al. 2009; Kirchler et al. 2010).

12.4.2 Homeodomain-Leucine Zipper Transcription Factors

Higher plants contain four families (classes I to IV) of HD-Zip proteins. Individual families can be distinguished by conservation of HD-Zip domains and by preservation of other motifs in the sequences, and also by specific intron and exon distributions (Ariel et al. 2007). When the phylogeny of each family were analysed in the rice and *Arabidopsis* genomes, it became evident that specific sub-groups of genes associated within each family.

Many HD-Zip sequences were isolated from plants, which contain a 60-amino-acid-long homeobox domain (HD) that consists of three characteristic α-helices (Fig. 12.3, right upper panel). The α-helix 3, considered the recognition α-helix, is the most conserved across HD proteins and is responsible for the specificity of binding between HD-Zip TF and DNA. The α-helix 3 typically lies within the major groove of DNA next to the core sequence ATTA (or TAAT), bound by most HD proteins. On the other hand, a leucine zipper that forms a second α-helical structural element is comprised of seven amino acid (heptad) repeats. The residues of the heptad are designated as a_n, $b_n \ldots g_n$ (n being the number of the heptad), where residue d is leucine. During dimerisation through the monomeric α-helices, the HD-Zip structure forms a coiled coil (Fig. 12.3, right upper panel). The leucine zipper of HD-Zip proteins is immediately downstream of the HD and enables dimerisation of HD-Zip factors, which is a prerequisite for DNA binding. This leads to a tandem binding arrangement of HDs and the advantage is that it provides the opportunity for a larger and more specific DNA sequence read-out. Many HD proteins bind strongly as monomers to DNA, but HD-Zip proteins possess a very weak affinity for DNA as monomers and require dimerisation for efficient DNA binding. The 3D information on HD-Zip proteins is currently unavailable, although the crystal structures of HD domains from insects and mammals are available.

When investigating the specificity of the HD-ZipI and HD-ZipII *cis*-element binding interactions *in vitro*, the results generally yield binding sequence 1 [CAAT(A/T)ATTG] or binding sequence 2 [CAAT(C/G)ATTG]. It is considered that these two sequences are composed of two overlapping, yet different HD related *cis*-elements composed of 5'-T**A**ATTG-3' and 3'-GTTA**T**T-5', or 5'-T**G**ATTG-3' and 3'-GTTA**C**T-5', respectively (the difference in the central nucleotide is highlighted in bold). The consequence of this is that, each HD of a dimer could interact with a different sequence, depending on the orientation of the dimer relative to the DNA. Also, only one HD of any dimer makes specific contacts with the central nucleotide and each monomer has specific preferences for the orientation of a *cis*-element (Tron et al. 2005). The key residues of the HD domains, which contribute to binding at the

central nucleotides, are conserved in the members of each family (Sessa et al. 1997). However, it is now evident, that these conservations are not solely responsible for distinctions between binding sites. There are many examples of members from each family that bind to both sequences *in vitro*, albeit with varying degrees of efficacy.

In this context, establishing the promoters that are directly regulated by HD-ZipI and HD-ZipII TFs will enable the precise identification of target *cis*-elements. This analysis can be achieved through: (i) identification of genes, which are coordinately regulated with HD-ZipI and HD-ZipII transcripts in different plant tissues or under the specific induction conditions, or (ii) identification of potential target genes in transgenic plants through a constitutive over-expression of HD-ZipI or HD-ZipII TFs. Once the potential downstream genes are revealed, their promoter regions can be investigated for the presence of specific *cis*-elements, functionality of which could be validated by using single base pair mutations and deletions in combination with the promoter activation studies in transient expression assays. Surprisingly, few downstream genes encoding HD-ZipI and HD-ZipII TFs have so far been described. This observation suggests that cooperation of HD-Zip TFs with other TFs may be critical to ensure that they bind DNA strongly, although other factors might be necessary for transcription to commence.

12.4.3 WRKY Transcription Factors

WRKY TFs are known to play roles during plant growth and development (Wei et al. 2012a) and plant responses to biotic stress (Pandey and Somssich 2009). These factors have recently been found to be switched on in both the ABA-dependent and ABA-independent signal transduction pathways (cf. Fig. 12.1) in response to drought, high temperature, high ozone concentrations and salt stress (Duan et al. 2007; Rushton et al. 2010; Chen et al. 2012; Rushton et al. 2012). It was demonstrated that some WRKY TFs are upstream regulators of ABA- and stress-responsive genes such as *ABF2, ABF4, ABI4, ABI5, MYB2, DREB1a, DREB2a, RAB18, RD29A* and *COR47* (Rushton et al. 2010, 2012). Ten genes encoding wheat and maize WRKY TFs have recently been isolated using yeast one-hybrid (Y1H) screen of cDNA libraries prepared from roots of drought tolerant cultivars subjected to drought (Cormack et al. 2002; Lopato et al. unpublished). Amongst these genes were identified homologues of *Arabidopsis* WRKY TFs that were shown to be involved in ABA responses (Chen et al. 2010), as well as several other genes of unknown functions (Pyvovarenko and Lopato 2011). Finally, specific *cis*-elements and target genes of WRKY TFs have been described (Cai et al. 2008; Niu et al. 2012; Rushton et al. 2012).

From the structural point of view, WRKY TFs are defined by the conserved WRKYGQK motif that is followed by the zinc-finger-like motifs C_2H_2 or C_2HC (C and H indicate cysteine and histidine amino acid residues, respectively) (Wei et al. 2012a). In *Arabidopsis*, 72 *WRKY* genes have been sub-classified in three groups based on the number of WRKY domains and patterns of zinc-finger-like motifs. WRKY TFs that belong to the first group contain two WRKY domains,

whereas the two other groups have only one domain. Wei et al. (2012a) have recently identified 132 WRKY members in maize and established that WRKY domain gain and loss have been a divergent force for expansion of the WRKY family. The authors concluded that monocotyledonous plants have developed larger families of WRKY TFs than dicotyledonous plants, and that rapid duplication of the WRKY family members may have helped to increase adaptability and to establish signal transduction pathway networks in monocotyledonous plants to overcome environmental adversity (Wei et al. 2012a).

The 3D structure of the AtWRKY C-terminal domain elucidated by X-ray crystallography (Duan et al. 2007) showed that this domain folds into a globular structure with five β-strands, forming an antiparallel β-sheet (Fig. 12.3, right bottom panel). A zinc-binding site was located in the AtWRKY structure at one end of the β-sheet, between the β4 and β5 strands. This zinc-binding site was crucial to shaping the overall structure and to underlie stability of the WRKY proteins. Further, the 3D structure of the AtWRKY C-terminal domain revealed that the DNA-binding residues were positioned at the β2 and β3 strands (Fig. 12.3, right bottom panel). In summary, the 3D structure of the AtWRKY C-terminal domain suggested a potential mechanism of transcriptional control by this family of TFs and how the signal transduction events could be mediated (Duan et al. 2007).

12.4.4 Transcription Factors are Involved in Gene Activation in the Form of Protein Complexes

Plant traits that are associated with stress tolerance are often multi-genic and thus problematic to manipulate and engineer. Research has shown that products of the drought-inducible genes could be identified through a variety of genomics approaches, such as sequencing of stress-inducible cDNA pools, expression analysis using RNA, cDNA and GeneChip micro- and oligo-arrays (Rabbani et al. 2003; Shinozaki and Yamaguchi-Shinozaki 2003, 2007), proteomics (Valliyodan and Nguyen 2006; Wani and Udgaonkar 2012) and metabolomics (Chaves et al. 2009; Widodo et al. 2009; Bowne et al. 2011) investigations, bioinformatics predictions (Goyal et al. 2005), analyses of spatial and temporal expression profiles of genes (Mitsuda and Masaru Ohme-Takagi 2009), and protein-DNA and protein-protein interactions in yeast one-hybrid and two-hybrid systems (Lopato et al. 2006; Garcia et al.2008; Lee et al. 2010). Function of identified stress-responsive factors has been investigated using experimental (Goyal et al. 2005) as well as computational tools (Das et al 2012). The latter platforms allow us to evaluate the properties of potential variants or designer proteins *in silico* before they could be tested *in planta* (Wise et al. 2003).

Most of the attempts to use TFs for genetic engineering of plants with increased tolerance to environmental stresses include up-regulation or silencing of TFs. The manipulation of a single gene may not give sustained tolerance to abiotic stresses (Agarwal et al. 2006), since TFs do not act alone, but are part of protein-protein and protein-DNA complexes (Kumimoto et al. 2010; Joshi-Saha et al. 2011). The

most straightforward of these complexes represent the homodimer and heterodimer complexes. However, physical interactions with other TFs, co-factors and modifying enzymes also take place. To this end, it is attractive to unravel the detailed structure and function of these TF complexes, as this knowledge can be used to design TF complexes for deployment by genetic engineering. For example, the knowledge of an active repressor motif of TF would permit manipulation of this signature to convert a repressor protein into an activator of transcription. In another case, when a short EDLL motif (present in AtERF98/TDR1 factor) was added to the proximal or distal parts of AtBH2 TF (which contains a repressive EAR motif), the TF was converted from a repressor to an activator (Tiwari et al. 2012).

Furthermore, a TF variant with increased specificity of binding to a particular *cis*-element can be generated. The mutated TF gene expressed under its own promoter can be used instead of overexpressing the original gene. This approach would provide more precise spatial and temporal patterns of gene expression and thus minimise aberrant effects on the development of transgenic plants. In contrast, a non-functional TF molecule with specific affinity to a particular *cis*-element can act as a competitor of the respective functional TF, if the non-functional TF variant is expressed under its own or strong stress-inducible promoter. The non-functional TF-DNA interaction could lead to a similar effect as that introduced by the RNAi-mediated silencing in specific tissues or at a particular time. This strategy can be most useful in plants with polyploid genomes such as wheat, potato or cotton, where other methods of gene silencing may be inefficient. Finally, identification of post-translationally modified residues within TF sequences would allow modification of the molecular structure of a TF so that strong DNA-TF interaction would be achieved without the need for post-translational modification. Use of a gene encoding such a molecular variant with low constitutive or early inducible promoters could provide a survival or adaptive advantage to a transgenic plant under stress.

To date, little data has been presented *in vitro* or *in planta* that demonstrates the significance of the protein-protein interactions, formed by the variety of plant bZIP and HD-Zip dimerisation partners, with regards to protein-DNA interactions (Lee et al. 2010). The mechanism by which bZIP, HD-Zip and WRKY TFs exert their control over plant development has not been fully described, although at least some members of bZIP, HD-Zip and WRKY families are believed to bind common *cis*-elements specific for each family (Frank et al. 1998; Lopato et al. 2006; Zhou et al. 2008). It was also shown *in vitro* that dimerisation is a prerequisite for DNA binding of the bZIP and HD-Zip families of TFs (e. g. Zhou et al. 2008). These observations suggest that large networks may potentially exist, where different dimerisation partners confer different transcriptional configurations and that members of the same family of TFs may compete for the same *cis*-element. This complexity of protein-protein and protein-DNA interactions increases the potential for integrating environmental and endogenous signals that regulate fluxes in signalling pathways. Although, such homo- and hetero-oligomerisation of TFs and competition for the same *cis*-element take place in each specific case, remains to be demonstrated. Finally, all three possibilities might depend on a type of post-translational modification of TFs. Revealing

the molecular mechanisms of binding of TFs to their individual *cis*-elements to activate or repress transcription will be of value in agricultural biotechnology.

The mechanisms discussed above could be analysed *in silico*, which would allow assessment of the significance of each potential phosphorylation and/or other site of post-translational modification on bZIP, HD-Zip and WRKY TFs. This knowledge could be used for site-directed mutagenesis to generate variant proteins with desirable features. The newly designed TF protein variants could then be again tested *in silico*, to see if modified residues mediate direct contacts with *cis*-elements *in vitro* through the DNA-binding domains of TFs. These interactions could also be evaluated by electrophoretic mobility shift assays or *in vivo* using Y1H assays. An obvious choice would be to substitute serine, threonine or tyrosine residues in TFs by aspartic or glutamic acids, to substitute for the negative charge imposed by phosphate groups. Finally, functionality of such interactions could be tested in transient expression assays in plant cells or protoplasts by using promoters of downstream genes coupled to a reporter gene as a prelude to stable transformation.

To illustrate the feasibility of studying sequence motifs of TFs, comparative studies with OsbZIP46 and OsbZIP23, which are members of the 3rd and 2nd subfamilies of bZIP TFs, showed promising results (Xiang et al. 2008; Tang et al. 2012). Five phosphorylation sites were detected in OsbZIP46 distributed on highly conserved motifs A-D, and their mutations to Asp indicated that they were all required for transactivation activity detected by Y1H assays. One D motif in the coding region of *OsbZIP46* was shown to confer a negative effect on the transactivation activity of OsbZIP46, and its deletion led to a positive transactivation effect and also increased the tolerance of rice plants to drought. It remains to be seen which part of the domain D is essential for the negative effect on transactivation activity, although a signature LxxxxLxxxL is suspected to play a role. Notably, it cannot be ruled out that the deletion of domain D may cause a conformational change that mimics the effects of post-translational modification, such as phosphorylation of OsbZIP46. Thus, the explanation of transactivation activities of OsbZIP46 and OsbZIP23 TFs detected in Y1H assays lies in the molecular structure of both TFs (Tang et al. 2012).

In summary, site-directed mutagenesis and post-translational modifications that would lead to alterations in structure and hence to modulations of the properties of TFs can potentially be used as an alternative to over-expression or silencing of TF genes in transgenic plants. Changes in primary structures of TF proteins would lead to increasing or decreasing strength of binding to DNA, modifications of their binding selectivity and/or adjustments of their protein-protein interaction properties. Consequently, these alterations would make binding of DNA to promoters and/or promoter activation independent from stress-inducible protein modifications and would lead to changes in protein-protein interaction preferences. If concurrent phosphorylation of TFs is required for DNA binding, then these protein variations can easily be tested *in vitro* through, for example, electrophoretic mobility shift assays (Kirchler et al. 2010). However, if the phosphorylation site is situated in other than DNA binding region of TFs and if phosphorylation has no effect on DNA binding properties of factors, then the influence of the introduced mutations in TFs on the activation properties of these protein variants can be tested using transient expression assays in wheat cell cultures.

12.5 Natural Variation of Transcription Factors Involved in Drought and Other Stress Responses

Understanding the ecological and evolutionary significance of natural variation in TF sequence and expression requires evaluation of the precise roles of each TF in adaptation to particular environments. These types of studies are just beginning to appear (Alonso-Blanco et al. 2009). The question then arises, have responses and interactions of TFs involved in drought and other stresses already been optimised in crop plants through the long history of selection of native variants, or are there still opportunities for a significant gain through allelic mining (Langridge et al. 2006; Alonso-Blanco et al. 2009; Qin et al. 2011)? Although information on natural variation of TFs involved in drought and other stresses is scarse, a few landmark studies have emerged.

Chen et al. (2005) have shown in *Arabidopsis* that genes showing substantial genetic variation in mRNA levels are those with functions in signal transduction, transcription and stress responses, suggesting the existence of variations in the regulatory mechanisms for these genes. Of the genes analysed, a number of differences were identified in the upstream promoter regions, several of which altered potential *cis*-regulatory elements. Chen et al. (2005) suggested that nucleotide polymorphisms in regulatory elements of genes encoding controlling factors (such as TFs) could be primary targets for natural selection and a driving force behind the evolution of *Arabidopsis* accessions.

Recently, Mondini et al. (2012) used high-resolution melting curve technology, which represents a powerful tool for detecting single nucleotide polymorphisms (SNPs) and insertion/deletion variations, in investigating the *HKT1, DREB1* and *WRKY1* genes. Notably, most of the identified SNPs were found in salt and drought tolerant durum wheat genotypes. Further, Das et al. (2012) used structural knowledge to explain how natural sequence variations of bZIP TFs lead to evolution of specificity in molecular recognition through intrinsically disordered regions of bZIP TFs. In the latter study, Das et al. (2012) used 15 yeast, mammalian and plant bZIP-basic regions (bZIP-BRs) and a combination of atomistic simulations with circular dichroism measurements to show that bZIP-bRs had quantifiable preferences for α-helical conformations in their unbound monomeric forms, which would decide their potential to bind strongly to specific sequences of DNA. These conformations varied from one bZIP-bR to another despite significant sequence similarity in their DNA binding motifs. Das et al. (2012) also used molecular dynamics simulations and revealed that intramolecular interactions between DNA binding motifs and short amino acid residue segments within bZIP-bR were the primary modulators of bZIP-bR helicities. This precise structural knowledge led to the design of novel bZIP-bRs chimeras and to formulation of relationships between bZIP-bR sequence and the degree of intrinsic disorder of these TFs.

In conclusion, we need to have access to systematic studies of natural variants of TFs involved in drought and other stress responses, as the presence of SNPs and

insertion/deletion variation leads to flexible conformations of TFs that could be important in their oligomerisation, thermodynamics and kinetics of DNA binding. The information of natural variation could be exploited through continued advancement of simulation methodologies and the synergy between experimental and computational approaches. For example, if this idea applies to bZIP TFs, then the variable helicities of unbound monomeric bZIP-bRs might influence the allosteric coupling between DNA binding and bZIP dimerisation.

12.6 Role of Genetic Engineering of Stress Tolerance in Agricultural Biotechnology is Rapidly Growing

Genetic engineering of plants for tolerance to extreme abiotic stresses could be achieved by the regulated expression of stress-inducible TFs, which in turn would regulate the expression of a large number of relevant downstream genes. For these reasons, TFs are powerful tools for genetic engineering as their overexpression can lead to the up-regulation of a whole array of genes under their control. The discovery of the plant bZIP, HD-Zip, WRKY TFs has contributed to the identification of pathways and cascades that control plasticity of plant growth and that are responsible for the modulation of plant development in response to environmental stress (Agalou et al. 2008). However, much work remains to be done. For example, the precise molecular descriptions of function of plant bZIP, HD-Zip and WRKY TFs are not available and their 'modi operandi' from the molecular level up to the ensemble molecular levels are also inaccessible. It will be important to answer the following questions: how the transcription activation and repression complexes are formed, what is the influence of post-translational modifications on the formation of such complexes, what are the mechanisms of activation and repression of the target genes, how these TFs form oligomeric assemblies, how various families of TFs are folded, what structural determinants play roles in recognition of the DNA activation complexes and ultimately, how properties of these complexes can be regulated? Answers to these questions would allow us to develop modified versions of TFs with improved DNA-binding properties. It would also allow us to generate TFs that are independent from protein kinase cascades or other upstream regulatory pathways, and finally, the knowledge would clarify the mechanisms of formation of functional complexes of TFs, particularly the mechanisms of gene repression or activation through the same *cis*-elements. Most importantly, new knowledge will help to make decisions about suitable applications of TFs to engineer plants for enhanced tolerance to abiotic stresses.

Acknowledgments We are grateful to Professor Peter Langridge for many fruitful and valuable discussions, suggestions and ideas throughout this work. We also thank Mr. John Harris for critically reading the manuscript. This work was supported by the Australian Centre for Plant Functional Genomics-II funded by the grants from the Australian Research Council and the Grains Research & Development Corporation, and by the South Australian Government. Additionally, this work

was supported by the DP120100900 and LP120100201 grants awarded by the Australian Research Council to MH, and MH and SL, respectively.

References

Abe H, Urao T, Ito T et al (2003) *Arabidopsis* AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signalling. Plant Cell 15:63–78

Acharya A, Ruvinov SB, Gal J et al (2002) A heterodimerizing leucine zipper coiled coil system for examining the specificity of a position interactions: amino acids I, V, L, N, A, and K. Biochemistry 41:14122–14131

Addicott FT, Lyon JL (1969) Physiology of abscisic acid and related substances. Annu Rev Plant Physiol 20:139–164

Agalou A, Purwantomo S, Overnaes E et al (2008) A genome-wide survey of HD-Zip genes in rice and analysis of drought-responsive family members. Plant Mol Biol 66:87–103

Agarwal PK, Agarwal P, Reddy MK, Sopory SK (2006) Role of DREB transcription factors in abiotic and biotic stress tolerance in plants. Plant Cell Rep 25:1263–1274

Alonso-Blanco C, Aarts MGM, Bentsink L et al (2009) What has natural variation taught us about plant development, physiology, and adaptation? Plant Cell 21:1877–1896

Ariel FD, Manavella PA, Dezar CA, Chan RL (2007) The true story of the HD-Zip family. Trends Plant Sci 12:419–426

Australian Government and Bureau of Meteorology (2012) http://www.bom.gov.au/

Bai L, Zhang G, Zhou Y et al (2009) Plasma membrane-associated proline-rich extensin-like receptor kinase 4, a novel regulator of Ca signalling, is required for abscisic acid responses in Arabidopsis thaliana. Plant J 60:314–327

Bowne J, Bacic A, Tester M, Roessner U (2011) Abiotic stress and metabolomics. In: Hall R (ed) Annual plant reviews 43, biology of plant metabolomics, chapter 3. Wiley-Blackwell Publishing

Boudsocq M, Laurière C (2006) Osmotic signaling in plants: multiple pathways mediated by emerging kinase families. Plant Physiol 138:1185–1194

Cai M, Qiu D, Yuan T et al (2008) Identification of novel pathogen-responsive cis-elements and their binding proteins in the promoter of *OsWRKY13*, a gene regulating rice disease resistance. Plant Cell Environ 31:86–96

Chaves MM, Flexas J, Pinheiro C (2009) Photosynthesis under drought and salt stress: regulation mechanisms from whole plant to cell. Ann Bot 103:551–560

Chen WJ, Chang SH, Hudson ME et al (2005) Contribution of transcriptional regulation to natural variations in Arabidopsis. Genome Biol 6:R32

Chen L, Song Y, Li S et al (2012) The role of WRKY transcription factors in plant abiotic stresses. Biochim Biophys Acta 1819:120–128

Chen H, Lai Z, Shi J et al (2010) Roles of Arabidopsis WRKY18, WRKY40, and WRKY60 transcription factors in plant responses to abscisic acid and abiotic stress. BMC Plant Biol 10:281

Cormack RS, Eulgem T, Rushton PJ et al (2002) Leucine zipper-containing WRKY proteins widen the spectrum of immediate early elicitor-induced WRKY transcription factors in parsley. Biochim Biophys Acta 1576:92–100

Coello P, Hirano E, Hey SJ et al (2012) Evidence that abscisic acid promotes degradation of SNF1-related protein kinase (SnRK) 1 in wheat and activation of a putative calcium-dependent SnRK2. J Exp Bot 63:913–924

Cutler SR, Rodriguez PL, Finkelstein RR, Abrams SR (2010) Abscisic acid: emergence of a core signaling network. Annu Rev Plant Biol 61:651–679

D'Angelo C, Weinl S, Batistic O et al (2006) Alternative complex formation of the Ca-regulated protein kinase CIPK1 controls abscisic acid-dependent and independent stress responses in Arabidopsis. Plant J 48:857–872

Das RK, Crick SL, Pappu RV (2012) N-Terminal segments modulate the α-helical propensities of the intrinsically disordered basic regions of bZIP proteins. J Mol Biol 416:287–299

Dekkers B, Schuurmans J, Smeekens S (2008) Interaction between sugar and abscisic acid signalling during early seedling development in Arabidopsis. Plant Mol Biol 67:151–167

Deppmann CD, Acharya A, Rishi V et al (2004) Dimerization specificity of all 67 B-ZIP motifs in *Arabidopsis thaliana*: a comparison to *Homo sapiens* B-ZIP motifs. Nucleic Acids Res 32:3435–3445

Duan M-R, Nan J, Liang Y-H et al (2007) DNA binding mechanism revealed by high resolution crystal structure of *Arabidopsis thaliana* WRKY1 protein. Nucleic Acids Res 35:1145–1154

Ellenberger TE, Brandl CJ, Struhl K, Harrison SC (1992) The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted α-helices: crystal structure of the protein-DNA complex. Cell 71:1223–1237

Frank W, Phillips J, Salamini F, Bartels D (1998) Two dehydration-inducible transcripts from the resurrection plant *Craterostigma plantagineum* encode interacting homeodomain-leucine zipper proteins. Plant J 15:413–421

Fuji H, Verslues PE, Zhu JK (2007) Identification of two protein kinases required for abscisic acid regulation of seed germination, root growth, and gene expression in Arabidopsis. Plant Cell 19:485–494

Fujita Y, Nakashima K, Yoshida T et al (2009) Three SnRK2 protein kinases are the main positive regulators of abscisic acid signaling in response to water stress in Arabidopsis. Plant Cell Physiol 50:2123–2132

Fujita Y, Fujita M, Shinozaki K, Yamaguchi-Shinozaki K (2011) ABA-mediated transcriptional regulation in response to osmotic stress in plants. J Plant Res 124:509–525

Fujita Y, Yoshida T, Yamaguchi-Shinozaki K (2012) Pivotal role of the AREB/ABF-SnRK2 pathway in ABRE-mediated transcription in response to osmotic stress in plants. Physiol Plant 147:15–27

Goyal K, Walton LJ, Tunnacliffe A (2005) LEA proteins prevent protein aggregation due to water stress. Biochem J 388:151–157

Harmon AC (2003) Calcium-regulated protein kinases of plants. Gravit Space Biol Bull 16:83–90

Harris J, Hrmova M, Lopato S, Langridge P (2011) Modulation of plant growth by HD-Zip class I and II transcription factors in response to environmental stimuli. Tansley review. New Phytol 190:823–837

Hirayama T, Shinozaki K (2010) Research on plant abiotic stress responses in the post-genome era: past, present and future. Plant J 61:1041–1052

Hirsch JA, Aggarwal AK (1995) Structure of even skipped homeodomain complexed to AT rich DNA: new perspectives on homeodomain specificity. EMBO J 14:6280–6291

Hong Y, Zhang W, Wang X (2010) Phospholipase D and phosphatidic acid signalling in plant response to drought and salinity. Plant Cell Environ 33:627–635

Huang GT, Ma SL, Bai LP et al (2012) Signal transduction during cold, salt, and drought stresses in plants. Mol Biol Rep 39:969–987

Intergovernmental Panel on Climate Change (IPCC) (2007) Climate change 2007-the physical science basis. Contribution of working group I to the fourth assessment report of the IPCC. Cambridge University Press, UK

Jain M, Tyagi AK, Khurana JP (2008) Differential gene expression of rice two-component signaling elements during reproductive development and regulation by abiotic stress. Funct Integr Genomics 8:175–180

Jewell MC, Campbell BC, Godwin ID (2010) Transgenic plants for abiotic stress resistance. In: Kole C, Michler CH, Abbott AG, Hall TC (eds) Transgenic crop plants chapter 2. Springer-Verlag Berlin Heidelberg, New York, pp 67–132

Joshi-Saha A, Valon C, Leung J (2011) Abscisic acid signal off the STARting block. Mol Plant 4:562–580

Kagaya Y, Hobo T, Murata M et al (2002) Abscisic acid-induced transcription is mediated by phosphorylation of an abscisic acid response element binding factor, TRAB1. Plant Cell 14:3177–3189

Kirchler T, Briesemeister S, Singer M et al (2010) The role of phosphorylatable serine residues in the DNA-binding domain of Arabidopsis bZIP transcription factors. Eur J Cell Biol 89:175–183

Kline KG, Barrett-Wilt GA, Sussman MR (2010) *In planta* changes in protein phosphorylation induced by the plant hormone abscisic acid. Proc Natl Acad Sci USA 107:15986–15991

Kumar SV, Wigge PA (2010) H2A.Z-containing nucleosomes mediate the thermosensory response in Arabidopsis. Cell 140:136–147

Kobayashi Y, Murata M, Minami H et al (2005) Abscisic acid-activated SNRK2 protein kinases function in the gene-regulation pathway of ABA signal transduction by phosphorylating ABA response element-binding factors. Plant J 44:939–949

Kulik A, Wawer I, Krzywińska E et al (2011) SnRK2 protein kinases-key regulators of plant response to abiotic stresses. OMICS 15:859–872

Kumimoto RW, Zhang Y, Siefers N, Holt BF (2010) NF-YC3, NF-YC4 and NF-YC9 are required for CONSTANS-mediated, photoperiod-dependent flowering in *Arabidopsis thaliana*. Plant J 63:379–391

Langridge P, Paltridge N, Fincher G (2006) Functional genomics of abiotic stress tolerance in cereals. Brief Funct Genomic Proteomic 4:343–354

Lee SJ, Kang JY, Park HJ et al (2010) DREB2C interacts with ABF2, a bZIP protein regulating abscisic acid-responsive gene expression, and its overexpression affects abscisic acid sensitivity. Plant Physiol 153:716–727

Li J, Wang XQ, Watson MB, Assmann SM (2000) Regulation of abscisic acid-induced stomatal closure and anion channels by guard cell AAPK kinase. Science 287:300–303

Lopato S, Bazanova N, Morran S et al (2006) Isolation of plant transcription factors using a modified yeast one-hybrid system. Plant Meth 2:3

Lopez-Perez L, Martinez-Ballesta MC, Maurel C, Carvajal M (2009) Changes in plasma membrane lipids, aquaporins and proton pump of broccoli roots, as an adaptation mechanisms to salinity. Phytochemistry 70:492–500

Luan S (2003) Protein phosphatases in plants. Annu Rev Plant Biol 54:63–92

Maali Amiri R, Yureva NO, Shimshilashvili KR et al (2010) Expression of acyl-lipid D12-desaturase gene in prokaryotic and eukaryotic cells and its effect on cold stress tolerance of potato. J Integr Plant Biol 52:289–297

Matsui A, Ishida J, Morosawa T et al (2008) Arabidopsis transcriptome analysis under drought, cold, high-salinity and ABA treatment conditions using a tiling array. Plant Cell Physiol 49:1135–1149

Melcher K, Ng LM, Zhou XE et al (2009) A gate-latch-lock mechanism for hormone signalling by abscisic acid receptors. Nature 462:602–608

Mitsuda N, Ohme-Takagi M (2009) Functional analysis of transcription factors in Arabidopsis. Plant Cell Physiol 50:1232–1248

Miyazono KI, Koura T, Kubota K et al (2012) Purification, crystallization and preliminary X-ray analysis of OsAREB8 from rice, a member of the AREB/ABF family of bZIP transcription factors, in complex with its cognate DNA. Acta Crystallogr Sect F Struct Biol Cryst Commun F68:491–494

Mondini L, Nachit N, Porceddu E, Pagnotta MA (2012) Identification of SNP mutations in *DREB1*, *HKT1*, and *WRKY1* genes involved in drought and salt stress tolerance in durum wheat (*Triticum turgidum* L. var *durum*). Omics 16:178–187

Nakashima K, Ito Y, Yamaguchi-Shinozaki K (2009) Transcriptional regulatory networks in response to abiotic stresses in Arabidopsis and grasses. Plant Physiol 149:88–95

Nambara E, Marion-Poll A (2005) Abscisic acid biosynthesis and catabolism. Annu Rev Plant Biol 56:165–185

Nemhauser JL, Hong F, Chory J (2006) Different plant hormones regulate similar processes through largely nonoverlapping transcriptional responses. Cell 126:467–475

Nikolaev Y, Pervushin K (2012) Structural basis of RNA binding by leucine zipper GCN4. Protein Sci 21:667–676

Nijhawan A, Jain M, Tyagi AK, Khurana JP (2008) Genomic survey and gene expression analysis of the basic leucine zipper transcription factor family in rice. Plant Physiol 146:333–350

Niu CF, Wei W, Zhou QY et al (2012) Wheat WRKY genes *TaWRKY2* and *TaWRKY19* regulate abiotic stress tolerance in transgenic Arabidopsis plants. Plant Cell Environ 35:1156–1170

Pandey SP, Somssich IE (2009) The role of WRKY transcription factors in plant immunity. Plant Physiol 150:1648–1655

Prasad PVV, Staggenborg SA, Ristic Z (2008) Impacts of drought and/or heat stress on physiological, developmental, growth, and yield processes of crop plants. In: Ahuja LR, Reddy VR, Saseendran SA, Yu Q (eds) Advances in agricultural systems modeling series 1. ASA, CSSA, SSSA, Madison, USA, pp 301–355

Pyvovarenko T, Lopato S (2011) Isolation of plant transcription factors using a yeast one-hybrid system. Methods Mol Biol 754:45–66

Qin F, Shinozaki K, Yamaguchi-Shinozaki K (2011) Achievements and challenges in understanding plant abiotic stress responses and tolerance. Plant Cell Physiol 52:1569–1582

Rabbani MA, Maruyama K, Abe H et al (2003) Monitoring expression profiles of rice genes under cold, drought, and high-salinity stresses and abscisic acid application using cDNA microarray and RNA gel-blot analyses. Plant Physiol 133:1755–1767

Rushton PJ, Somssich IE, Ringler P, Shen QJ (2010) WRKY transcription factors. Trends Plant Sci 15:247–258

Rushton DL, Tripathi P, Rabara RC et al (2012) WRKY transcription factors: key components in abscisic acid signalling. Plant Biotechnol J 10:2–11

Saidi Y, Peter M, Finka A et al (2010) Membrane lipid composition affects plant heat sensing and modulates $Ca2^{+}$-dependent heat shock response. Plant Sig Behavior 5:1530–1533

Saidi Y, Finka A, Goloubinoff P (2011) Heat perception and signalling in plants: a tortuous path to thermotolerance. New Phytol 190:556–565

Sessa G, Morelli G, Ruberti I (1993) The ATHB-1 and –2 HD-Zip domains homodimerize forming complexes of different DNA-binding specificities. EMBO J 12:3507–3517

Sessa G, Morelli G, Ruberti I (1997) DNA-binding specificity of the homeodomain leucine zipper domain. J Mol Biol 274:303–309

Schutze K, Harter K, Chaban C (2008) Post-translational regulation of plant bZIP factors. Trends Plant Sci 13:247–255

Shinozaki K, Yamaguchi-Shinozaki K, Seki M (2003) Regulatory network of gene expression in the drought and cold stress responses. Curr Opin Plant Biol 6:410–417

Shinozaki K, Yamaguchi-Shinozaki K (2007) Gene networks involved in drought stress response and tolerance. J Exp Bot 58:221–227

Siberil Y, Benhamron S, Memelink J et al (2001) Catharanthus roseus G-box binding factors 1 and 2 act as repressors of strictosidine synthase gene expression in cell cultures. Plant Mol Biol 45:477–488

Sirichandra C, Davanture M, Turk BE et al (2010) The Arabidopsis ABA-activated kinase OST1 phosphorylates the bZIP transcription factor ABF3 and creates a 14–3-3 binding site involved in its turnover. PLoS One 5:e13935

Tang N, Zhang H, Li X et al (2012) Constitutive activation of transcription factor OsbZIP46 improves drought tolerance in rice. Plant Physiol 158:1755–1768

The Food and Agriculture Organization of the United Nations (2002) http://www.fao.org/

Tiwari SB, Belachew A, Ma SF et al (2012) The EDLL motif: a potent plant transcriptional activation domain from AP2/ERF transcription factors. Plant J 70:855–865

Tolleter D, Hincha DK, Macherel D (2010) A mitochondrial late embryogenesis abundant protein stabilizes model membranes in the dry state. Biochim Biophys Acta 1798:1926–1933

Tran L-S, Nakashima K, Sakuma Y et al (2007) Co-expression of the stress-inducible zinc finger homeodomain ZFHD1 and NAC transcription factors enhances expression of the *ERD1* gene in Arabidopsis. Plant J 49:46–63

Tron AE, Comelli RN, Gonzalez DH (2005) Structure of homeodomain-leucine zipper/DNA complexes studied using hydroxyl radical cleavage of DNA and methylation interference. BioChemistry 44:16796–16803

Tuberosa R, Salvi S (2006) Genomics-based approaches to improve drought tolerance of crops. Trends Plant Sci 11:405–412

Tunnacliffe A, Wise MJ (2007) The continuing conundrum of the LEA proteins. Naturwissenschaften 94:791–812

Umezawa T, Fujita M, Fujita Y et al (2006) Engineering drought tolerance in plants: discovering and tailoring genes unlock the future. Curr Opin Biotech 17:113–122

Valdés AE, Overnäs E, Johansson H et al (2012) The homeodomain-leucine zipper (HD-Zip) class I transcription factors ATHB7 and ATHB12 modulate abscisic acid signalling by regulating protein phosphatase 2C and abscisic acid receptor gene activities. Plant Mol Biol 80:405–418

Valliyodan B, HT Nguyen (2006) Understanding regulatory networks and engineering for enhanced drought tolerance in plants. Curr Opin Plant Biol 9:1–7

Vlad F, Droillard M-J, Valot B et al (2010) Phospho-site mapping, genetic and *in planta* activation studies reveal key aspects of the different phosphorylation mechanisms involved in activation of SnRK2s. Plant J 21:3170–3184

Wan J, Griffiths R, Ying J et al (2009) Development of drought-tolerant canola (*Brassica napus L.*) through genetic modulation of ABA-mediated stomatal responses. Crop Sci 49:1539–1554

Wang J, Zhou J, Zhang B et al (2011) Genome-wide expansion and expression divergence of the basic leucine zipper transcription factors in higher plants with an emphasis on sorghum. J Integr Plant Biol 53:212–231

Wani AH, Udgaonkar JB (2012) Mass spectrometry studies of protein folding. Curr Sci 102:1–21

Wei K-F, Chen J, Chen Y-F et al (2012a) Molecular phylogenetic and expression analysis of the complete WRKY transcription family in maize. DNA Res 19:153–164

Wei K, Chen J, Wang Y et al (2012b) Genome-wide analysis of bZIP-encoding genes in maize. DNA Res 19:463–476

Wieczorek A (2003) Use of biotechnology in agriculture-benefits and risks. Governance for industrial transformation. Proceedings of the 2003 Berlin conference on the human resources. University of Hawaii Press, USA, pp 1–6

Widodo, Patterson JH, Newbigin E et al (2009) Metabolic responses to salt stress of barley (*Hordeum vulgare* L.) cultivars, Sahara and Clipper, which differ in salinity tolerance. J Exp Bot 60:4089–4103

Wise MJ (2003) LEAping to conclusions: a computational reanalysis of late embryogenesis abundant proteins and their possible roles. BMC Bioinform 4:52

Xiang Y, Tang N, Du H et al (2008) Characterization of OsbZIP23 as a key player of the basic leucine zipper transcription factor family for conferring abscisic acid sensitivity and salinity and drought tolerance in rice. Plant Physiol 148:1938–1952

Yamaguchi-Shinozaki K, Shinozaki K (2006) Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses. Annu Rev Plant Biol 57:781–803

Yang S, Vanderbelt B, Wan J, Huang Y (2010) Narrowing down the targets: towards successful genetic engineering of drought-tolerant crops. Mol Plant 3:469–490

Yoshida T, Fujita Y, Sayama H et al (2009) AREB1, AREB2, and ABF3 are master transcription factors that cooperatively regulate ABRE-dependent ABA signaling involved in drought stress tolerance and require ABA for full activation. Plant J 61:672–685

Zeller G, Henz SR, Widmer CK et al (2009) Stress-induced changes in the *Arabidopsis thaliana* transcriptome analyzed using whole-genome tiling arrays. Plant J 58:1068–1082

Zhang H, Mao X, Jing R et al (2011) Characterization of a common wheat (*Triticum aestivum* L.) TaSnRK2.7 gene involved in abiotic stress responses. J Exp Bot 62:975–988

Zhang XX, Chan CS, Bao H et al (2012a) Nanodiscs and SILAC-based mass spectrometry to identify a membrane protein interactome. J Proteome Res 11:1454–1459

Zhang S, Haider I, Kohlen W, Jiang L, Bouwmeester H, Meijer AH, Schluepmann H, Liu CM, Ouwerkerk PB (2012b) Function of the HD-Zip I gene *Oshox22* in ABA-mediated drought and salt tolerances in rice. Plant Mol Biol 80:571–585

Zhao Y, Zhou Y, Jiang H et al (2011) Systematic analysis of sequences and expression patterns of drought-responsive members of the HD-Zip gene family in maize. PLoS One 6:e28488

Zhou QY, Tian AG, Zou HF et al (2008) Soybean WRKY-type transcription factor genes, *GmWRKY13*, *GmWRKY21*, and *GmWRKY54*, confer differential tolerance to abiotic stresses in transgenic Arabidopsis plants. Plant Biotechnol J 6:486–503

Zhou X, Yuan F, Wang M et al (2013) Molecular characterization of an ABA insensitive 5 orthologue in *Brassica oleracea*. Biochem Biophys Res Commun 430:1140–1146

Chapter 13
The Borlaug Global Rust Initiative: Reducing the Genetic Vulnerability of Wheat to Rust

Sarah Davidson Evanega, Ravi P. Singh, Ronnie Coffman and Michael O. Pumphrey

Abstract The Borlaug Global Rust Initiative (BGRI) is a collaboration of scientists from around the world inspired by Dr. Normal E. Borlaug to combat the threat of dangerous rusts of wheat. The BGRI's Durable Rust Resistance in Wheat project is aimed specifically at mitigating the threat of wheat stem rust Ug99 through an interdisciplinary approach that includes pathogen surveillance and diagnostics, gene discovery efforts, rapid breeding and deployment of new varieties of durably resistant wheat, gene stewardship, seed multiplication and delivery, and human and infrastructure capacity building. Responsible gene stewardship and genomic approaches are guiding the delivery of varieties that carry diverse race-specific resistance gene combinations and complex multigenic race non-specific "adult plant resistant" gene combinations to protect the world wheat crop against potential epidemics of wheat stem rust.

S. D. Evanega (✉)
College of Agriculture and Life Sciences, Cornell University,
251 Emerson Hall, Ithaca, NY 14853, USA
e-mail: snd2@cornell.edu

R. P. Singh
International Maize and Wheat Improvement Center (CIMMYT),
Apdo. Postal 6–641, Mexico, DF 06600, USA
e-mail: r.singh@cgiar.org

R. Coffman
College of Agriculture and Life Sciences, Cornell University,
252 Emerson Hall, Ithaca, NY 14853, USA
e-mail: wrc2@cornell.edu

M. O. Pumphrey
Washington State University, 291D Johnson Hall, Pullman, WA 99164, USA
e-mail: m.pumphrey@wsu.edu

R. Tuberosa et al. (eds.), *Genomics of Plant Genetic Resources,*
DOI 10.1007/978-94-007-7575-6_13,

13.1 Introduction

Stem rust is potentially the most devastating of the three wheat rusts, which include stem (black) rust, stripe (yellow) rust and leaf (brown) rust. Stripe rust epidemics have been more frequent and widespread in recent years and under severe epidemics yields losses are generally higher than usually recorded in large-scale production statistics because the cool and wet conditions favorable to disease development are also favorable for wheat productivity. Stem rust, in contrast, is a biological firestorm with the potential to completely devastate an otherwise healthy crop just three weeks before harvest (Herrera-Foessel et al. 2011; Singh et al. 2006) at warmer temperatures less conducive for wheat productivity. It is arguably the most feared disease of wheat on all continents where wheat is grown.

In 1998 William Wagoire, a Ugandan native and CIMMYT-trained wheat breeder was inspecting test plots in southwest Uganda's Kalengyere Highland Crop Research Centre when he first spotted the stem rust pustules that would later be named Ug99 (Ug for Uganda and 99 for the year in which the race was named). Wagoire was testing his wheat lines for stripe rust, *Puccinia striiformis*. To his surprise, he found they were infected with *Puccinia graminis* or stem rust—a disease thought to be largely under control in most of the world since the 1970s. Wagoire's finding was confirmed by South African rust pathologist Zak Pretorius who gave the stem rust race the informal name of Ug99 (Pretorius et al. 2000).

Ug99, more formally known in the scientific literature as TTKSK (Jin et al. 2007, 2008), was remarkable in that, at the time of its discovery, it was the only known race of *P. graminis* to overcome the race-specific stem rust resistance gene, *Sr31* after more than 30 years of widespread deployment, leading to a wrong sense of "durability" associated with this resistance gene. Ug99 was uniquely virulent on *Sr31* as well as most of the resistance genes of wheat origin and other important genes found in CIMMYT, European, North American and Australian wheat germplasm (Reynolds and Borlaug 2006; Jin et al. 2007). Today, the area under immediate threat along the projected migration pathway in North Africa, the Middle East and Asia (excluding China) amounts to 50 million ha of wheat—25 % of the world's wheat area and 19 % of global production—about 117 million t (Reynolds and Borlaug 2006). An epidemic in this region would have a serious impact on the 1 billion people living within this zone (Fig. 13.1). Subsequent estimates warned that over 90 % of the world's wheat varieties are at risk imposed by Ug99 (Singh et al. 2006).

13.1.1 Current Distribution of the Ug99 Race Group

In retrospect, there is some evidence that Ug99 may have been present in Kenya as early as 1993—prior to its identification in Uganda. By 2001 it was confirmed in Kenya, where it is now endemic. In 2003 it was widely reported in Ethiopia and was poised to move across the Red Sea to the Arabian Peninsula and beyond to the world's bread basket in South Asia. That jump happened in 2006 when virulence to *Sr31* was

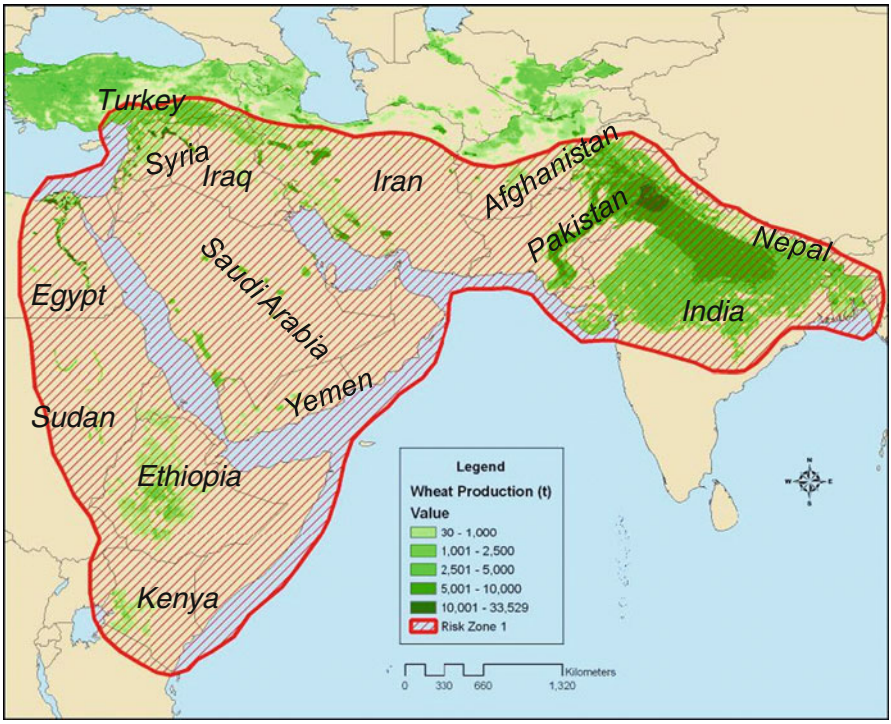

Fig. 13.1 Population and wheat production in Risk Area 1, where Ug99 is present or an immediate risk. Ug99 could affect the lives and livelihoods of more than 1 billion people in Risk Zone 1 where Ug99 is present or is at immediate risk of emerging

identified in Yemen. By 2007 it was carried into Iran, apparently by a tropical storm. To date, there are no reports that the pathogen has moved further into South Asia. However, the Ug99 family has also moved south from Kenya and is now present as far as South Africa. Other countries where Ug99 is present are Eritrea, Sudan, Tanzania, Mozambique and Zimbabwe, totaling 11 countries. The pathogen is not only moving, but it is evolving, overcoming additional stem rust resistance genes across the eastern areas of Africa. Race TTKSK, the original Ug99, with virulence to *Sr31* and *Sr21*, is now present in Tanzania, Kenya, Ethiopia, Sudan, Yemen and Iran. In addition, other virulence combinations of *Sr31, Sr21, Sr24* and *Sr36* have been identified throughout eastern and southern Africa. In total, the Ug99 family now has eight members. Surveillance data indicate that Ug99 variants with combined virulence to *Sr31* and *Sr24* are spreading rapidly. An update of the occurrence of the Ug99 family was compiled by D. Hodson (Singh et al. 2011a) and is reproduced in Fig. 13.2.

The widely deployed resistance gene *Sr31*, which Ug99 overcame, is located in the 1BL.1RS translocation from "Pektus" rye, which was originally associated with increased grain yields and resistance to all three wheat rusts and powdery mildew

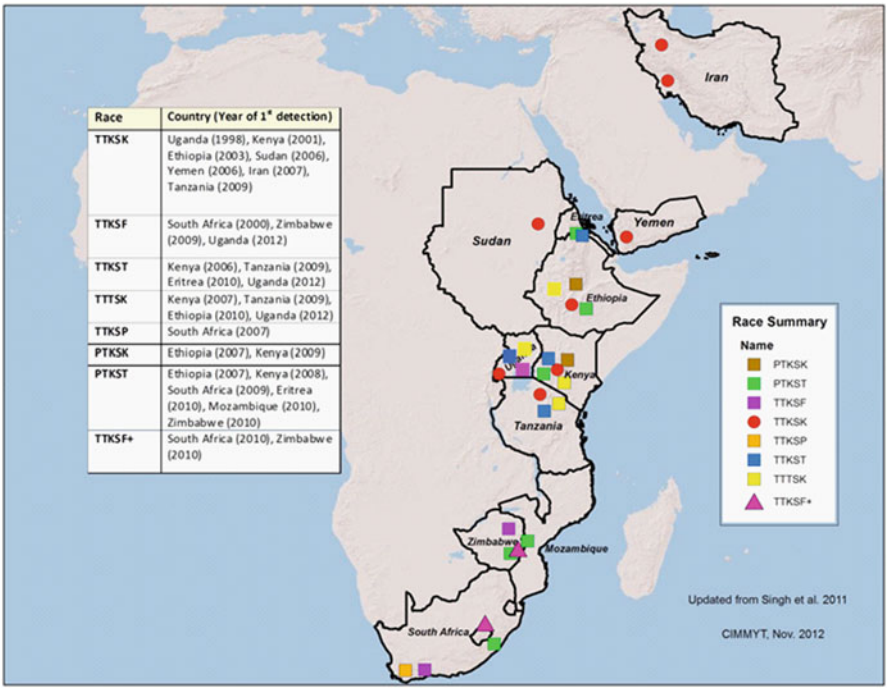

Fig. 13.2 Spread of the Ug99 race group, 1998–2012 by race. The presence of the eight-member Ug99 race group ranges from Iran to South Africa as of June 2011. Source: D. Hodson, updated from (Singh et al. 2011a)

(Singh et al. 2008). Most wheats with *Sr31* were derived from wheat x rye crosses produced in the 1930s in Germany (McIntosh et al. 1995). Since the 1980s and until the discovery of Ug99, the *Sr31*-harboring translocation was used in bread-wheat breeding programs worldwide, with the exception of those in Australia due to potential problems associated with bread making (McIntosh et al. 1995; Singh et al. 2008). The frequent use of this translocation in the CIMMYT breeding program (as much as 70 % at its peak) helped facilitate its widespread deployment, which is thought to have reduced the stem rust survival to almost non-existent levels in the 1990s in many wheat growing regions of the world—so much that increasingly constrained resources and research priorities shifted to focus on other diseases of wheat (Singh et al. 2008).

In retrospect, *Sr31* was a relatively "durable" race-specific major resistance gene, in that it fit most of Roy Johnson's criteria for durable resistance—it remained effective during its relatively prolonged and widespread use in environments favorable to the disease (Johnson 1984). *Sr31* remained effective for over 30 years before the shifty pathogen overcame it in East Africa. The East African highlands are a hot spot for the evolution of stem rust races. A combination of factors—high altitude with strong UV, year-round wheat cultivation in countries like Kenya, and the presence

of barberry, which hosts the sexual stage of stem rust, all probably facilitate the evolution of new stem rust races.

Throughout breeding history, single race-specific major genes rarely stand the test of time in the face of more rapidly evolving pathogens. This was proven true once again in 2006 when *Sr24* virulence was detected in the Ug99 lineage in Kenya after widespread deployment of the *Sr24*-carrying variety, Kenya Mwamba, which was planted on 30 % of Kenya's wheat growing area at the time. The virulence was endemic by 2007, and today *Sr24* virulence is virtually fixed in the Kenyan stem rust population (Coffman 2011).

13.2 The Borlaug Global Rust Initiative and the Durable Rust Resistance in Wheat project

Since the 1980s stem rust research declined worldwide. This was in part due to complacency afforded by the widespread deployment of *Sr31* and other major resistance genes that temporarily thwarted stem rust outbreaks in most of the world, which resulted in a shift in resources to other diseases of wheat (Singh et al. 2006). The emergence of *Sr31* virulence in East Africa concerned Nobel Peace Prize winner and famine fighter, Dr. Norman E. Borlaug, who consequently sounded the alarm on the threat of Ug99 to global food security.

The Borlaug Global Rust Initiative (BGRI), first formed as the Global Rust Initiative in 2005, started in response to Norman E. Borlaug's call to arms following the alarming discovery of Ug99 in East Africa and as an outcome of a subsequent Expert Panel Report led by CIMMYT (CIMMYT 2005). With the support of the Rockefeller Foundation and at the urging of Dr. Borlaug, CIMMYT held a Global Rust Summit in Nairobi, Kenya later that year. There, scientists, policy makers and other key stakeholders from 18 different countries saw firsthand the devastating effects of Ug99 in Kenyan wheat fields. Subsequent to the summit, the BGRI was born, open to any interested participants under a charter "with the objectives of systematically reducing the world's vulnerability to stem, yellow, and leaf rusts of wheat and advocating/facilitating the evolution of a sustainable international system to contain the threat of wheat rusts and continue the enhancements in productivity required to withstand future global threats to wheat" (2005). Among the 10 recommendations made by the Expert Panel, and upon which the BGRI is based, recommendations two and three address the need for increased and diverse genetic resources that withstand the threat of Ug99 and future races of wheat stem rust.

- Recommendation #2. Because diverse sources of resistance would be necessary for all genetic control strategies, the Panel recommends that diverse genetic resistance be identified in global wheat germplasm by testing in Kenya and Ethiopia.

- Recommendation #3. Because modern cultivars currently grown in northern Africa and Asia are susceptible to race Ug99, the Panel recommends that a breeding strategy be implemented to incorporate diverse genetic resistance to Ug99 into such germplasm before the race migrates to those areas. DNA-marker assisted selection should be utilized where feasible.

Borlaug began his career working on stem rust resistance—one of the key traits of the Green Revolution wheats—and spent his last years inspiring the world's wheat breeders and pathologists to wake up from an *Sr31*-dependent complacency and join together to mitigate the threat of stem rust to the world's vulnerable wheat varieties.

The Durable Rust Resistance in Wheat project is one of the primary sources of funds fueling stem rust research at international centers and within key national programs such as the Ethiopian Institute for Agricultural Research (EIAR) and the Kenyan Agricultural Research Institute (KARI). The project was first funded through a $ 26 million grant to Cornell University by the Bill & Melinda Gates Foundation (BMGF). In the project's second phase, an additional $ 40 million was granted to carry the work through 2015 with funds from BMGF and the United Kingdom Department for International Development (DFID). The DRRW project is carried out by over 20 institutions around the world and is managed at Cornell University. The project directors serve as the secretariat of the Borlaug Global Rust Initiative, ensuring the wheat rust community continues to work together to carry out Borlaug's legacy.

The Durable Rust Resistance in Wheat project includes activities that aim to identify and responsibly steward new genetic resources for wheat breeders that will help them mitigate the threat of Ug99 and other threatening races of stem rust. In the first and second phases of the project, these activities include, but are not limited to,

- Discovering of new resistance genes in the primary, secondary and tertiary gene pools,
- Selecting of resistant lines resulting from wild relative introgressions,
- Mapping, validation, and delivery of marker-selectable resistance genes,
- Use of genomic selection models to breed for adult plant resistance (minor gene resistance),
- Collecting, preserving and responsibly distributing genetic resources and information.

In 1993 Roy Johnson wrote "It seems reasonable to argue that genetic complexity, presented by combining many genes each of small effect is more likely to achieve durable resistance, but this should not lead to the neglect of the opportunities presented by the increasing number of major genes implicated as playing an important role in durable resistance" (Jacobs and Parlevliet 1993). The DRRW project shares this view and is pursuing multiple approaches based on both minor small effect genes that permit some small amount of retarded rust development, and large effect major genes that elicit the hypersensitive defense response.

13.2.1 Breeding for Adult Plant Resistance

The project's breeding activities are largely being carried out at CIMMYT in partnership with KARI and EIAR, where a commitment to breeding for adult plant (APR)/minor gene resistance predominates in the wheat breeding program. CIMMYT has chosen to pursue the APR gene-based breeding approach for several reasons, many of which are integral to the global service they provide of sharing their materials with the world's national programs. About half of the spring wheat varieties in Asia, Middle East, Africa and Latin America are derived from international agricultural research centers such as CIMMYT and ICARDA. Thus, the lines distributed by CIMMYT have the potential to be deployed on large areas throughout the world. Mega-varieties such as the *Sr31*-dependent PBW343, which comprises over 6 million ha of wheat in India, continue to dominate despite the availability of other varieties. In most countries, especially developing countries that lack a competitive seed sector, variety turnover is slow. These are compelling reasons for an international distributor, such as CIMMYT, to adopt an APR strategy in which wheat lines are, by necessity, bred for genetic complexity of resistance, rather than the more vulnerable approach of using race-specific "major" genes that may be easily deployed singly due to their large phenotypic effects. The danger posed by inadequate monitoring of rapidly mutating and migrating stem rust races is further support for the APR approach. Lastly, the use of minor genes based resistance by CIMMYT provides opportunities to other breeding programs to utilize race-specific resistance and further enhance diversity for resistance in farmers' fields.

Sr2 is one of the best characterized APR genes that confers resistance to stem rust. It is arguably the most important stem rust resistance gene and is closely associated with pseudo black chaff, which offers a morphological marker for breeders working with *Sr2*. *Sr2* was introduced into hexaploid wheat in the 1920s and has remained durable since (McIntosh et al. 1995). Less is known about other genes that contribute to adult plant resistance, but widely used APR genes *Lr34/Yr18* (recently designated *Sr57*) and *Lr46/Yr29* also contribute to stem rust APR (Bhavani et al. 2011) in combination with QTLs at various other genomic locations identified through biparental and association mapping (Bhavani et al. 2011; Yu et al. 2011). However, early work by Knott and revisited by Singh indicate that the accumulation of between four and five minor effect genes renders a plant near-immune (Singh et al. 2000). For CIMMYT, concentrating on an APR approach, this finding provides its breeders, and the farmers they serve an escape from re-occurring boom-and-bust cycles and ensures increased productivity in the short and long term.

Breeding efforts that targeted minor-genes based APR to Ug99 and other rust threats began in 2006 following Borlaug's call to arms and the formation of the Borlaug Global Rust Initiative (Singh et al. 2011b). Identification of some key parental resistance sources allowed incorporation of such resistance through targeted crossing and multiple-location shuttle breeding, CIMMYT spring-bread wheat breeders have released important APR lines that are already incorporated into the breeding scheme of national programs. To date, 13 CIMMYT lines that possessed high yield potential

and non-race specific resistance to stem and stripe rusts are being adopted by national programs in eight at-risk countries (see Table 13.1). In addition the completion of 1st breeding cycle from targeted crosses made in 2006 has quickly changed the resistance status of new CIMMYT germplasm being distributed through international yield trials and screening nurseries from about 90 % with inadequate resistance (40 % or higher stem rust severities) in 2006 to over 80 % with adequate resistance based on APR genes in 2011 (Table 13.2). Use of these high-yielding materials in current crossing programs promises to simplify the further accumulation of multiple minor genes with superior yields as most parents will possesses high to adequate APR to stem rust.

One emerging success story is in Ethiopia where wheat is an important staple for resource-poor farmers who grow 95 % of the country's wheat crop. There, about 3 million t of wheat is produced on 1.7 million ha. In districts where newly released CIMMYT varieties are being planted, progressive farmers are claiming 5–6 t/ha and even 7 t/ha by farmers in the Arsirobe district.Although Ug99 has been present in Ethiopia since 2003, climatic conditions have not favored severe stem rust epidemics. Wheat varieties commonly grown in Ethiopia are susceptible to the Ug99 race group and favorable climatic conditions could promote a serious stem rust epidemic. This would compound the already devastating effects of recent stripe rust epidemics in Ethiopia.

Dr. Bedada Girma Buta, a scientist from the Ethiopian Institute of Agricultural Research (EIAR), working in collaboration with CIMMYT and other EIAR colleagues, released two semi-dwarf, high yielding varieties of wheat that are resistant to Ug99 and the yellow rust race(s) that were widespread in Ethiopia in 2010. The varieties, Kakaba and Danda, are derived from CIMMYT lines Picaflor and Danphe, respectively, and are protected by non-race specific resistance based on the *Sr2* complex and other unidentified minor genes (Table 13.1). In 2010 and 2011 EIAR provided 10 farmers with enough seed to each plant a quarter-hectare with the new varieties. Kakaba, released in May 2010, was not affected by the yellow rust epidemic that devastated popular Ethiopian varieties such as Kubsa and Galama during the main season that year. Upon harvest, the farmers served as a very efficient seed distribution network, selling seeds to other farmers. At present, the area planted with Kakaba and Danda is not more than 15,000 ha, but in the face of growing disease pressure, increasing farmer awareness of the new varieties, and national and international seed multiplication efforts, the wheat area covered by these varieties should rapidly increase.

13.2.1.1 Hypersensitive Response (HR) Gene Approaches to Stem Rust Resistance

A large gene-discovery effort is integral to the DRRW project in order to generate new wheat varieties with resistance gene combinations that minimize the likelihood of progressive mutation of stem rust races. Major genes have the desirable characteristic of limiting or eliminating inoculum production during early season and

Table 13.1 The Ug99 resistant wheat varieties and promising lines derived from CIMMYT germplasm and in advanced stages of testing by country. Data provided by R. Singh

			Stem rust		Yellow rust	
Country	Variety	CIMMYT cross or name	9 Oct 2010	14 Oct 2010	9 Oct 2010	Ug99 resistance
India	Super 152	Pfau/Seri.1B//Amad/3/Waxwing	20 MSS	30 MSS	15 MR	APR (Sr2 +)
	Super 172	Munal#1	10 MSS	15 MSS	5 MR	APR (Sr2 +)
	Baj	Waxwing/4/Sni/Trap#1/3/Kauz*2/Trap//Kauz	5 M	10 M	5 MR	APR (Sr2 +)
		Frankolin#1	5 MSS	10 MSS	5 MR	APR (Sr2 +)
Bangladesh		Frankolin#1	5 MSS	10 MSS	5 MR	APR (Sr2 +)
Nepal		Frankolin#1	5 MSS	10 MSS	5 MR	APR (Sr2 +)
		Picaflor#1	1 MS	1 MS	10 MR	APR (Sr2 +)
		Danphe#1	1 MS	1 MS	5 MR	APR (Sr2 +)
Pakistan	NARC2011	Oasis/SKauz//4*Bcn/3/2*Pastor	5 R	10 R	15 MR	Sr2 + Sr25
Afghanistan	Koshan 09	Quaiu #1	30 MR	40 MR	1 R	Sr2 + SrTmp
	Muqawim 09	Oasis/SKauz//4*Bcn/3/2*Pastor	5 R	10 R	15 MR	Sr2 + Sr25
	Baghlan 09	Picaflor#1	1 MS	5 MS	10 MR	APR (Sr2 +)
	To be named	Chonte#1	1 MS	5 MS	10 MR	APR (Sr2 +)
Egypt	Misr 1	Oasis/SKauz//4*Bcn/3/2*Pastor	5 R	10 R	20 MR	Sr2 + Sr25
	Misr 2	Skauz/Bav92	5 R	10 R	5 MR	Sr2 + Sr25
Ethiopia	Danda	Danphe#1	1 MS	1 MS	5 MR	APR (Sr2 +)
	Kakaba	Picaflor#1	1 MS	5 MS	10 MR	APR (Sr2 +)
Kenya	Robin	Babax/Lr42//Babax*2/3/Tukuru	10 RMR	15 RMR	10 RMR	Sr2 + SrTmp
	Eagle 10	ND643/2*Weebill	10 R	10 RMR	5 RMR	SrND643

Table 13.2 Resistance to Ug99 group of races of stem rust fungus in various international yield trials and screening nurseries distributed/under distribution in 2011–2012 (resistance categories based on data from Njoro, Kenya 2009-main season and 2010-off-season and recorded when the susceptible check Cacuke had dried out following 100 % stem rust severity)

Resistance Category	Stem rust severity (%)	Entries in International Trials and Nurseries									
		33rd ESWYT[1]		45th IBWSN[2]		6th SRRSN[3]		6th EBWYT[4]		3rd CSISA-HT-EM[5]	
		No.	%	No.	%	No.	%	No.	%	No.	%
Adult Plant											
Near-immune resistant	1	10	21.7	55	16.6	57	47.1	10	35.7	3	10.7
Resistant	5–10	17	37.0	99	29.9	38	31.4	12	42.9	7	25.0
Resistant-Moderately Res	15–20	12	26.1	99	29.9					13	46.4
Moderately resistant	30			22	6.6			2	7.1	1	3.6
Mod. Resistant-Mod. Sus	40			13	3.9					1	3.6
Moderately susceptible	50–60									1	3.6
Mod. Susceptible-Sus	70–80			4	1.2						
Susceptible	90–100			1	0.3					1	3.6
Race-specific											
Sr25	5–10	2	4.3	5	1.5	6	5.0	1	3.6		
Sr26	1–5	1	2.2	5	1.5	6	5.0	2	7.1		
SrHuw234	30			1	0.3						
SrSha7	5–10			7	2.1	11	9.1				
SrTmp	10–50	4	8.7	19	5.7	1	0.8	1	3.6	1	3.6
SrUnknown	1–10			1	0.3	2	1.7				

[1] 33rd Elite Spring Wheat Yield Trial
[2] 45th International Bread Wheat Screening Nursery
[3] 6th Stem Rust Resistance Screening Nursery
[4] 6th Elite Bread Wheat Yield Trial
[5] 3rd Cereal Systems Initiative for South Asia-Heat Tolerant-Early Maturity Yield Trial

over-wintering/over-summering periods, and tend to be less temperature sensitive. As such, they continue to play a critical role in further reducing rust populations. Routine deployment of hypersensitive response (HR) genes in combinations or pyramids has been less practical in the past due to the small number of broadly effective genes available at any given time and the lack of diagnostic markers needed to pyramid genes. A first step in achieving wheat lines with multiple sources of resistance is to have more genes at hand with which to work. Four years into the DRRW project, approximately 30 major genes have been identified as sources of resistance to Ug99 and its variants, many of which are the focus of DRRW research efforts. Alongside gene-discovery efforts are activities to supply breeders with user-friendly molecular markers for efficient gene pyramiding, and better diagnostics to determine the genetic basis of resistance of deployed varieties. Stem rust resistance genes that are effective against Ug99 based on seedling and/or field testing (Singh et al. 2011b) and have been tagged by molecular markers include: *Sr2* (Mago et al. 2011), *Sr13* (Admassu et al. 2011; Simons et al. 2011), *Sr22* (Olson et al. 2010; Periyannan et al. 2011), *Sr25* (Liu et al. 2010), *Sr26* (Liu et al. 2010), *Sr32* (Bariana et al. 2001), *Sr33* (Sambasivam et al. 2008), *Sr35* (Zhang et al. 2010), *Sr39* (Mago et al. 2009; Niu et al. 2011), *Sr40* (Wu et al. 2009), *Sr44* (Liu et al. 2012), *Sr45* (Sambasivam et al. 2008), *Sr46* (Lagudah, personal communication), *Sr47* (Faris et al. 2008), *Sr50* (Anugrahwati et al. 2008), *Sr51* (Liu et al. 2011a), *Sr52* (Qi et al. 2011), *Sr53* (Liu et al. 2011b), *SrCad* (Hiebert et al. 2011), *Sr57* (synonym *Lr34/Yr18*), and *SrWeb* (Hiebert et al. 2010).

Several Ug99-effective genes originated from wild relatives of wheat; genetic stocks with small translocation segments harboring the resistance genes have been developed (e.g. *Sr22*, *Sr32*, *Sr39*, *Sr40*, *Sr47*, and *Sr53*), or are under development (e.g. *Sr35*, *Sr37*, *Sr43*, *Sr44*, *Sr51*) to minimize potential for linkage with undesirable traits. In addition, translocation stocks with minimal alien chromatin are being developed to introduce additional newly discovered stem rust resistance genes from *Thinopyrum intermedium*, *Aegilops caudata*, *Th. ponticum*, *Th. junceum*, *Ae. sharonenesis*, *Ae. speltoides*, and *Haynaldia villosa*.

Screening for Ug99 resistance in the secondary and tertiary gene pools of wheat has indicated a number of genes (Pretorius et al. 2000) are likely accessible (Jin et al. 2009; Rouse and Jin 2011; Rouse et al. 2011; Xu et al. 2009) and additional resistance screening and introgression efforts are ongoing through the DRRW project. New genetic resources under development include materials derived from wild emmer and other tetraploids, *H. villosa*, *Triticum dicoccoides*, *T. monococcum*, *Thinopyrum* spp.-derived amphiploids, *Secale cereal*, *Sitopsis*section species, *Ae. tauschii* and other *Aegilops* species.

13.2.1.2 Responsible Gene Stewardship

As new sources of rust resistance are obtained in a breeder-ready form, it is critical that genes be deployed responsibly. Limiting the stem rust population is the most practical means of reducing the probability of virulent mutations and epidemic

development within a season, over time, and across the globe. Both APR and major gene resistance can significantly contribute to this goal. With this in mind, DRRW project partners advocate for responsible use of APR and major gene types of resistance. Deploying varieties with genetic complexity of resistance and a high level of resistance that limits inoculum production are primary DRRW goals. Widespread replacement of susceptible varieties is another critical component of reducing the stem rust population advocated through the BGRI and the DRRW project.

While genes such as *Sr2/Yr30*, *Lr34/Yr18/Sr57*, *Lr46/Yr29*, and *Lr67/Yr46/Sr55* (Herrera-Foessel et al. 2011) appear to be "durable," with no current evidence to suggest that they will break down, most other APR loci are poorly characterized and it is difficult to carry out research that adequately tests race specificity. Leaf rust and stripe rust genetic systems have clearly demonstrated the presence of race-specific resistance genes that provide moderate protection and are not readily identified in seedling resistance tests. These observations suggest that even when deploying lines with stem rust resistance based on APR, breeders should strive for near-immunity, where presumably four or more APR genes have been combined to achieve the high level of resistance, even if moderate levels of APR are considered adequate. Increased complexity of resistance should reduce inoculum production and the probability of virulent mutations if uncharacterized APR loci are race-specific. The pleiotropic effects of APR genes identified to date on other wheat rusts adds additional impetus to combining four or more whenever possible. The CIMMYT program has rapidly progressed in routine selection of elite germplasm with near-immunity to Ug99 stem rust based on APR, in addition to leaf and stripe rust races found throughout the world (Singh et al. 2011b). Fixation of *Sr2/Yr30*, *Lr34/Yr18/Sr57*, *Lr46/Yr29*, and *Lr67/Yr46/Sr55* resistance loci may be an attractive goal for many breeding programs throughout the world, upon which further improvements in rust resistance can be targeted.

Based on the fate of essentially all rust resistance major genes deployed to date, a singly deployed race-specific gene is an ephemeral effort and perpetuates the boom-and-bust cycles too familiar to wheat breeders and growers alike. The DRRW project partners advocate for responsible gene stewardship by deploying three or more major genes in combination whenever major genes are used. Towards this goal, one approach of the DRRW project is to assemble closely linked resistance genes into linkage blocks. The linkage blocks will reduce the likelihood that the resistance genes will segregate after materials are recombined in future breeding efforts, which would result in a dangerous scenario of single gene deployment. These linkage blocks will then be incorporated into promising CIMMYT materials and, ultimately, land in the hands of breeders in national programs. Chromosome arms with strong candidate combinations have been identified, and work toward this approach is now underway. Multiple linked-gene combinations are needed in order to prevent widespread deployment of a single linkage block. A goal of project partners is to develop distinct gene combinations for different user groups defined by geography, agroecology, or grain type preferences.

Although there is some evidence in other species that a minor gene may help prolong the effectiveness of a single major gene (Brun et al. 2010), there is no

concrete evidence of this in wheat resistance to rust. Since the molecular mechanism of major gene function in the hypersensitive response is thought to be independent of the slow rusting response characteristic of APR materials, the most responsible gene deployment strategies at present should not rely on single major genes in an APR background, but rather should strive to maximize complexity of each type of resistance by maintaining three or more major genes in combination and/or near-immunity based on APR.

13.3 Conclusion

The BGRI has been successful in raising the awareness of a potential worldwide threat to wheat production from Ug99 and its more virulent derivatives that have now colonized Eastern and Southern Africa, Yemen in Arabian Peninsula and Iran in West Asia. Through funding from BMGF, joined by DFID in the 2nd phase, the DRRW project has been successful in developing global partnership involving 20 institutions worldwide that has identified new genetic diversity for both race-specific and minor genes based resistance and making progress towards identifying breeders-friendly molecular markers for their proper utilization and deployment in combinations. Releases of resistant varieties in various countries, development and distribution of high-yielding germplasm with high to adequate adult-plant resistance through CIMMYT International Yield Trials and Screening Nurseries promises further releases of new resistant varieties and utilization in breeding programs. Reducing worldwide area sown to susceptible wheat and promoting varieties that carry diverse race-specific resistance gene combinations and complex multigenic APR is the only environmentally sound way forward to secure wheat crop from potential epidemics of stem rust.

Acknowledgements The authors would like to thank Cally Arthur for carefully reviewing the manuscript and Dave Hodson for sharing Fig. 13.2. The BGRI is supported by funds provided through a grant from the Bill & Melinda Gates Foundation and the UK Department for International Development to Cornell University for the Durable Rust Resistance in Wheat (DRRW) project.

References

Admassu B, Perovic D, Friedt W, Ordon F (2011) Genetic mapping of the stem rust (*Puccinia graminis* f. sp *tritici* Eriks. & E. Henn) resistance gene *Sr13* in wheat (*Triticum aestivum* L.). Theor Appl Genet 122:643–648

Anugrahwati DR, Shepherd KW, Verlin DC et al (2008) Isolation of wheat-rye 1RS recombinants that break the linkage between the stem rust resistance gene *SrR* and secalin. Genome 51: 341–349

Bariana HS, Hayden MJ, Ahmed NU et al (2001) Mapping of durable adult plant and seedling resistances to stripe rust and stem rust diseases in wheat. Aust J Agric Res 52:1247–1255

Bhavani S, Singh RP, Argillier O et al (2011) Mapping durable adult plant stem rust resistance to the race Ug99 group in six CIMMYT wheats. In: McIntosh RA (ed) BGRI 2011 Technical Workshop. Borlaug Global Rust Initiative, St Paul, pp 43–53

Brun H, Chevre AM, Fitt BDL et al (2010) Quantitative resistance increases the durability of qualitative resistance to *Leptosphaeria maculans* in *Brassica napus*. New Phytol 185:285–299

CIMMYT (2005) Sounding the alarm on global stem rust. CIMMYT, Mexico, D.F.

Coffman R (2011) Durable rust resistance in wheat interim progress report. Cornell University

Faris JD, Xu SS, Cai XW et al (2008) Molecular and cytogenetic characterization of a durum wheat-*Aegilops speltoides* chromosome translocation conferring resistance to stem rust. Chromosome Res 16:1097–1105

Herrera-Foessel SA, Lagudah ES, Huerta-Espino J et al (2011) New slow-rusting leaf rust and stripe rust resistance genes *Lr67* and *Yr46* in wheat are pleiotropic or closely linked. Theor Appl Genet 122:239–249

Hiebert CW, Fetch TG, Zegeye T (2010) Genetics and mapping of stem rust resistance to Ug99 in the wheat cultivar Webster. Theor Appl Genet 121:65–69

Hiebert CW, Fetch TG, Zegeye T et al (2011) Genetics and mapping of seedling resistance to Ug99 stem rust in Canadian wheat cultivars 'Peace' and 'AC Cadillac'. Theor Appl Genet 122:143–149

Jacobs TH, Parlevliet JE (1993) Durability of disease resistance. Kluwer Academic Publishers, Dodrecht

Jin Y, Pretorius ZA, Singh RP (2007) New virulence within race TTKS (Ug99) of the stem rust pathogen and effective resistance genes. Phytopathology 97:S137–S137

Jin Y, Szabo LJ, Pretorius ZA et al (2008) Detection of virulence to resistance gene *Sr24* within race TTKS of *Puccinia graminis* f. sp *tritici*. Plant Dis 92:923–926

Jin Y, Szabo LJ, Rouse MN et al (2009) Detection of virulence to resistance Gene *Sr36* within the TTKS race lineage of Puccinia graminis f. sp tritici. Plant Dis 93:367–370

Johnson R (1984) A critical analysis of durable resistance. Annu Rev Phytopathol 22:309–330

Liu SX, Yu LX, Singh RP et al (2010) Diagnostic and co-dominant PCR markers for wheat stem rust resistance genes *Sr25* and *Sr26*. Theor Appl Genet 120:691–697

Liu WX, Jin Y, Rouse M et al (2011a) Development and characterization of wheat-*Ae. searsii* Robertsonian translocations and a recombinant chromosome conferring resistance to stem rust. Theor Appl Genet 122:1537–1545

Liu WX, Rouse M, Friebe B et al (2011b) Discovery and molecular mapping of a new gene conferring resistance to stem rust, *Sr53*, derived from *Aegilops geniculata* and characterization of spontaneous translocation stocks with reduced alien chromatin. Chromosome Res 19:669–682

Liu WX, Danilova TV, Rouse M et al (2013) Development and characterization of a compensating wheat-*Thinopyrum intermedium* Robertsonian translocation with *Sr44* resistance to stem rust (Ug99). Theor Appl Genet, Jan 29 (e-pub ahead of print)

Mago R, Zhang P, Bariana HS et al (2009) Development of wheat lines carrying stem rust resistance gene *Sr39* with reduced *Aegilops speltoides* chromatin and simple PCR markers for marker-assisted selection. Theor Appl Genet 119:1441–1450

Mago R, Brown-Guedira G, Dreisigacker S et al (2011) An accurate DNA marker assay for stem rust resistance gene *Sr2* in wheat. Theor Appl Genet 122:735–744

McIntosh RA, Wellings CR, Park RF (1995) Wheat rusts: an Atlas of resistance genes. CSIRO Publications, Victoria

Niu ZX, Klindworth DL, Friesen TL et al (2011) Targeted Introgression of a wheat stem rust resistance gene by DNA marker-assisted chromosome engineering. Genetics 187:1011– 1068

Olson EL, Brown-Guedira G, Marshall D et al (2010) Development of wheat lines having a small introgressed segment carrying stem rust resistance Gene *Sr22*. Crop Sci 50:1823–1830

Periyannan SK, Bansal UK, Bariana HS et al (2011) A robust molecular marker for the detection of shortened introgressed segment carrying the stem rust resistance gene *Sr22* in common wheat. Theor Appl Genet 122:1–7

Pretorius ZA, Singh RP, Wagoire WW, Payne TS (2000) Detection of virulence to wheat stem rust resistance gene *Sr31* in *Puccinia graminis* f. sp. *tritici* in Uganda. Plant Dis 84:203

Qi LL, Pumphrey MO, Friebe B et al (2011) A novel Robertsonian translocation event leads to transfer of a stem rust resistance gene (*Sr52*) effective against race Ug99 from *Dasypyrum villosum* into bread wheat. Theor Appl Genet 123:159–167

Reynolds MP, Borlaug NE (2006) Applying innovations and new technologies for international collaborative wheat improvement. J Agric Sci 144:95–110

Rouse MN, Jin Y (2011) Genetics of Resistance to Race TTKSK of *Puccinia graminis* f. sp. *tritici* in *Triticum monococcum*. Phytopathology 101:1418–1423

Rouse MN, Wanyera R, Njau P, Jin Y (2011) Sources of resistance to stem rust race Ug99 in Spring Wheat Germplasm. Plant Dis 95:762–766

Sambasivam PK, Bansal UK, Hayden MJ et al (2008) Identification of markers linked with stem rust resistance genes *Sr33* and *Sr45*. Proceedings of the 11th International Wheat Genetics Symposium, Brisbane, Australia

Simons K, Abate Z, Chao SM et al (2011) Genetic mapping of stem rust resistance gene *Sr13* in tetraploid wheat (*Triticum turgidum* ssp *durum* L.). Theor Appl Genet 122:649–658

Singh RP, Huerta-Espino J, Rajaram S (2000) Achieving near-immunity to leaf and stripe rusts in wheat by combining slow rusting resistance genes. Acta Phytopathol Hun 35:133–139

Singh RP, Hodson DP, Jin Y et al (2006) Current Status, likely migration and strategies to mitigate the threat to wheat production from race Ug99 (TTKS) of stem rust pathogen. CAB reviews: perspectives in agriculture, veterinary science, nutrition and natural resources 1:1–13

Singh RP, Hodson DP, Huerta-Espino J et al (2008) Will stem rust destroy the world's wheat crop? In: Sparks DL (ed) Advances in Agronomy, Vol 98. Elsevier Academic Press Inc, San Diego, pp 271–309

Singh RP, Hodson DP, Huerta-Espino J, Jet al. (2011a) The emergence of Ug99 races of the stem rust fungus is a threat to world wheat production. In: VanAlfen NK, Bruening G, Leach JE (eds) Annual Review of Phytopathology, Vol 49. Annual Reviews, Palo Alto, pp 465–481

Singh RP, Huerta-Espino J, Bhavani S et al (2011b) High yielding CIMMYT spring wheats with resistance to Ug99 and other rusts developed through targeted breeding. In: McIntosh RA (ed) BGRI 2011 Technical Workshop. Borlaug Global Rust Initiative, St. Paul, pp 98–104

Wu SY, Pumphrey M, Bai GH (2009) Molecular mapping of stem-rust-resistance Gene *Sr40* in wheat. Crop Sci 49:1681–1686

Xu SS, Jin Y, Klindworth DL et al (2009) Evaluation and characterization of seedling resistances to stem rust Ug99 races in wheat-alien species derivatives. Crop Sci 49:2167–2175

Yu LX, Lorenz A, Rutkoski J et al (2011) Association mapping and gene-gene interaction for stem rust resistance in CIMMYT spring wheat germplasm. Theor Appl Genet 123:1257–1268

Zhang WJ, Olson E, Saintenac C et al (2010) Genetic maps of stem rust resistance Gene *Sr35* in diploid and hexaploid wheat. Crop Sci 50:2464–2474

Chapter 14
Genomes, Chromosomes and Genes of the Wheatgrass Genus *Thinopyrum*: the Value of their Transfer into Wheat for Gains in Cytogenomic Knowledge and Sustainable Breeding

Carla Ceoloni, Ljiljana Kuzmanović, Andrea Gennaro, Paola Forte, Debora Giorgi, Maria Rosaria Grossi and Alessandra Bitti

Abstract Perennial wheatgrass species of the genus *Thinopyrum* possess several appealing attributes for wheat improvement, contributing to tolerance to biotic and abiotic stresses, as well as to quality and even to yield increase. Major genes or QTLs underlying such traits have been identified on numerous chromosomes of both diploid (*Th. elongatum* and *Th. bessarabicum*) and polyploid (mainly *Th. intermedium* and *Th. ponticum*) representatives of the genus, having different genome origin (E, J, St/S) and involving several homoeologous groups. *Thinopyrum* chromosomes sharing homoeology with wheat group 7 chromosomes turned to be particularly rich in beneficial genes; among them, a *Th. ponticum* group 7 chromosome referred to as 7Ag or 7el has been extensively targeted in various successful attempts of harnessing its attractive gene content. A survey of the several wheat translocation/recombinant lines involving this chromosome in the background of both bread and durum wheat is given. Such lines are described as highly valuable tools for a variety of studies, from development of integrated genetic and physical maps, to the analysis of structural and functional characteristics associated with defined alien chromosome subregions. The validity of *Th. ponticum* group 7 transfers as breeding materials (notable genes and traits including *Lr19*, *Sr25*, Fusarium head blight resistance, yellow pigment content, and even yield) is also highlighted. Finally, examples are given of pyramiding of group 7 *Thinopyrum* genes through 'precision' breeding strategies of chromosome

C. Ceoloni (✉) · L. Kuzmanović · A. Gennaro · P. Forte · D. Giorgi · M. R. Grossi · A. Bitti
Department of Agriculture, Forestry, Nature and Energy (DAFNE), University of Tuscia, 01100 Viterbo, Italy
e-mail: ceoloni@unitus.it

A. Gennaro
Present address: GMO Unit, European Food Safety Authority (EFSA), 43126 Parma, Italy

D. Giorgi
Present address: Italian National Agency for New Technologies, Energy and the Environment (ENEA), Casaccia Research Center, Rome, Italy

R. Tuberosa et al. (eds.), *Genomics of Plant Genetic Resources,*
DOI 10.1007/978-94-007-7575-6_14,

engineering, which, efficiently aided by current genetic, cytogenetic and genomic (or, collectively, 'cytogenomic') technologies, enable a multifaceted and sustainable improvement of the wheat crop based on the use of the wealth of natural genetic resources of its related gene pools.

14.1 The Vast *Thinopyrum* Genus

The wheatgrass genus *Thinopyrum* Löve (Löve 1982; Dewey 1984), belonging to the tertiary gene pool of wheat relatives, is a particularly large reservoir of desirable traits for improvement of cultivated *Triticum* species. The genus encompasses a large number of perennial species, exhibiting a wide range of ploidy levels, from diploidy to decaploidy. The diploid (2n = 2x = 14) *Th. elongatum* (Host) D.R. Dewey [syn. *Agropyron elongatum* (Host) P. Beauv., *Lophopyrum elongatum* (Host) A. Löve] and *Th. bessarabicum* (Savul. & Rayss) Löve [syn. *Agropyron bessarabicum* (Savul. & Rayss)], as well as the hexaploid (2n = 6x = 42) *Th. intermedium* (Host) Barkworth & D.R. Dewey [syn. *Agropyron intermedium* (Host) P. Beauv.), *Elytrigia intermedia* (Host) Nevski] and the decaploid (2n = 10x = 70) *Th. ponticum* (Popd.) Barkworth & D.R. Dewey [syn. *Agropyron elongatum* (Host) P. Beauv., *Lophopyrum ponticum* (Popd.) A. Löve, *Elytrigia pontica* (Popd.) Holub] are among the most extensively exploited in wheat breeding. They have been used for more than half a century to enrich wheat germplasm with a plentiful array of genes for disease and pest resistance (e.g. Li and Wang 2009), for tolerance to environmental stresses, such as salinity (Colmer et al. 2006), high temperature, strong light and drought at the grain filling stage (Li et al. 2008), as well as for processing quality (Liu et al. 2008), and even yield-related traits (Singh et al. 1998; Kuzmanović et al. 2013).

While for *Th. elongatum* and *Th. bessarabicum* assignment of the E (= E^e) and of the closely related, yet not homologous, J (= E^b) genome, respectively, is relatively well established (e.g., Jauhar 1990; Wang 1992, 2011), genomic composition of polyploid representatives of the genus has been long debated (Zhang et al. 1996a, b; Chen et al. 1998; Chen 2005; Arterburn et al. 2011; Wang 2011). Nonetheless, a shared belief is that a third genome type, named St or S, characterizing the *Pseudoroegneria* genus, definitely enters in the genomic make-up of both *Th. intermedium* and *Th. ponticum*, possibly representing the core genome in the perennial genera of Triticeae tribe (Wang et al. 2010a). Since the St/S genome, in turn, shows close relatedness with the E and J genomes, as proved by extensive autosyndetic pairing (Jauhar 1995; Cai and Jones 1997) and cross-hybridization in genomic *in situ* hybridization (GISH) experiments (Zhang et al. 1996a), a conclusive definition of the genome formulas of *Th. intermedium* and *Th. ponticum* has not been reached. Therefore, the former has been described with various genome formulas, including E^eE^bSt (Wang and Zhang 1996) and E_1E_2St (Zhang et al. 1996b) or JJ^sS (Chen et al. 1998), while $E^eE^bE^xStSt$ (reviewed in Li and Zhang 2002) or JJJ^sJ^s (Chen et al. 1998) have been indicated for the latter. The controversy focuses in particular on the distinction between complete St/S-genome chromosomes on one hand (Zhang et al. 1996a), and chromosomes with presence of St/S genomic DNA confined to

pericentromeric regions (i.e. J^s type chromosomes, see Chen et al. 1998) on the other, hypothesized to result from intergenomic rearrangements in the course of polyploid evolution. At any rate, hybridization to the St/S genomic DNA, whether complete or segmental, represents a distinctive mark of the genomic origin of the *Thinopyrum* chromosome(s) involved.

In determining the genomic structure of polyploid grass species, and its relationships with that of other *Thinopyrum* as well as *Triticum* species, an important contribution has come from creation and analysis of wheat-*Thinopyrum* hybrids and complete or, more frequently, partial amphiploids (e.g. Charpentier 1992; Jauhar 1995; Zhang et al. 1996a, b; Chen et al. 1998; Fedak et al. 2000; Fedak and Han 2005; Oliver et al. 2006; Li et al. 2008; Sepsi et al. 2008). These often possess several appealing attributes for wheat breeding, including perennial growth habit (Tsitsin 1965; Jauhar 1995; Cox et al. 2002; Bell et al. 2010); they also represented valuable starting material for the isolation of single alien chromosome addition and substitution lines into the wheat background, enabling genome and homoeologous group attribution of the specific *Thinopyrum* chromosome, besides that chromosomal assignment of genes of interest (e.g. Forster et al. 1987; Charpentier 1992; Friebe et al. 1992; Zhang et al. 1992; Larkin et al. 1995; Chen et al. 1999). This also applies to hybrids and amphiploids involving the diploid species of the genus, i.e. *Th. elongatum* (e.g. Dvorak and Knott 1974; Jauhar et al. 2009) and *Th. bessarabicum* (William and Mujeeb-Kazi 1995). Although the picture may be complicated by intergenomic rearrangements rather frequently occurring in such amphiploids (reviewed in Fedak and Han 2005), and hence maintained in derived addition and substitution lines (see, e.g., Wang et al. 2010b), the latter materials represent a step forward the reduction of unwanted alien genetic material that characterizes a breeding-friendly approach of harnessing useful alien variability for wheat improvement (see, for review, Ceoloni and Jauhar 2006; Qi et al. 2007).

14.2 Valuable *Thinopyrum* Genes Transferred to Wheat

Major genes or QTLs underlying traits capable to enhance wheat performance have been found on numerous *Thinopyrum* chromosomes, having different genome origin (where identified) and wheat-alien homoeologous relationships (Wang 2011). To recall a few, most relevant cases, two different group 2 *Th. intermedium* chromosomes derived from the Zhong 5 partial amphiploid were found to carry a gene conferring resistance to BYDV (Tang et al. 2000) and a highly effective resistance gene toward *Fusarium graminearum* infection (Fedak and Han 2005), respectively. On the long (L) arm of a group 3 *Th. ponticum* chromosome, assigned to a J^s genome (Li et al. 2003), the leaf rust (*Lr24*) and the stem rust (*Sr24*) resistance genes are located, which, following isolation of various types of translocations (reviewed in Friebe et al. 1996), were extensively deployed in wheat breeding worldwide (Friebe et al. 1996; Bariana et al. 2007; Kolmer et al. 2007). Another widely exploited *Th. ponticum* gene, located on the L arm of a group 6 homoeologous chromosome (Knott 1961; Friebe

et al. 1994), is *Sr26*, highly effective to the wheat stem rust disease (Park 2007), and even against the recently emerged and threatening pathogen race Ug99 (Liu et al. 2010). Of unknown genomic origin but still on a group 6 *Thinopyrum* chromosome is also a gene conferring resistance to cephalosporium stripe, an important disease of winter wheat, with no resistance source available within the species gene pool (Cai et al. 1996). As to genes related to tolerance to abiotic stresses, the widely spread halophytic behaviour among *Thinopyrum* species, both diploid and polyploid, is likely ascribable to genes on more than one chromosome (reviewed in Colmer et al. 2006). However, the distal end of the long arm of chromosome 3E of *Th. elongatum* contains a gene(s) exerting a major enhancing effect on the leaf sodium 'exclusion' trait in wheat (Mullan et al. 2009).

14.2.1 Transfers Involving Homoeologous Group 7 Chromosomes

A *Thinopyrum* chromosome group that was found to be particularly rich in valuable genes for wheat improvement is the one sharing homoeology with the wheat group 7 chromosomes. Chromosomes 7 of the polyploid *Th. intermedium* and *Th. ponticum* were greatly exploited for the transfer of numerous beneficial traits. As to the former species, genes for resistance to the three wheat rust diseases were identified on a Zhong 5-derived 7Ai-2 chromosome of seemingly J^s genomic origin, with a leaf rust and a stem rust resistance gene being located in the distal region of the L arm, and a stripe rust resistance gene(s) in the short (S) arm or in the proximal region of the same L arm (Tang et al. 2000). To a 7Ai-1 chromosome of probably different genomic origin (reviewed in Ayala-Navarrete et al. 2007), the *Sr44* stem rust resistance gene (McIntosh et al. 1995; Xu et al. 2009), as well as one or possibly two genes for BYDV resistance (*Bdv2* + ?) were allocated and introduced into various wheat substitution and translocation lines (Banks et al. 1995; Sharma et al. 1995; Crasta et al. 2000; Ayala-Navarrete et al. 2007, 2009). As to the 7 L arm of the diploid *Th. elongatum* (7EL), it was shown in various studies to carry a highly effective resistance gene toward Fusarium head blight (FHB), an important wheat disease causing yield and grain quality losses worldwide (Fedak et al. 2003; Oliver et al. 2005; Miller et al. 2011).

Perhaps the most extensively targeted group 7 *Thinopyrum* chromosome is the one belonging to *Th. ponticum*, originally named 7Ag (Sears 1973) or 7el (Sharma and Knott 1966; Knott et al. 1977). Its genomic origin remains to be ascertained, although preliminary evidence (Ceoloni et al. unpublished) is suggestive of a St/J^s derivation (see § 14.1). Two largely homologous 7el chromosomes, designated $7el_1$ and $7el_2$, started to be exploited more than 30 years ago. In addition to a nearly complete pairing (Dvorak 1975; Forte et al. 2011), their homology, probably traceable back to two accessions of the same *Th. ponticum* species, is also substantiated by considerable correspondence in gene content, particularly at the L arm level. When introduced into wheat cultivars in the form of substitution and translocation lines, both 7el chromosomes revealed the presence of genes controlling resistance to leaf- and stem

rust, as well as of genes determining yellow flour pigmentation (*Yp*) and Segregation distortion (*Sd*) (Kim et al. 1993; see also ahead). Distinct phenotypes for reaction to *Fusarium* infection differentiate the two 7el sources, $7el_1$ being susceptible and $7el_2$ bearing a major QTL in the distal end of its long arm (Shen and Ohm 2007).

Although $7el_2$ was shown to cause moderate leaf rust resistance in adult wheat plants (Knott et al. 1977), it was for the stem rust resistance gene *Sr43*, located on its L arm, that this chromosome was induced to pair and recombine with wheat chromosomes (Kibirige-Sebunya and Knott 1983). Despite the validity still exhibited by *Sr43* toward stem rust races spread in different parts of the world, including the African Ug99 (e.g. Xu et al. 2009; Sivasamy et al. 2010), the available wheat transfer lines, all with sizable alien introgressions (Kim et al. 1993) and also negatively affected by the presence of *Yp* and *Sd* genes (Kibirige-Sebunya and Knott 1983), have had so far limited use in breeding. Although, as above mentioned, genes with similar effects also occur on the $7el_1$ chromosome, and hence on at least some of the wheat transfer lines bearing portions of it, they have not hampered as drastically as for $7el_2$ its exploitation for wheat improvement, particularly when durum wheat represented the target species (see ahead).

14.3 Wheat-*Th. ponticum* $7el_1$ Transfer Lines: Highly Valuable Mapping and Breeding Tools

Extensive work has been performed on genes and traits of $7el_1$ derivation, thanks to the availability of a wide array of translocation and recombinant lines involving this chromosome, produced both in bread wheat (*Triticum aestivum* L., $2n = 6x = 42$, genome AABBDD) and in durum wheat (*Triticum durum* Desf., $2n = 4x = 28$, genome AABB) backgrounds.

No doubt, the consistent interest for $7el_1$ transfers was primarily addressed to the *Lr19* gene, conferring a largely effective resistance to wheat leaf rust across time and space (Gennaro et al. 2009). Agrus, a $7el_1$(7D) substitution line, was the starting material initially used to incorporate *Lr19* into bread wheat cultivars through both irradiation (Sharma and Knott 1966; Knott 1968) and induced homoeologous recombination (Sears 1973, 1978). Among the radiation-induced translocations, the one named T4 (=Agatha), consisting of a 70 % long $7el_1L$ segment inserted onto the wheat 7DL arm (Dvorak and Knott 1977; Friebe et al. 1996), proved to have a good compensating ability (Friebe et al. 1994). An additional resistance gene, *Sr25*, conferring resistance to several races of wheat stem rust (McIntosh et al. 1976; Knott 1989a), and recently shown to be effective even against Ug99 (Li and Wang 2009; Liu et al. 2010), enhanced the validity of T4. As such, this sizable translocation has been incorporated into several bread wheat varieties, notably the CIMMYT cultivar Oasis 86 and various derivatives still bred in several world Countries endangered by Ug99 (e.g. DRRW project 2011). To eliminate the associated yellow flour pigmentation defect, several attempts were carried out starting from line T4. The first ended up with two mutants exhibiting differently reduced pigment content (Knott 1980, 1984)

but poor agronomic performance (Knott 1989b), one (Agatha 28-4) carrying a point mutation for a candidate gene for the *Yp* phenotype (see ahead), the other (Agatha 235-6) having apparently lost a terminal portion of the original $7el_1L$ segment, also including *Sr25* (Friebe et al. 1994). Additional white-endosperm T4 derivatives with shortened amounts of alien chromatin were obtained by *ph1*-induced homoeologous recombination (Marais 1992a; Marais et al. 2001). Although their use in breeding was hampered by negative effects caused by the presence of at least one *Sd* gene (Marais et al. 2001; see also ahead, § 14.4.3), they represented valuable materials to assign the critical alien genes and a number of molecular markers of the RFLP type to defined intervals along the $7el_1L$ arm, hence to start assessing its synteny and colinearity with wheat 7L arms. Previously, the first series of 7AgL (= $7el_1L$)-7DL homoeologous recombinants (Sears 1973, 1978) had been discriminated on the basis of their *Lr19* and *Sd* phenotypes, combined with results of meiotic pairing analyses and of the distribution pattern of a *Thinopyrum*-specific repetitive sequence (Zhang and Dvorak 1990). Based on these parameters, only an indirect and comparative estimate of the relative length of the alien and wheat chromosome segments in the individual recombinant chromosomes could be established. This was somewhat true also for a set of deletions induced along the T4 segment (Marais 1992b), which were an additional, useful tool in the effort of determining the linear order and approximate physical location of the relevant *Thinopyrum* genes (centromere-*Sd1*-*Lr19*-*Sr25*/*Yp*), as well as of some RLFP (Prins et al. 1996, 1997; Prins and Marais 1998) and several AFLP (Groenewald et al. 2005) markers, but not to establish a precise correspondence between their genetic and physical map positions along the $7el_1L$ arm. This was partly accomplished by Zhang et al. (2005), who assigned ten RFLPs, the STS-$Lr19_{130}$ marker (Prins et al. 2001) and, based on phenotypes, the *Lr19* and *Yp* genes, to subregions of a sizable $7el_1L$ segment, originally present on the 7DL arm of Sears' line Transfer#1 (Sears 1973) and fractionated into various homoeologous recombinant chromosomes. One of the selected recombinants in the course of the work of Zhang et al. (2005) was proved to have lost the very distal *Yp* gene, yet retaining *Lr19* and the proximally associated *Sd1* gene (see § 4.3).

On the other hand, since the *Lr19* + *Sr25* + *Yp* association is advantageous for durum wheat, a *ph1*-mediated chromosome engineering strategy was specifically targeted to the concomitant increase in novel breeding materials of leaf and stem rust resistance as well as of pigment content of their semolina and pasta products. The latter, corresponding to an increase of seed carotenoids, could not only respond to a relevant aesthetic requirement by a large portion of consumers, but also provide with added value in terms of health beneficial antioxidants (Adom et al. 2003). As donor line, a tetraploid primary recombinant was employed, derived from Sears' bread wheat 7A-$7el_1$ Transfer#12 (Sears 1973), carrying an exceedingly large amount of $7el_1$ chromatin to be tolerated at the 4x ploidy level (Ceoloni et al. 1996). Various secondary recombinants with exchanges involving the 7AL and $7el_1L$ homoeologous arms were isolated, and the amount of $7el_1$ chromatin introgressed in each of them assessed by GISH (Fig. 14.1, see also Ceoloni et al. 2000, 2005). One of them, namely R5-2-10, carrying 23 % of distal $7el_1L$ including the *Lr19* + *Sr25* + *Yp* genes, and showing no sign of any *Sd* effect (Ceoloni et al. 2005), exhibited very good agronomic

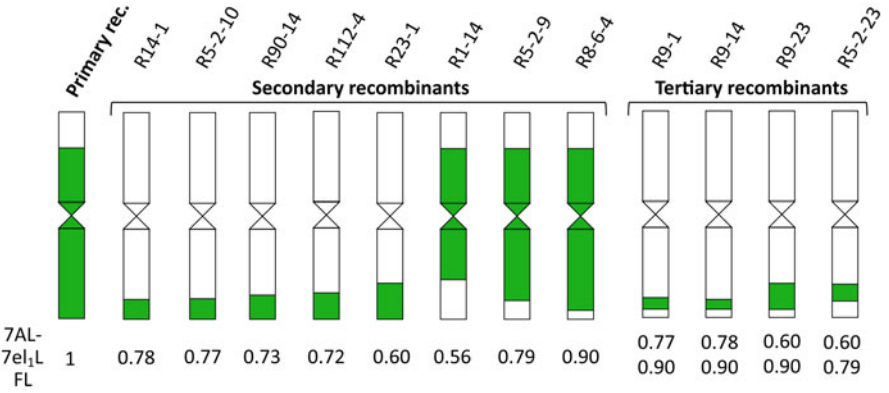

Fig. 14.1 Schematic representation of GISH-based physical maps of 7A/7el$_1$ chromosomes, individually present in different durum wheat-*Th. ponticum* recombinant lines. The green colour corresponds to *Th. ponticum* chromatin. FL = position of the recombinant breakpoints, expressed as fractional length of the 7AL/7el$_1$L arm

and quality performance (Gennaro et al. 2003). Following a conventional breeding process with adapted varieties, a derivative of R5-2-10 has been released in Italy in 2010 with the name of Cincinnato.

From homologous recombination between secondary recombinant chromosomes having overlapping 7el$_1$L regions, tertiary recombinants with 7el$_1$L intercalary segments were also obtained (Fig. 14.1; see also Ceoloni et al. 2005; Gennaro et al. 2009; Kuzmanovic 2011).

Altogether, the 7AL-7el$_1$L durum wheat recombinants, for several of which near-isogenic lines (NILs) have been obtained, represent a highly valuable tool to carry out a variety of studies, partly illustrated in the following, starting from integrated genetic and physical mapping of the 7L critical arms, with assignment of numerous markers and genes to several 7L subregions, to the analysis of some structural and functional characteristics associated with defined 7el$_1$L portions.

14.3.1 A Durum Wheat-Th. ponticum 7AL-7el$_1$L Cytogenetic Map

As revealed by GISH, the durum wheat-*Th. ponticum* recombinant chromosomes encompass a series of physical breakpoints (BPs) concentrated in the distal half of the 7AL-7el$_1$L arms, from fractional length (FL) 0.56 up to the telomere (Fig. 14.1; see also Ceoloni et al. 2005). This distribution confirms the highly preferential occurrence of crossovers in distal chromosomal regions in the Triticeae (e.g. Saintenac et al. 2009). The number and distribution of the durum wheat-*Th. ponticum* 7AL-7el$_1$L BPs usefully widen and complement previously obtained ones along the same arms, including the most frequently employed for deletion bin mapping of the wheat

7AL arm (FLs 0.39, 0.71, 0.74, 0.86 and 0.90; see, e.g., Hossain et al. 2004; Sourdille et al. 2004). The eight of them consistently used in mapping analyses can be grouped into two configuration types, five (R14-1, R5-2-10, R90-14, R112-4 and R23-1) having a 7A centromere and distal $7el_1L$ segments (FLs at 0.78, 0.77, 0.73, 0.72 and 0.60, respectively; Fig. 14.1), and three (R1-14, R5-2-9 and R8-6-4) with a $7el_1$ centromere but with distal 7AL segments (FLs at 0.56, 0.79 and 0.90, respectively; Fig. 14.1). Although the indicated FLs of the recombinant 7AL-$7el_1L$ arms have to be considered somewhat different from those of wheat 7AL deletion lines, since length estimates showed the alien arm to be at least 10 % shorter than the wheat arm, it seems obvious that a particular concentration of BPs involves the subterminal 20 % of the arms. In this region, where a high density of EST loci was observed (Hossain et al. 2004), the *Lr19*, *Sr25* and *Yp* target genes are located.

A detailed cytogenetic map was developed by mapping several RFLP, SSR, STS and EST markers, as well as established alien genes and putative genes/QTLs for phenotypes of interest, to the eight 7 L subregions defined by the 7AL-$7el_1L$ recombination BPs indicated above (Gennaro et al. 2010; Kuzmanovic 2011). A summary version of such map is given here (Fig. 14.2), in which landmark and other frequently used markers across a number of Triticeae group 7 maps (e.g. Hohmann et al. 1994, 1995; Gale et al. 1995; Nelson et al. 1995; de la Peña et al. 1997; Hossain et al. 2004; Somers et al. 2004; Xue et al. 2008) are mainly reported.

RFLP markers, whose linear order along 7AL mostly confirms previous evidence, have been indicative of a good colinearity between $7el_1L$ and the 7AL of the background durum wheat genotypes of the recombinant lines. However, thanks to the codominant 7AL/$7el_1L$ polymorphism frequently detected by these markers, a clear $7el_1L$-specific duplication was found to involve the *Xwg380* locus, present in both the FL 0.72-0.60 (*Xwg380.1*, Fig. 14.2) and 0.73-0.72 (*Xwg380.2*, Fig. 14.1) intervals, as proved by the R23-1 recombinant showing both loci and R112-4 with *Xwg380.2* only. This duplication does not seem to be unique to the *Th. ponticum* $7el_1L$ arm; in fact, in contrast to the case of wheat 7A and *T. monococcum* $7A^m$ (Dubcovsky et al. 1996), duplication of the syntenic region including *Xwg380* was observed in other Triticeae, including *Ae. squarrosa* (Gale and Devos 1996) and *H. vulgare* (Heun et al. 1991). On the other hand, a 7AL-specific duplication concerned the *Xpsr687* locus, with the most frequently detected very distal location (e.g. Hohmann et al. 1994; Gale et al. 1995; Zhang et al. 2005) being accompanied by a second, more proximal one on the 7AL of the durum wheat materials (Fig. 14.2). Duplications included as well EST sequences, as in the case of BE426802 (FL 0.77-0.78, Fig. 14.2), turned out to be duplicated both on 7A and 7B of the durum wheat recombinants (Fig. 14.3). The same EST also highlights duplicated loci in adjacent bins at the distal end of bread wheat 7AL and 7BL (http://wheat.pw.usda.gov/wEST; Fig. 14.3), where duplications for the *XBF482714* and *XBE426274* loci, not detected in the durum materials, were identified too (http://wheat.pw.usda.gov/wEST).

Such duplications are part of complex rearrangements which apparently involve the long arms of wheat homoeologous group 7 chromosomes and break them into discontinuous syntenic units relative to the rice genome (Francki et al. 2004). The

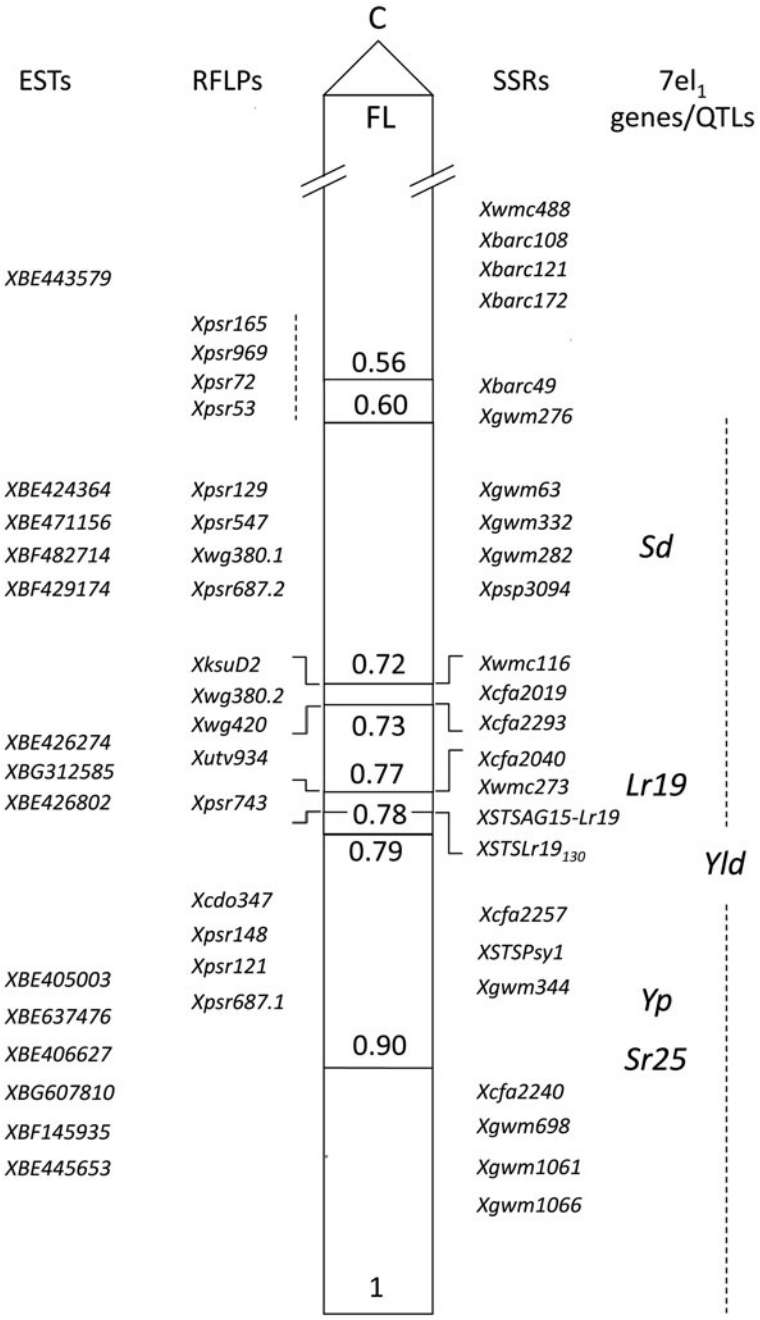

Fig. 14.2 A 7AL-7el$_1$L cytogenetic map divided into subregions defined by the physical breakpoints of durum wheat-*Th. ponticum* recombinant chromosomes (see Fig. 14.1), expressed as fractional lengths (FL) of the arm (C = centromere). Several molecular markers and relevant 7el$_1$L genes/QTLs have been associated to the physical intervals (vertical dotted lines indicate an approximate location)

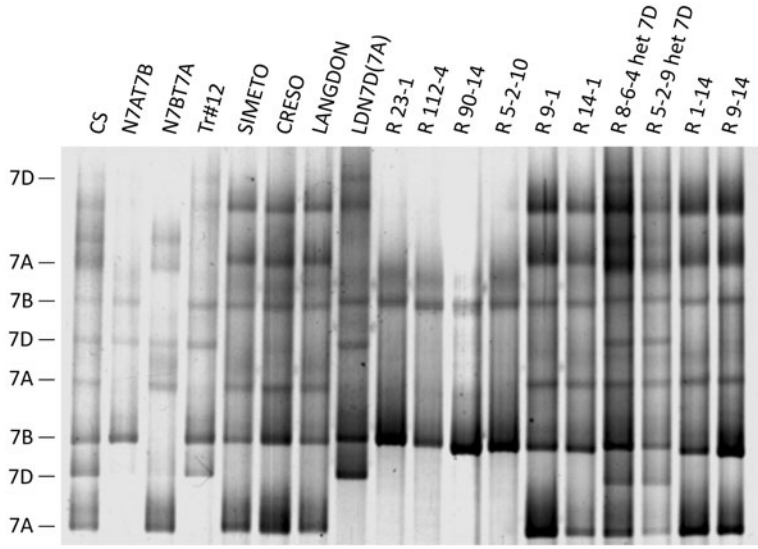

Fig. 14.3 Polyacrylamide gel separation of amplification products derived from the wheat BE426802 EST sequence. Both in durum and in bread wheat genotypes at least two bands could be unequivocally assigned to all the group 7 homoeologous chromosomes, which indicates a conserved locus duplication, both within and between species. Genotypes tested (from left to right): CS = bread wheat cv. Chinese Spring; N7AT7B, N7BT7A = CS nullisomic-tetrasomic lines; Tr#12 = CS 7A/7el$_1$ recombinant line; Simeto, Creso, Langdon (LDN) = durum wheat cultivars; LDN7D(7A) = LDN substitution line with a pair of 7D chromosomes in place of a 7A pair; R23-1 to R9-14 = durum wheat-*Th. ponticum* recombinants (see also Figs. 14.1 and 14.2), all homozygous for a 7A-7el$_1$ recombinant chromosome, except for R8-6-4 and R5-2-9, heterozygous for a 7A-7el$_1$ and a 7D chromosome

presence of noncoding, repetitive DNA, ubiquitous and abundant in plant genomes (e.g. Heslop-Harrison and Schwarzacher 2011), may have favoured such rearrangements, contributing to speciation. Various *Thinopyrum*-specific sequences of this type were isolated (McIntyre et al. 1988; Zhang and Dvorak 1990; Bournival et al. 1994), and the presence of some of them ascertained along the 7el$_1$L chromosome arm (Zhang and Dvorak 1990; Bournival et al. 1994). Although in lower copy number, some of these sequences were also detected in the genomes of other Triticeae species, including wheat (Wang and Wei 1995). Closely related, repeated DNA sequences, both coding and noncoding, distally located on wheat 7AL and *Th. ponticum* 7el$_1$L arms, might have represented the "substrate" for unequal crossing-over events, particularly under less efficient homology scrutinizing mechanisms, as it possibly happens in the absence of the wheat *Ph1* gene (Corredor et al. 2007; Moore and Shaw 2009). An unequal crossing-over probably caused the duplication, including the *Wsp-1* locus (co-presence of *Wsp-D1* and *Wsp-B1* homoeoloci), occurred in the bread wheat *Lr19*-149 white-endosperm recombinant during induced homoeologous recombination (*Ph1*$^-$), which led to its relocation from 7DL to 7BL (Prins et al. 1997).

A homoeo-duplication was also observed in one of the durum wheat-*Th. ponticum* secondary recombinants, namely R8-6-4, and in its derived types R9-1, R9-14 and R9-23 (Fig. 14.1). All these recombinants share a terminal 7AL segment (FL 0.90-1) in which, in contrast with the marker order exhibited by the other recombinants, 7AL alleles for a set of distal RFLP loci (from *Xpsr743* up to *Xpsr687.1*) were detected together with the corresponding $7el_1$ homoeoalleles, located in the proximally adjacent segment. While co-dominant markers, like most of the RLFPs used, provided with strong, supporting evidence of the duplication and of its size, only for some SSR loci (i.e. *Xcfa2040* and *Xwmc273*) inclusion into the potentially duplicated region could be inferred from their presence in R8-6-4 and derived genotypes, incompatible with their physical position in all other recombinants (FL 0.77-0-78). Likewise, the BE426802 EST sequence (FL 0.77-0-78, Fig. 14.1) was unexpectedly present in the distal 7AL segment of R8-6-4 and its derivatives.

Similarly to what observed for RFLP markers, linear order of SSR markers highlighted by the durum wheat-*Th. ponticum* recombinants generally confirmed what reported in previous wheat maps (e.g. Röder et al. 1998; Somers et al. 2004; Xue et al. 2008; Francki et al. 2009), though discrepancies were observed with the results of microsatellite deletion bin mapping by Sourdille et al. (2004). As to ESTs mapped onto the 7AL-$7el_1$L durum wheat recombinant arms, the majority of them have been assigned by deletion bin mapping to the most distal portion of bread wheat 7AL (FL 0-86-0.90 and 0.90-1, see http://wheat.pw.usda.gov/wEST), while in the *Th. ponticum* arm they appear to be scattered in four BP intervals, spanning from FL 0.60 to 1 (Fig. 14.2). This can be partly due to the size difference between the wheat and alien arm ($7el_1L < 7AL$, see above), and also to a possibly different structural organization leading to variations in marker/gene density. Data from genetic maps indicate the length of the entire 7el chromosome (95.76 cM in Zhang et al. 2011) to be shorter than the consensus map of wheat 7A (131 cM in Somers et al. 2004), with corresponding average genetic distances between markers of 1.47 cM vs. 2.1 cM, respectively (Zhang et al. 2011).

14.4 Recent Insights into $7el_1L$-linked Genes

14.4.1 Rust Resistance Genes: Lr19 and Sr25

Lr19 could be precisely located within the 1 % $7el_1L$ portion differentiating the two durum wheat recombinant lines carrying the smallest $7el_1L$ distal portions, i.e. R5-2-10 (FL 0.77, resistant to leaf rust) and R14-1 (FL 0.78, susceptible). Near-isogenic lines (NILs) of this pair of contrasting recombinant lines represented ideal material to exploit a PCR-based DNA fingerprinting technique (NBS profiling, see van der Linden et al. 2004) to identify NBS-containing DNA sequences linked to *Lr19* (Gennaro et al. 2009). The choice of this approach was based on the reasoning that since *Lr19* confers a hypersensitive response to the pathogen (Saini et al. 1998; Plotnikova 2008), it was considered likely that the gene would be a member of the major

nucleotide-binding site (NBS)-leucine-rich repeat (LRR) plant *R* gene family. The successful application of NBS profiling to the durum wheat–*Th. ponticum* recombinant NILs led to the enrichment of the wheat and alien target regions with new, effective markers (Gennaro et al. 2009). In particular, one differential PCR product (AG15), tightly linked to *Lr19* (Fig. 14.2), was isolated and sequenced, representing the first full-length resistance gene analog isolated from *Th. ponticum* and from distant wheat relatives in general. From the AG15 sequence, a codominant STS marker was developed for the presence of *Lr19* (STSAG15-*Lr19*, see Fig. 14.2; Gennaro et al. 2009). Since AG15, proved to be constitutively expressed, as is common for *R* genes, represents a candidate sequence for *Lr19*, a stable transformation assay has been carried out to test whether the sequence confers resistance to leaf rust. Results of leaf rust inoculation of transgenic plants raise some doubts on the actual AG15-*Lr19* equivalence (Gennaro et al. unpublished).

The chromosomal location of the AG15 sequence makes it unlikely that it represents *Sr25*, which is distal to *Lr19* and so tightly linked to *Yp* that repeated rounds of homoeologous recombination and radiation-induced deletions failed to separate the two genes (Marais et al. 2001; Groenewald et al. 2005). In fact, the two yellow pigment mutants of Agatha (=T4) translocation line (Knott 1980), one carrying and the other lacking *Sr25* (Knott 1984; see also § 3), both give an amplification product with AG15-specific primers (Gennaro et al. 2009). Therefore *Sr25* is likely to reside in the $7el_1L$ segment shared by R5-2-10 and R14-1, both of which are thought to carry *Yp*, since their endosperm is pigmented (Gennaro et al. 2003) and both carry the $7el_1L$ allele for the a strong candidate for the *Yp* phenotype, i.e. the *Psy1* gene coding for phytoene synthase (Gennaro et al. 2008; see also ahead). As above recalled, *Sr25* is one of the few, still effective genes providing an effective resistance to the stem rust race Ug99. Use of a codominant STS marker developed from the wheat group 7 EST BF145935, located at the distal end of wheat 7L arms and of *Th. ponticum* $7el_1L$ (Fig. 14.2; Ayala-Navarrete et al. 2007; Kuzmanovic 2011), provides a robust assay for *Sr25* deployment in breeding programs (Liu et al. 2010).

14.4.2 Yellow Pigment Gene(s)

The R14-1 and R5-2-10 recombinants exhibited similar levels of pigment content and yellow index (Gennaro et al. 2003), confirming the more distal location of the *Yp* gene(s) compared to *Lr19*. Actually, the recombinant R14-1 chromosome is the only one, within the durum wheat-*Th. ponticum* set, in which the *Lr19-Yp* gene association is broken, as it includes *Yp* and lacks *Lr19*. As mentioned above, also $7el_2$-derived substitution and translocation lines into bread wheat exhibit an increased yellow flour pigmentation (Kibirige-Sebunya and Knott 1983), probably determined by an allele of the $7el_1$ gene(s). Location of *Yp* gene(s)/QTL(s) in distal regions of the long arms of homoeologous group 7 chromosomes appears to be highly conserved in the Triticeae, being common to wild grasses, such as *Th. ponticum* and *Hordeum chilense* (Alvarez et al. 1998), as well as to the 7A and 7B chromosomes of bread (Parker et al.

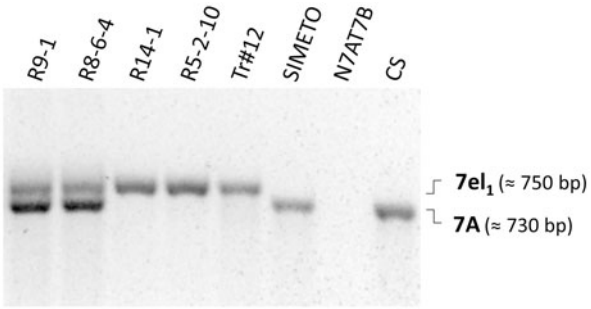

Fig. 14.4 7AL/7el$_1$L polymorphism for the *Psy-1* gene as highlighted by an STS-*Psy1* codominant marker. The 7AL *Psy1* allele, present in the durum wheat cv. Simeto and in the bread wheat cv. Chinese Spring (CS), is replaced by the 7el$_1$L orthologue in R5-2-10 and R14-1 durum wheat recombinants and in CS recombinant Tr#12, while both alleles are present in the durum wheat recombinant R8-6-4 and in its R9-1 derivative, carrying a homoeoduplication in this region (see text). N7AT7B = CS nullisomic 7A-tetrasomic 7B control line

1998; Mares and Campbell 2001) and durum wheat (Elouafi et al. 2001; Pozniak et al. 2007; Singh et al. 2009). As mentioned above for the durum wheat 7AL-7el$_1$ recombinants, in the same regions the *Psy1* gene is located, coding for phytoene synthase, a key enzyme in carotenoid biosynthesis, hense a likely candidate gene for the yellow pigment trait. The strong association between *Psy1* polymorphism and yellow pigment content has been demonstrated both in wheat, notably durum wheat (Pozniak et al. 2007; Zhang and Dubcovsky 2008; Singh et al. 2009), as well as in *H. chilense* (Atienza et al. 2007) and *Th. ponticum* (Gennaro et al. 2008; Zhang and Dubcovsky 2008). Mapping of *Psy1-A1* on durum wheat 7AL localized the gene to a 4–8 cM interval at the distal end of the arm, in close proximity to the SSR marker locus *Xcfa2257* (Singh et al. 2009). This evidence is confirmed by the cytogenetic maps of durum wheat-*Th. ponticum* recombinants, with the R14-1 7el$_1$ segment (FL 0.78-1 on the 7AL-7el$_1$L arm, Fig. 14.1) containing a *Psy1-el$_1$1* specific allele and a null allele for a number of SSR loci, including *Xcfa2257* (Fig. 14.2). The *Psy1-el$_1$1* allele has been identified by use of a codominant marker (STS*Psy1*, see Fig. 14.2), developed by exploiting the available information on wheat *Psy1* sequence variation (Gennaro et al. 2008 and unpubl.; Fig. 14.4). While in distal recombinants, such as R14-1 and R5-2-10 (Fig. 14.1), the 7AL allele is replaced by the colinear 7el$_1$L allele, both are present in other recombinant types, i.e. R8-6-4 and its derivative R9-1 (Fig. 14.4). This result shows that the homoeoduplication which characterizes R8-6-4 and lines derived from it (see § 3.1) includes the *Psy1* locus.

The peculiar structure of such recombinants allows to delimit the physical location of the *Psy1-el$_1$1* locus to the FL 0.78-0.90 (the former being the 7el$_1$L-7AL breakpoint of R14-1, the latter that of R8-6-4 and derivatives, all containing a *Psy1-el$_1$1* locus). In the 7AL segment occupying the terminal 10 % of the arm in such recombinants, the *Psy1-A1* locus is located. The simultaneous presence of the two

Psy1 genes, which can be easily followed by the codominant *Psy1* marker, could have beneficial, additive effects on durum wheat semolina colour.

While *Psy1* appears to have a main role in the expression of the yellow pigment phenotype, a second QTL affecting pigment content in the durum wheat grain appears to be present on 7AL, not associated to *Psy1-A1*, and more proximally located, near the *Xwmc116* locus (Zhang and Dubcovsky 2008; Singh et al. 2009). That differences in yellow pigment content could not be fully explained by the *Psy1*-linked gene was also put forward for *Th. ponticum*; in this case, the possible existence of a second *Yp* gene on the $7el_1L$ arm was firstly suggested on the basis of identification of mutant (Marais 1992b) or recombinant (Ceoloni et al. 2000) genotypes with intermediate levels of yellow pigment content. In a recent investigation of the two "white-endosperm" Agatha (=T4) mutants (Zhang and Dubcovsky 2008), the evidence concerning Agatha 235-6, which, contrary to Agatha 28-4, showed an unchanged *Psy1* sequence, led to propose that the mutation might affect a different gene within the same T4 segment. No candidate is so far available for the second *Yp* gene, both in wheat and in *Th. ponticum*, although the apparent widespread duplication of *Psy* in the Poaceae (e.g. Gallagher et al. 2004) supports the speculation, at least in the durum wheat case, that another, yet unidentified *Psy* paralog might be located at or near *Xwmc116* (Singh et al. 2009).

14.4.3 Segregation Distortion Genes

Segregation distortion (SD), i.e. the deviation of observed genetic ratios from the expected Mendelian ratios within a segregating population, has been observed in a wide variety of organisms, including fungi, plants, insects, and mammals (Lyttle 1991; Taylor and Ingvarsson 2003). In fact, genetic elements that cause SD, hence referred to as *Sd* genes, may be potent evolutionary forces, particularly in terms of species differentiation and genome restructuring (Hurst and Werren 2001). Genomic regions harboring markers with abnormal segregation ratios have been reported in many crop species including wheat (Kumar et al. 2007 and references therein). *Sd* genes have also been identified in several wild wheat relatives, with their effects, mostly on male gametogenesis, revealed upon hybridization with wheat (e.g. Endo 1990, 2007; Faris et al. 1998). In wheat-alien combinations, the effect of the *Sd*-bearing chromosome can vary from its selective or even exclusive retention in the wheat background through generations (preferential transmission), apparently achieved by causing various abnormalities to gametes lacking it (e.g. Finch et al. 1984; Endo 1990; Nasuda et al. 1998), to a more or less dramatic self-elimination (e.g. Prins and Marais 1999). Within the *Thinopyrum* genus, there seems to be a conserved tendency for group 7 chromosomes to harbor *Sd* genes among polyploid species, including *Th. intermedium* (Kong et al. 2008) and *Th. ponticum* (Kibirige-Sebunya and Knott 1983; Scoles and Kibirige-Sebunya 1983; Prins and Marais 1999; Cai et al. 2011), though not in the diploid *Th. elongatum* (Dvorak 1980).

The *Th. ponticum* 7elL arm of both el_1 and el_2 origin contains one or more *Sd* genes, which have been rather intensively studied, primarily because of the linked, beneficial genes of breeding relevance. Preferential transmission through the female gametes of heterozygous plants was firstly observed for the entire $7el_2$ or portions of it carrying the stem rust resistance gene *Sr43* (Kibirige-Sebunya and Knott 1983). The extent of distortion turned out to be dependent on the genetic background of different varieties used. Similar observations were made on the $7el_1$-derived T4 translocation (Knott 1971; Marais 1992c; McIntosh et al. 1995), although the effects were predominantly on the male gametophyte. However, progressively shortened T4 segments, relocated to chromosome 7B in the course of repeated rounds of homoeologous recombination, often showed the opposite behaviour, i.e. from mild to strong self-elimination (Prins and Marais 1999; Marais et al. 2001; Groenewald et al. 2005). The genetic structure and the SD phenotype of such materials suggested the hypothesis of two *Sd* genes being located along $7el_1$L, one, named *Sd1*, proximal to the *Xpsr129* locus (see Fig. 14.2), the other, *Sd2*, being distally placed, in rather close linkage with *Lr19* (Prins and Marais 1999; Marais et al. 2001; Groenewald et al. 2005). The presence of *Sd2* was not confirmed in a recent study, based on a *Th. ponticum* 7el genetic map developed from a cross between a 7D($7el_1$) and a 7D($7el_2$) substitution line into bread wheat cv. Thatcher (Cai et al. 2011). In that study, three chromosome regions with skewed segregation were found along 7el, altogether spanning almost the entire short arm and the proximal half of the long arm. Of the long arm portions involved in SD, one, named SDR3, covers a few markers, including PSR129, and the more proximal SDR2 region includes the remaining part of the arm; therefore, either one might coincide with the *Sd1* locus mentioned above. Cai et al. (2011) tended to exclude the SDR3 to *Sd1* correspondence, because of the different behaviour of the $7el_1$ and $7el_2$ *Sd* genes with respect to previous findings (e.g. Kibirige-Sebunya and Knott 1983; Prins and Marais 1999). However, the simultaneous presence in their material of *Sd* genes from the two sources, as well as differences in the genetic backgrounds of the various materials are likely causes of the variable effects observed, in terms of strength and also in the type of mostly affected gametophyte.

As to effect and behaviour of the $7el_1$ *Sd* gene(s) in the durum wheat recombinants, where a different group 7 homoeologue from most bread wheat transfers, i.e. 7A, is involved, it was initially observed that, with female transmission being normal in all cases, that through the male germline seemed to be correlated with the size of the alien segment, resulting normal for $7el_1$L segments spanning up to 28 % of the recombinant arm length (e.g. lines R5-2-10 and R112-4, Fig. 14.1; Ceoloni et al. 2005). In fact, based on previous mapping of *Sd1* (Marais et al. 2001; Groenewald et al. 2005), the gene was not expected to be included in any of the $7el_1$L distal segments inserted in the various recombinant chromosomes (Fig. 14.1), the most proximal BP (R23-1) including the *Xpsr129* locus, but not *Xpsr165* (Fig. 14.2). Nonetheless, male transmission of the R23-1 recombinant chromosome was somewhat below normal (Ceoloni et al. 2005), and the defect was initially attributed to the excessive length of its $7el_1$L portion in a tetraploid background. However, the reduced transmission and associated abnormalities, mainly affecting male gametogenesis, do not seem to have been eliminated in new, interstitial recombinant chromosomes, such as R9-23

and R5-2-23 (Fig. 14.1), obtained from the cross of R23-1 with R9-1 and R5-2-9, respectively, whose segment size is definitely within the range of those well tolerated even in durum wheat (Ceoloni and Jauhar 2006). All these recombinant chromosomes share the most proximal $7el_1L$ portion, delimited by the R23-1 breakpoint; it seems therefore likely that a segregation distortion factor is contained in that segment, possibly corresponding to the SDR3 region of Cai et al. (2011). Segregation ratios ranged from normal to highly distorted (always in the direction of self -elimination) when the R23-1 chromosome was introduced in the background of different durum wheat varieties, but also showed heterogeneity among progeny of the same cross and in the course of successive generations (Grossi et al. 2009 and unpubl.). Further, analysis of the meiotic and post-meiotic stages of pollen development in R23-1 as compared to lines with 28 % (R112-4) and 23 % (R5-2-10) of $7el_1L$ and to their controls lacking any alien segment showed R23-1 to be the only line presenting various irregularities, confirming the assignment of the *Sd* factor(s) to the 12 % $7el_1L$ differentiating R112-4 from R23-1 (FL 0.60-0.72; Fig. 14.2). R23-1 homozygotes exhibiting relatively normal fertility have been obtained, but this achievement appears to be rather inconsistent, definitely influenced by the recipient background, hence with possible negative consequences in a breeding perspective.

Further analyses on the interstitial recombinants, carrying different amounts of distal alien chromatin (Fig. 14.1), might shed some light on the possible existence of a second *Sd* gene on $7el_1L$ (Marais et al. 2001; Groenewald et al. 2005). In this hypothesis, it could be speculated that, at least in a durum wheat background, both genes are needed for the segregation distortion phenotype to be expressed (as in R23-1), while the absence of the more proximal and likely stronger one (*Sd1*?) would prevent it (as in R112-4 and R5-2-10, hypothetically carrying *Sd2* only).

14.4.4 Yield QTL(s)

The existence of a QTL positively affecting yield (here called *Yld-7el*$_1$*L*) in wheat-*Th. ponticum* stocks, determining an increase in grain yield and biomass, was originally suggested by CIMMYT, on the basis of results obtained using NILs of the original T4 (= Agatha) translocation (70 % of $7el_1$ L arm onto wheat 7DL) into various bread wheat backgrounds (Singh et al. 1998; Reynolds et al. 2001; Monneveaux et al. 2003).

On the other hand, a strong yield QTL, mainly affecting grain number per ear and expressed mostly under stressed conditions, was identified on distal 7AL (Quarrie et al. 2005). Its expression was suggested to be associated to a gene(s) controlling flag-leaf width and chlorophyll content, indirectly affecting the amount of assimilates transferred to the spike (Quarrie et al. 2006). Such QTL effects were associated with a cluster of markers including SSR loci *Xwmc273*, *Xgwm332* and *Xpsp3094* (Fig. 14.2). Prompted by this evidence, NILs of some of the durum wheat 7AL-$7el_1L$ recombinant lines, with wheat-alien breakpoints around the suggested location of the 7AL yield QTL, i.e. R23-1, R112-4 and R5-2-10, have been and currently are employed in

comparative field experiments. As a whole, the results of the first 2 years of trials carried out in Central Italy confirm the positive effects on yield originally observed from the introgression into bread wheat of the sizable T4 segment; moreover, different yield-related traits have been associated to defined $7el_1L$ sub-regions (Kuzmanović et al. 2013). In fact, the three recombinant lines showed differential phenotypes for traits such as seed and tiller number/plant, above-ground biomass, flag-leaf width, and grain yield. For line R112-4, in particular, significantly increased values were recorded for tiller number at heading (+ 27 %) and at harvest (+ 25 %), as well as for biomass (+ 28 %) and seed number/plant (+ 27 %), altogether contributing to a significantly higher yield (+ 35 %), at least in the more favourable conditions of 2010. Line R23-1 also exhibited a significant increase in seed number/plant, though accompanied by a decrease in total grain yield, presumably due to negative effects on grain weight determined by presence of *Sd* gene(s) (see above). Thus, in the $7el_1L$ segment shared by R112-4 and R23-1 and absent from R5-2-10 (FL 0.72-0.77), a putative QTL for seed number/plant may be located.

While field trials of the 7AL-$7el_1L$ recombinant lines will be extended in time and space for further verification and validation of the expression of the various QTLs, the information so far obtained allows drawing a "first draft" of the association of yield-related traits with specific $7el_1L$ segments already incorporated in recombinant/translocation lines of durum and bread wheat. The combination of this knowledge with genetic and physical mapping data (Fig. 14.2; Kuzmanović 2011; Kuzmanović et al. 2013; Zhang et al. 2011) is expected to make marker-assisted selection of $7el_1L$ yield-enhancing segments (and genes therein) a feasible wheat breeding target in the near future.

14.5 Gene Pyramiding in Wheat Involving 7L *Thinopyrum* Traits

The examples of *Lr19*, *Sr25* and *Yp* genes, and also of the *Yld-7el*$_1$*L* QTL(s) are illustrative of how the combined use of recent methods of analysis is making detection and characterization of wheat-alien introgression products much more efficient and accurate than ever so far, thus strengthening the possibility of a finely tuned and breeding-responsive makeup of the wheat genome through the chromosome engineering approach (see, e.g., Ceoloni and Jauhar 2006; Qi et al. 2007). Moreover, recent achievements demonstrate that such a 'precision' breeding strategy also enables development of more "sophisticated" products, in which different alien genes are pyramided. *Thinopyrum* genes belonging to homoeologous group 7 have been targeted for this goal; one such case is a "trigenomic" combination in a bread wheat background, involving wheat chromosome 7D and recombinant segments between a *Th. intermedium* and a *Th. ponticum* translocation, the former carrying the *Bdv2* gene for BYDV resistance (see § 2.1), the latter being a derivative of the Agatha 28-4 mutant, therefore retaining the *Lr19* and *Sr25* genes but with a reduced yellow

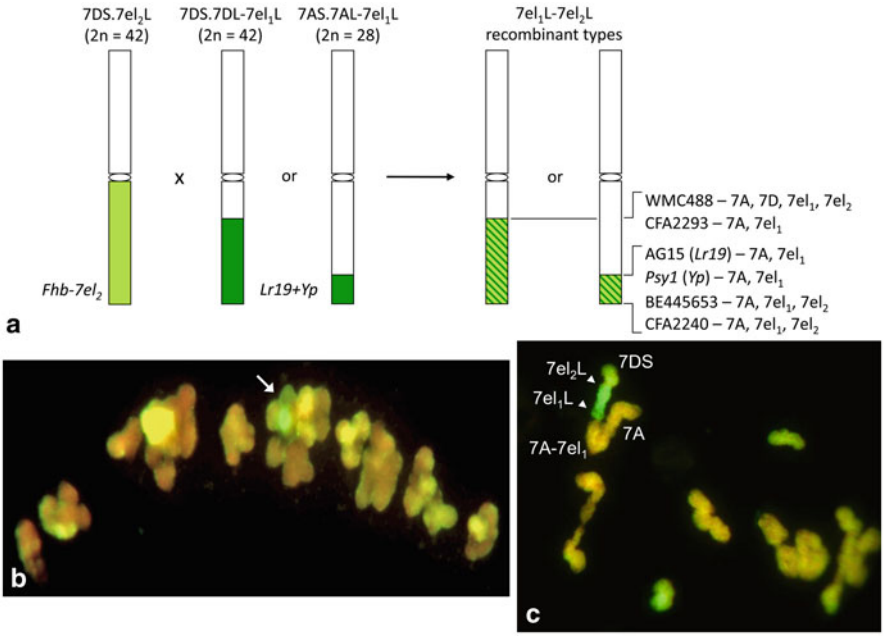

Fig. 14.5 Pyramiding useful genes from different *Th. ponticum* 7el sources (*Lr19+Yp* from $7el_1$ and FHB resistance from $7el_2$) in bread and durum wheat. **a** representative chromosomes of wheat genotypes bearing $7el_1L$ or $7el_2L$ portions and of their $7el_1L$-$7el_2L$ recombinant products. **b** GISH of a meiotic metaphase I cell of a bread wheat F_1 hybrid (2n = 42) with a ring bivalent (arrowed, green fluorescence) resulting from homologous pairing between a 7DS.$7el_2L$ and a 7DS.7DL-$7el_1L$ chromosome. **c** GISH of a partial meiotic metaphase I cell of a bread x durum wheat hybrid in which the trivalent configuration includes the critical pairing partners, i.e. a 7AS.7AL-$7el_1L$ and a 7DS.$7el_2L$ chromosome, being bound at the level of their el_1 and el_2 portions (green fluorescence)

pigment content (Ayala-Navarrete et al. 2007). Further, recombinant inbred lines derived from the cross between a 7D($7el_1$) and a 7D($7el_2$) substitution line, primarily developed to construct a 7el genetic map (Shen and Ohm 2007; Zhang et al. 2011), could also serve as pre-breeding materials for the simultaneous introgression into bread wheat of positive traits from the two 7el sources.

More directly aiming at potential use in breeding is the cross between a bread wheat 7DS.$7el_2$ L centric translocation line (Kim et al. 1993), used as donor of an effective QTL for FHB resistance located in the distal portion of the alien arm (Shen and Ohm 2007; Zhang et al. 2011), and the bread wheat 7DL-$7el_1L$ T4 translocation line (§ 14.3), as well as some of the 7AL-$7el_1L$ durum wheat recombinant lines (§ 14.3). All the $7el_1L$ recombinant lines contain the *Lr19* + *Yp* + *Sr25* genes, and also yield-related QTL(s), but lack any FHB resistance (Fig. 14.5).

The multi-targeted transfers are being aided by suitable polymorphic markers and application of GISH in somatic and meiotic cells (Fig. 14.5; Kuzmanovic 2011; Forte et al. 2011). In spite of the complete homology relating the $7el_1L$ and $7el_2L$

portions bearing the above genes, their position either on corresponding 7D chromosomes in the bread wheat cross combination, or on homoeologous 7DL and 7AL arms in the pentaploid F1's from crosses involving the durum wheat recombinants, affected the occurrence of el_1-el_2 pairing and recombination. In fact, compared to the recovery of 49 % recombinant types following nearly full pairing between the 7DL-7el_1L and 7el_2L arms, pairing involving the 7AL-7el_1L and 7DL-7el_2L arms showed around 35 % frequency, with less than 20 % recombinant types consequently recovered (Fig. 14.5; Forte et al. 2011). Nonetheless, the first recombinants with the desired $el_1 + el_2$ gene combinations have been isolated in both cases, and they are being backcrosses with adapted varieties of bread and durum wheat.

The latter, which has a relatively 'sensitive' genome to chromosome manipulations, has been simultaneously engineered with short chromosomal segments from up to three related species, derived from the wild gene pool (*Th. ponticum* and *Ae. longissima*) and from the D genome of bread wheat, to widen its genetic base with novel disease resistance and quality-related genes (Gennaro et al. 2007; Gennaro et al. 2012). These successful achievements, efficiently aided by current genetic, cytogenetic and genomic (or, collectively, 'cytogenomic') technologies, provide convincing evidence of the feasibility of a multifaceted and sustainable improvement of the wheat crop through a directed and knowledgeable use of the wealth of the natural genetic resources of its related gene pools.

References

Adom KK, Sorrells ME, Hai LR (2003) Phytochemical profiles and antioxidant activity of wheat varieties. J Agric Food Chem 51:7825–7834

Alvarez JB, Martin LM, Martin A (1998) Chromosomal localization of genes for carotenoid pigments using addition lines of *Hordeum chilense* in wheat. Plant Breed 117:287–289

Arterburn M, Kleinhofs A, Murray T, Jones S (2011) Polymorphic nuclear gene sequences indicate a novel genome donor in the polyploidy genus *Thinopyrum*. Hereditas 148:8–27

Ayala-Navarrete L, Bariana HS, Singh RP et al (2007) Trigenomic chromosomes by recombination of *Thinopyrum intermedium* and *Th. ponticum* translocations in wheat. Theor Appl Genet 116:63–75

Ayala-Navarrete L, Tourton E, Mechanicos AA, Larkin PJ (2009) Comparison of *Thinopyrum intermedium* derivatives carrying barley yellow dwarf virus resistance in wheat. Genome 52:537–546

Banks PM, Larkin PJ, Bariana HS et al (1995) The use of cell culture for subchromosomal introgressions of barley yellow dwarf virus resistance from *Thinopyrum intermedium* to wheat. Genome 38:395–405

Bariana HS, Brown GN, Bansal UK et al (2007) Breeding triple rust resistant wheat cultivars for Australia using conventional and marker-assisted selection technologies. Austr J Agric Res 58:576–587

Bell LW, Wade LJ, Ewing MA (2010) Perennial wheat: a review of environmental and agronomic prospects for development in Australia. Crop Pasture Sci 61:679–690

Bournival B, Obanni M, Abad A et al (1994) Isolation of a new species-specific repetitive sequence from *Thynopyrum elongatum* and its use in studies of alien translocations. Genome 37:97–104

Cai J, Zhang X, Wang B et al (2011) A genetic analysis of segregation distortion revealed by molecular markers in *Lophopyrum ponticum* chromosome 7E. J Genet 90:373–376

Cai X, Jones S (1997) Direct evidence for high level of autosyndetic pairing in hybrids of *Thinopyrum intermedium* and *Th. ponticum* with *Triticum aestivum*. Theor Appl Genet 95:568–572

Cai X, Jones S, Murray TD (1996) Characterization of an *Agropyron elongatum* chromosome conferring resistance to cephalosporium stripe in common wheat. Genome 39:56–62

Ceoloni C, Jauhar PP (2006) Chromosome engineering of the durum wheat genome: strategies and applications of potential breeding value. In: Singh RJ, Jauhar PP (eds) Genetic resources, chromosome engineering, and crop improvement: cereals. CRC Press, Taylor & Francis Group, pp 27–59

Ceoloni C, Biagetti M, Ciaffi M et al (1996) Wheat chromosome engineering at the 4x level: the potential of different alien gene transfers into durum wheat. Euphytica 89:87–97

Ceoloni C, Forte P, Ciaffi M et al (2000) Chromosomally engineered durum wheat: the potential of alien gene introgressions affecting disease resistance and quality. In: Royo C, Nachit MM, Di Fonzo N, Araus J-L (eds) Durum wheat improvement in the Mediterranean region: new challenges. Options Méditerranéennes A-40:363–371

Ceoloni C, Forte P, Gennaro A et al (2005) Recent developments in durum wheat chromosome engineering. Cytogenet Genome Res 109:328–344

Charpentier A (1992) Production of disomic addition lines and partial amphidiploids of *Thinopyrum junceum* on wheat. CR Acad Sci Paris 315:551–557

Chen Q (2005) Detection of alien chromatin introgression from *Thinopyrum* into wheat using S genomic DNA as a probe—A landmark approach for *Thinopyrum* genome research. Cytogenet Genome Res 109:350–359

Chen Q, Conner RL, Laroche A, Thomas JB (1998) Genome analysis of *Thinopyrum intermedium* and *Thinopyrum ponticum* using genomic *in situ* hybridization. Genome 41:580–586

Chen Q, Conner RL, Laroche A et al (1999) Genome *in situ* hybridization analysis of *Thinopyrum* chromatin in a wheat-*Th. intermedium* partial amphiploid and six derived chromosome addition lines. Genome 42:1217–1223

Colmer TD, Flowers TJ, Munns R (2006) Use of wild relatives to improve salt tolerance in wheat. J Exp Bot 57:1059–1078

Corredor E, Lukaszewski AJ, Pachón P et al (2007) Terminal regions of wheat chromosomes select their pairing partner in meiosis. Genetics 177:699–706

Cox CM, Murray TD, Jones SS (2002) Perennial wheat germplasm lines resistant to eyespot, *Cephalosporium stripe*, and wheat streak mosaic. Plant Dis 86:1043–1048

Crasta OR, Francki MG, Bucholtz DB et al (2000) Identification and characterization of wheat–wheatgrass translocation lines and localization of barley yellow dwarf virus resistance. Genome 43:698–706

de la Peña RC, Murray TD, Jones SS (1997) Identification of an RFLP interval containing *Pch2* on chromosome 7AL in wheat. Genome 40:249–252

Dewey DR (1984) The genomic system of classification as a guide to intergeneric hybridization with the perennial Triticeae. In: Gustafson JP (ed) Gene manipulation in plant improvement. Plenum Publ Corp, New York, pp 209–279

DRRW—Durable Rust Resistance in Wheat (2011) http://wheatrust.cornell.edu, Project impacts 2011: Appendix F—Project-Wide Impact Metrics and Varieties Released November 2011

Dubcovsky J, Luo M-C, Zhong G-Y et al (1996) Genetic map of diploid wheat, *Triticum monococcum* L., and its comparison with maps of *Hordeum vulgare* L. Genetics 143:983–999

Dvorak J (1975) Meiotic pairing between single chromosomes of diploid *Agropyron elongatum* and decaploid *A. elongatum* in *Triticum aestivum*. Can J Genet Cytol 17:329–336

Dvorak J (1980) Homoeology between *Agropyron elongatum* chromosomes and *Triticum aestivum* chromosomes. Can J Genet Cytol 22:237–259

Dvorak J, Knott DR (1974) Disomic and ditelosomic additions of diploid *Agropyron elongatum* chromosomes to *Triticum aestivum*. Can J Genet Cytol 16:399–417

Dvorak J, Knott DR (1977) Homoeologous chromatin exchange in a radiation-induced gene transfer. Can J Genet Cytol 19:125–131

Elouafi I, Nachit MM, Martin LM (2001) Identification of a microsatellite on chromosome 7B showing a strong linkage with yellow pigment in durum wheat (*Triticum turgidum* L. var. *durum*). Hereditas 135:255–261

Endo TR (1990) Gametocidal chromosomes and their induction of chromosome mutations in wheat. Jpn J Genet 65:135–152

Endo TR (2007) The gametocidal chromosome as a tool for chromosome manipulation in wheat. Chromosome Res 15:67–75

Faris JD, Laddomada B, Gill BS (1998) Molecular mapping of segregation distortion loci in *Aegilops tauschii*. Genetics 149:319–327

Fedak G, Han F (2005) Characterization of derivatives from wheat-*Thinopyrum* wide crosses. Cytogenet Genome Res 109:360–367

Fedak G, Chen Q, Conner RL et al (2000) Characterization of wheat-*Thinopyrum* partial amphiploids by meiotic analysis and genomic in situ hybridization. Genome 43:712–719

Fedak G, Han F, Cao W et al (2003) Identification and characterization of novel sources of resistance to Fusarium head blight. In: Pogna NE, Romanò M, Pogna EA, Galterio G (eds) Proceedings of the 10th International Wheat Genetics Symposium. SIMI, Rome, pp 354–356

Finch RA, Miller TE, Bennett MD (1984) "Cuckoo" *Aegilops* addition chromosome in wheat ensures its transmission by causing chromosome breaks in meiospores lacking it. Chromosoma 90:84–88

Forster BP, Reader SM, Forsyth SA et al (1987) An assessment of the homoeology of six *Agropyron intermedium* chromosomes added to wheat. Genet Res 50:91–97

Forte P, Kuzmanovic L, Gennaro A et al (2011) Using wild species of *Thinopyrum* genus in breeding wheat resistant to Fusarium head blight. Proceedings of the Joint Meeting AGI-SIBV-SIGA. Assisi, Italy. http://www.geneticagraria.it/attachment/Abstract_2011/6A_21.pdf

Francki M, Carter M, Ryan K et al (2004) Comparative organization of wheat homoeologous group 3S and 7L using wheat-rice synteny and identification of potential markers for genes controlling xanthophyll content in wheat. Funct Integr Genomics 4:118–130

Francki MG, Walker E, Crawford AC et al (2009) Comparison of genetic and cytogenetic maps of hexaploid wheat (*Triticum aestivum* L.) using SSR and DArT markers. Mol Genet Genomics 281:181–191

Friebe B, Mukai Y, Gill BS, Cauderon Y (1992) C-banding and in situ hybridization analysis of *Agropyron intermedium*, a partial wheat × *Ag. intermedium* amphiploid and six derived chromosome addition lines. Theor Appl Genet 84:899–905

Friebe B, Jiang J, Knott DR, Gill BS (1994) Compensation indices of radiation-induced wheat-*Agropyron elongatum* translocations conferring resistance to leaf rust and stem rust. Crop Sci 34:400–404

Friebe B, Jiang J, Raupp WJ et al (1996) Characterization of wheat-alien translocations conferring resistance to diseases and pests: current status. Euphytica 91:59–87

Gale MD, Devos KM (1996) Homoeologous group 7. In: McGuire PE, Qualset CO (eds) Progress in genome mapping of wheat and related species. Joint Proceedings of the 5th and 6th Workshops of The International Triticeae Mapping Initiative, Norwich (UK) and Sydney (Australia). Genetic Resources Conservation Program, University of California, pp 107–120

Gale MD, Atkinson MD, Chinoy RL et al (1995) Genetic maps of hexaploid wheat. Proceedings of the 8th International Wheat Genetics Symposium. China Agricultural Scientech Press, Beijing, pp 29–40

Gallagher CE, Matthews PD, Li F, Wurtzel ET (2004) Gene duplication in the carotenoid biosynthetic pathway preceded evolution of the grasses. Plant Physiol 135:1776–1783

Gennaro A, Borrelli GM, D'Egidio MG et al (2003) A chromosomally engineered durum wheat-*Thinopyrum ponticum* recombinant line with novel and promising attributes for varietal development. In: Pogna NE, Romanò M, Pogna EA, Galterio G (eds) Proceeding of the 10th International Wheat Genetics Symposium. SIMI, Rome, pp 881–883

Gennaro A, Forte P, Carozza R et al (2007) Pyramiding different alien chromosome segments in durum wheat: feasibility and breeding potential. Isr J Plant Sci 55:267–276

Gennaro A, Koebner RMD, Ceoloni C (2009) A candidate for *Lr19*, an exotic gene conditioning leaf rust resistance in wheat. Funct Integr Genomics 9:325–334

Gennaro A, Koebner RMD, Janni M, Ceoloni C (2008) Identification of candidate sequences for the *Lr19* and *Yp* genes transferred from *Thinopyrum ponticum* to durum wheat by chromosome engineering. In: Proceedings of the Plant and Animal Genomes XVI Conference. San Diego, CA, USA (abstract P312)

Gennaro A, Kuzmanovic L, Forte P et al (2010) Genetic and cytogenetic maps of durum wheat-*Thinopyrum ponticum* recombinant chromosomes: a tool to finely describe Triticeae 7 L arms and to target alien traits. Proceedings of the 54th Italian Society of Agricultural Genetics Annual Congress. Matera, Italy. Available at: http://www.geneticagraria.it/attachment/SIGA_2010/6_07.pdf

Gennaro A, Forte P, Panichi D et al (2012) Stacking small segments of the 1D chromosome of bread wheat containing major gluten quality genes into durum wheat: transfer strategy and breeding prospects. Mol Breed 30:149–167

Groenewald JZ, Fourie M, Marais AS, Marais GF (2005) Extention and use of a physical map of the *Thinopyrum*-derived *Lr19* translocation. Theor Appl Genet 112:131–138

Grossi MR, Gennaro A, Forte P et al (2009) Segregation distortion following introgression of *Thinopyrum ponticum* DNA into durum wheat. Proceedings of the 53rd Italian Society of Agricultural Genetics Annual Congress. Torino, Italy. Available at http://www.geneticagraria.it/attachment/SIGA_2009/7_20.pdf

Heslop-Harrison JS, Schwarzacher T (2011) Organisation of the plant genome in chromosomes. Plant J 66:18–33

Heun M, Kennedy AE, Anderson JA et al (1991) Construction of a restriction fragment length polymorphism map of barley (*Hordeum vulgare*). Genome 34:437–447

Hohmann U, Endo TR, Gill KS, Gill BS (1994) Comparison of genetic and physical maps of group 7 chromosomes from *Triticum aestivum* L. Mol Gen Genet 245:644–653

Hohmann U, Graner A, Endo TR et al (1995) Comparison of wheat physical maps with barley linkage maps for group 7 chromosomes. Theor Appl Genet 91:618–626

Hossain KG, Kalavacharla V, Lazo GR et al (2004) A chromosome bin map of 2148 expressed sequence tag loci of wheat homoeologous group 7. Genetics 168:687–699

Hurst GDD, Werren JH (2001) The role of selfish genetic elements in eukaryotic evolution. Nature Rev Genet 2:597–606

Jauhar PP (1990) Multidisciplinary approach to genome analysis in the diploid species, *Thinopyrum bessarabicum* and *Th. elongatum* (*Lophopyrum elongatum*), of the Triticeae. Theor Appl Genet 80:523–536

Jauhar PP (1995) Meiosis and fertility of F_1 hybrids between hexaploid bread wheat and decaploid tall wheatgrass (*Thinopyrum ponticum*). Theor Appl Genet 90:865–871

Jauhar PP, Peterson TS, Xu SS (2009) Cytogenetic and molecular characterization of a durum alien disomic addition line with enhanced tolerance to Fusarium head blight. Genome 52:467–483

Kibirige-Sebunya I, Knott DR (1983) Transfer of stem rust resistance to wheat from an *Agropyron* chromosome having a gametocidal effect. Can J Genet Cytol 25:215–221

Kim NS, Armstrong K, Knott DR (1993) Molecular detection of *Lophopyrum* chromatin in wheat-*Lophopyrum* recombinants and their use in physical mapping of chromosome 7D. Theor Appl Genet 85:561–567

Knott DR (1961) The inheritance of rust resistance. VI. The transfer of stem rust resistance from *Agropyron elongatum* to common wheat. Can J Plant Sci 41:109–123

Knott DR (1968) Translocations involving *Triticum* chromosomes and *Agropyron* chromosomes carrying rust resistance. Can J Genet Cytol 10:695–696

Knott DR (1971) The transfer of genes for disease resistance from alien species to wheat by induced translocations. In: Mutation breeding for disease resistance. IAEA, Vienna, pp 67–77

Knott DR (1980) Mutation of a gene for yellow pigment linked to *Lr19* in wheat. Can J Genet Cytol 22:651–654

Knott DR (1984) The genetic nature of mutations of a gene for yellow pigment linked to *Lr19* in 'Agatha' wheat. Can J Genet Cytol 26:392–393
Knott DR (1989a) Genetic analysis of resistance. In: The wheat rusts—breeding for resistance. Springer, Berlin, pp 58–83
Knott DR (1989b) The effect of transfers of alien genes for leaf rust resistance on the agronomic and quality characteristics of wheat. Euphytica 44:65–72
Knott DR, Dvorak J, Nanda JS (1977) The transfer to wheat and homoeology of an *Agropyron elongatum* chromosome carrying resistance to stem rust. Can J Genet Cytol 19:75–79
Kolmer JA, Jin Y, Long DL (2007) Wheat leaf and stem rust in the United States. Austr J Agric Res 58:631–638
Kong L, Anderson JM, Ohm HW (2008) Segregation distortion in common wheat of a segment of *Thinopyrum intermedium* chromosome 7E carrying *Bdv3* and development of a *Bdv3* marker. Plant Breed 128:591–597
Kumar S, Gill BS, Faris JD (2007) Identification and characterization of segregation distortion loci along chromosome 5B in tetraploid wheat. Mol Genet Genomics 278:187–196
Kuzmanovic L (2011) Cytogenetic, physiological and agronomic characterization of wheat-*Thinopyrum ponticum* recombinant lines carrying relevant breeding traits. PhD Thesis, Department of Agriculture, Forestry, Nature and Energy (DAFNE), University of Tuscia, Viterbo, Italy
Kuzmanović L, Gennaro A, Benedettelli S, Dodd IC, Quarrie SA, Ceoloni C (2013) Structural-functional dissection and characterization of yield-contributing traits originating from a group 7 chromosome of the wheatgrass species *Thinopyrum ponticum* after transfer into durum wheat. J Exp Bot. doi:10.1093/jxb/ert393
Larkin PJ, Banks PM, Lagudah ES et al (1995) Disomic *Thinopyrum intermedium* addition lines in wheat with barley yellow dwarf virus resistance and with rust resistances. Genome 38:385–394
Li D, Zhang XY (2002) Physical localization of the 18S-5.8S-26S rDNA sequence analysis of ITS regions in *Thinopyrum ponticum* (Poaceae: Triticeae): implications for concerted evolution. Ann Bot 90:445–452
Li H, Chen Q, Conner RL et al (2003) Molecular characterization of a wheat-*Thinopyrum ponticum* partial amphiploid and its derivatives for resistance to leaf rust. Genome 46:906–913
Li H, Wang X (2009) *Thinopyrum ponticum* and *Thinopyrum intermedium*: the promising source of resistance to fungal and viral diseases of wheat. J Genet Genomics 36:557–565
Li Z, Li B, Tong Y (2008) The contribution of distant hybridization with decaploid *Agropyron elongatum* to wheat improvement in China. J Genet Genomics 35:451–456
Liu S, Gao X, Xia G (2008) Characterizing HMW-GS alleles of decaploid *Agropyron elongatum* in relation to evolution and wheat breeding. Theor Appl Genet 116:325–334
Liu S, Yu L-X, Singh RP et al (2010) Diagnostic and co-dominant PCR markers for wheat stem rust resistance genes *Sr25* and *Sr26*. Theor Appl Genet 120:691–697
Löve A (1982) Generic evolution of the wheatgrasses. Biol Zentralbl 101:199–212
Lyttle TW (1991) Segregation distorters. Ann Rev Genet 25:511–557
Marais GF (1992a) The modification of a common wheat-*Thinopyrum distichum* translocated chromosome with a locus homoeoallelic to *Lr19*. Theor Appl Genet 85:73–78
Marais GF (1992b) Gamma-irradiation induced deletions in an alien chromosome segment of the wheat 'Indis' and their use in gene mapping. Genome 35:225–229
Marais GF (1992c) Genetic control of a response to the segregation distortion allele, *Sd-1d*, in the common wheat line 'Indis'. Euphytica 60:89–95
Marais GF, Marais AS, Groenwald JZ (2001) Evaluation and reduction of *Lr19*–149, a recombined form of the *Lr19* translocation of wheat. Euphytica 121:289–295
Mares DJ, Campbell AW (2001) Mapping components of flour and noodle colour in Australian wheat. Aust J Agric Res 52:1297–1309
McIntosh RA, Dyck PL, Green GJ (1976) Inheritance of leaf rust and stem rust resistances in wheat cultivars Agent and Agatha. Aust J Agric Res 28:37–45

McIntosh RA, Wellings CR, Park RF (1995) Wheat rusts: an atlas of resistance genes. CSIRO Publications, East Melbourne

McIntyre CL, Clarke BC, Appels R (1988) Amplification and dispersion of repeated DNA sequences in the Triticeae. Plant Syst Evol 160:39–59

Miller SS, Watson EM, Lazebnik J et al (2011) Characterization of an alien source of resistance to Fusarium head blight transferred to Chinese Spring wheat. Botany 89:301–311

Monneveaux P, Reynolds MP, Gonzalez Aguilar J, Singh RP (2003) Effects of the *7DL.7Ag* translocation from *Lophopyrum elongatum* on wheat yield and related morphophysiological traits under different environments. Plant Breed 122:379–384

Moore G, Shaw P (2009) Improving the chances of finding the right partner. Curr Opin Genet Dev 19:99–104

Mullan DJ, Mirzaghaderi G, Walker E et al (2009) Development of wheat–*Lophopyrum elongatum* recombinant lines for enhanced sodium 'exclusion' during salinity stress. Theor Appl Genet 119:1313–1323

Nasuda S, Friebe B, Gill BS (1998) Gametocidal genes induce chromosome breakage in the interphase prior to the first mitotic cell division of the male gametophyte in wheat. Genetics 149:1115–1124

Nelson JC, Sorrells ME, Van Deynze AE et al (1995) Molecular mapping of wheat: major genes and rearrangements in homoeologous groups 4, 5 and 7. Genetics 141:721–731

Oliver RE, Cai X, Xu SS et al (2005) Wheat-alien species derivatives: a novel source of resistance to Fusarium head blight in wheat. Crop Sci 45:1353–1360

Oliver RE, Xu SS, Stack RW et al (2006) Molecular cytogenetic characterization of four partial wheat-*Thinopyrum ponticum* amphiploids and their reactions to *Fusarium* head blight, tan spot, and *Stagonospora nodorum* blotch. Theor Appl Genet 112:1473–1479

Park RF (2007) Stem rust of wheat in Australia. Austr J Agric Res 58:558–566

Parker GD, Chalmers KJ, Rathjen AJ, Langridge P (1998) Mapping loci associated with flour colour in wheat (*Triticum aestivum* L.). Theor Appl Genet 97:238–245

Plotnikova LY (2008) Cellular features of immune reaction of common wheat with *Lr19* gene to brown rust fungus infection. Tsitologiya 50:124–131

Pozniak C, Knox R, Clarke F, Clarke J (2007) Identification of QTL and association of a phytoene synthase gene with endosperm colour in durum wheat. Theor Appl Genet 114:525–537

Prins R, Groenewald JZ, Marais GF et al (2001) AFLP and STS tagging of *Lr19*, a gene conferring resistance to leaf rust in wheat. Theor Appl Genet 103:618–624

Prins R, Marais GF (1998) An extended deletion map of the *Lr19* translocation and modified forms. Euphytica 103:95–102

Prins R, Marais GF (1999) A genetic study of the gametocidal effect of the *Lr19* translocation of common wheat. SAJ Plant Soil 16:10–14

Prins R, Marais GF, Janse BJH et al (1996) A physical map of the *Thinopyrum*-derived *Lr19* translocation. Genome 39:1013–1019

Prins R, Marais GF, Pretorius ZA et al (1997) A study of modified forms/ of the *Lr19* translocation of common wheat. Theor Appl Genet 95:424–430

Qi L, Friebe B, Zhang P, Gill BS (2007) Homoeologous recombination, chromosome engineering and crop improvement. Chromos Res 15:3–19

Quarrie SA, Steed A, Calestani C et al (2005) A high-density genetic map of hexaploid wheat (*Triticum aestivum* L.) from the cross Chinese Spring × SQ1 and its use to compare QTLs for grain yield across a range of environments. Theor Appl Genet 110:865–880

Quarrie SA, Pekic-Quarrie S, Radosevic R et al (2006) Dissecting a wheat QTL for yield present in a range of environments: from the QTL to candidate genes. J Exp Bot 57:2627–2637

Reynolds MP, Calderini DF, Condon AG, Rajaram S (2001) Physiological basis of yield gains in wheat associated with the *Lr19* translocation from *Agropyron elongatum*. Euphytica 119: 137–141

Röder MS, Korzun V, Wendehake K et al (1998) A microsatellite map of wheat. Genetics 149: 2007–2023

Saini RG, Kaur L, Kaur M (1998) Adult plant leaf rust (*Puccinia recondita tritici*) resistance of known *Lr* genes against three virulence variants of race 77 from Indian sub-continent. Indian J Agric Sci 68:776–779

Saintenac C, Falque M, Martin OC et al (2009) Detailed recombination studies along chromosome 3B provide new insights on crossover distribution in wheat (*Triticum aestivum* L.). Genetics 181:393–403

Scoles GJ, Kibirige-Sebunya IN (1983) Preferential abortion of gametes in wheat induced by an *Agropyron* chromosome. Can J Genet Cytol 25:1–6

Sears ER (1973). *Agropyron*-wheat transfers induced by homoeologous pairing. In: Sears ER, Sears LMS (eds) Proceeding of the 4th International Wheat Genetics Symposium. Univ. of Missouri, Columbia, MO, USA, pp 191–199

Sears ER (1978) Analysis of wheat-*Agropyron* recombinant chromosomes. In: Sanchez-Monge E, Garcia-Olmedo F (eds) Interspecific Hybridization in Plant Breeding, Proceedings of the 8th EUCARPIA Congress, Madrid, Spain. Escuela Técnica Superior de Ingenieros Agrónomos, Ciudad Universitaria, Madrid, pp 63–72

Sepsi A, Molnár I, Szalay D, Molnár-Lang M (2008) Characterization of a leaf rust-resistant wheat-*Thinopyrum ponticum* partial amphiploid BE-1, using sequential multicolour GISH and FISH. Theor Appl Genet 116:825–834

Sharma D, Knott DR (1966) The transfer of leaf rust resistance from *Agropyron* to *Triticum* by irradiation. Can J Genet Cytol 8:137–143

Sharma H, Ohm H, Goulart L et al (1995) Introgression and characterization of barley yellow dwarf virus resistance from *Thinopyrum intermedium* into wheat. Genome 38:406–413

Shen X, Ohm H (2007) Molecular mapping of *Thinopyrum*-derived Fusarium head blight resistance in common wheat. Mol Breed 20:131–140

Singh A, Reimer S, Pozniak CJ et al (2009) Allelic variation at *Psy1-A1* and association with yellow pigment in durum wheat grain. Theor Appl Genet 118:1539–1548

Singh RP, Huerta-Espino J, Rajaram S, Crossa J (1998) Agronomic effects from chromosome translocations 7DL.7Ag and 1BL.1RS in spring wheat. Crop Sci 38:27–33

Sivasamy M, Kumar J, Menon MK, Tomar SMS (2010) Developing elite, durable disease resistant wheat cultivars combining high grain yield and end-use quality by introgressing effective genes employing conventional and modern breeding approaches. Annu Wheat Newsl 56:87–94

Somers DJ, Isaac P, Edwards K (2004) A high-density microsatellite consensus map for bread wheat (*Triticum aestivum* L.). Theor Appl Genet 109:1105–1114

Sourdille P, Singh S, Cadalen T et al (2004) Microsatellite-based deletion mapping system for the establishment of genetic map-physical map relationships in wheat. Funct Integr Genomics 4:12–25

Tang S, Li Z, Jia X, Larkin PJ (2000) Genomic in situ hybridization (GISH) analyses of *Thinopyrum intermedium*, its partial amphiploid Zhong 5, and disease-resistant derivatives in wheat. Theor Appl Genet 100:344–352

Taylor DR, Ingvarsson PK (2003) Common features of segregation distortion in plants and animals. Genetica 117:27–35

Tsitsin NV (1965) Remote hybridization as a method of creating new species and varieties of plants. Euphytica 14:326–330

van der Linden CD, Wouters DCAE, Mihalka V et al (2004) Efficient targeting of plant disease resistance loci using NBS profiling. Theor Appl Genet 109:384–393

Wang Q, Xiang J, Gao A et al (2010a) Analysis of chromosomal structural polymorphisms in the St, P, and Y genomes of Triticeae (Poaceae). Genome 53:241–249

Wang RR-C (1992) Genome relationships in the perennial Triticeae based on diploid hybrids and beyond. Hereditas 116:133–136

Wang RR-C (2011) *Agropyron* and *Psathyrostachys*. In: Kole C (ed) Wild crop relatives: genomic and breeding resources, cereals. Springer, Berlin, pp 77–108

Wang RR-C, Larson SR, Jensen KB (2010b) Analyses of *Thinopyrum bessarabicum*, *T. elongatum*, and *T. junceum* chromosomes using EST-SSR markers. Genome 53:1083–1089

Wang RR-C, Wei J-Z (1995) Variations of two repetitive DNA sequences in several Triticeae genomes revealed by polymerase chain reaction and sequencing. Genome 38:1221–1229

Wang RR-C, Zhang XY (1996) Characterization of the translocated chromosome using fluorescence in situ hybridization and random amplified polymorphic DNA on two *Triticum aestivum-Thinopyrum intermedium* translocation lines resistant to wheat streak mosaic or barley yellow dwarf virus. Chromos Res 4:583–587

William MDHM, Mujeeb-Kazi A (1995) Biochemical and molecular diagnostics of *Thinopyrum bessarabicum* chromosomes in *Triticum aestivum* germ plasm. Theor Appl Genet 90:952–956

Xu SS, Jin Y, Klindworth D.L, Wang RR-C, Cai X (2009) Evaluation and characterization of seedling resistances to stem rust Ug99 races in wheat–alien species derivatives. Crop Sci 49:2167–2175

Xue S, Zhang Z, Lin F et al (2008) A high-density intervarietal map of the wheat genome enriched with markers derived from expressed sequence tags. Theor Appl Genet 117:181–189

Zhang HB, Dvorak J (1990) Characterization and distribution of an interspersed repeated nucleotide sequence from *Lophopyrum elongatum* and identification of a segregation distortion with it. Genome 33:927–936

Zhang W, Dubcovsky J (2008) Association between allelic variation at the *Phytoene synthase 1* gene and yellow pigment content in the wheat grain. Theor Appl Genet 116:635–645

Zhang W, Lukaszewski AJ, Kolmer J et al (2005) Molecular characterization of durum and common wheat recombinant lines carrying leaf rust resistance (*Lr19*) and yellow pigment (*Y*) genes from *Lophopyrum ponticum*. Theor Appl Genet 111:573–582

Zhang X, Shen X, Hao Y et al (2011) A genetic map of *Lophopyrum ponticum* chromosome 7E, harbouring resistance genes to Fusarium head blight and leaf rust. Theor Appl Genet 122: 263–270

Zhang XY, Zhensheng L, Shuyang C (1992) Production and identification of three 4Ag(4D) substitution lines of *Triticum aestivum-Agropyron*: relative transmission rate of alien chromosomes. Theor Appl Genet 83:707–714

Zhang XY, Dond YS, Wang RR-C (1996a) Characterization of genomes and chromosomes in partial amphiploids of the hybrids of *Triticum aestivum* x *Thinopyrum ponticum* by *in situ* hybridization, isozyme analysis, and RAPD. Genome 39:1062–1071

Zhang XY, Koul A, Petroski R et al (1996b) Molecular verification and characterization of BYDV-resistant germ plasms derived from hybrids of wheat with *Thinopyrum ponticum* and *Th. intermedium*. Theor Appl Genet 93:1033–1039

Chapter 15
Identification and Implementation of Resistance: Genomics-Assisted use of Genetic Resources for Breeding Against Powdery Mildew and Stagonospora Nodorum Blotch in Wheat

Liselotte L. Selter, Margarita Shatalina, Jyoti Singla and Beat Keller

Abstract Wheat belongs to the three most important cereal crops of the world and is grown under a wide variety of climatic and agricultural conditions. Fungal pathogens represent the most relevant biotic stresses for wheat. These include different rust species, powdery mildew, leaf spots, as well as a number of other diseases that result in reduced grain yield and quality. Recently developed genomic tools allow new approaches to improve breeding for resistance to these pathogens based on a more efficient use of genetic resources. In this chapter, we will focus on the powdery mildew and Stagonospora nodorum blotch diseases and discuss the successful identification of wheat genes determining the outcome of pathogen-host interaction and the development of perfect markers for them. Genomic approaches, including gene cloning, allele mining, transcriptomics and comparative genomics have greatly changed and improved our understanding of molecular wheat-powdery mildew interactions. For the necrotrophic pathogen *Stagonospora nodorum* much of the interaction was found to be based on pathogen toxins and host susceptibility genes. The work on specific gene-for-gene interactions opened new possibilities for more efficient resistance breeding. In addition, the molecular identification of quantitatively acting resistance loci in wheat has made important progress, although only few such genes have been cloned, only one of them each against mildew and Stagonospora nodorum blotch. However, even at this early stage it can be foreseen that the new knowledge might revolutionize breeding for durable resistance in the near future. The progress made towards a whole genome sequence of wheat together with ongoing developments of high throughput techniques provides a completely new perspective on resistance breeding against these two diseases.

B. Keller (✉) · L. L. Selter · M. Shatalina · J. Singla
Institute of Plant Biology, University of Zurich, Zollikerstrasse 107,
8008 Zurich, Switzerland
e-mail: bkeller@botinst.uzh.ch

L. L. Selter
e-mail: liselotte.selter@gmx.net

M. Shatalina
e-mail: margarita.shatalina@access.uzh.ch

J. Singla
e-mail: jyoti.singla@uzh.ch

R. Tuberosa et al. (eds.), *Genomics of Plant Genetic Resources,*
DOI 10.1007/978-94-007-7575-6_15,

15.1 Introduction

15.1.1 Emerging Challenges for Wheat Resistance Breeding Against Powdery Mildew and Stagonospora Nodorum Blotch: Changes in Agricultural Practice, Climate Change and Pathogen Adaptation

The powdery mildew disease occurs in crop growing regions worldwide. Before the green revolution, powdery mildew was found predominantly in regions with a cool, humid and semi-continental climate. However, with the introduction of new agricultural practices and intensified crop production during the last decades, powdery mildew has gained importance also in the more arid crop growing regions. Today, economically relevant powdery mildew epidemics cause serious yield losses in the cool and humid areas of China, North and South America, Northern Europe as well as in North and East Africa. The widespread use of irrigation systems and nitrogen fertilizers for yield improvement has created favourable conditions for this particular pathogen in additional agro-ecosystems. Wheat farmers can considerably influence powdery mildew epidemics by adapting appropriate agricultural practices, such as choosing the right sowing period, lower population densities or lower use of fertilizers. Breeding for genetic resistance to powdery mildew is nevertheless considered the most effective disease control strategy and will be discussed in detail below. Cultivar mixtures and low density planting are good strategies to slow disease development, but both have their specific problems and have been employed only occasionally so far. The application of foliar fungicides is often chosen as a last strategy if cultural practices are not able to control powdery mildew development. However, an intense use of fungicides can lead to fungicide resistance in the pathogen. This has become a major concern in Europe (Wolfe 1984).

Stagonospora nodorum blotch (SNB) affects wheat grown under humid conditions and mild temperatures in Europe, South America, Central Asia and North Africa. In North America, China and Europe, the genetic diversity of *S. nodorum* populations is very high. It was shown that populations in Europe, North America and China have no or relatively little subdivision and serve as donors for disease distribution to other continents (Stukenbrock et al. 2006). Therefore, the suggested center of origin for *S. nodorum* in the Fertile Crescent coincides with the center of origin for wheat (Balter 2007; Burger et al. 2008). The distribution of *S. nodorum* is mainly human-mediated, which is the main way of disease transport from North America, Europe and China to other parts of the world (Stukenbrock et al. 2006). A wide range of fungicides are efficiently applied in the areas with SNB infections. Reports about fungicide-resistant isolates of *S. nodorum* are very rare. However, the possibility of their emergence remains a threat in regions with extensive fungicide application (Oliver et al. 2012).

Modern bread wheat is a temperate crop adapted to regions with annual rainfall between 30 and 90 cm. It is nowadays cultivated on both hemispheres under a wide range of climatic conditions and different soils, making up 17 % of all crop acreage. Although the impact of climate change on crops shows complex regional patterns, significant yield losses have been predicted using the worst case CO_2 emission scenario of the Intergovernmental Panel on Climate Change (Luck et al. 2011). Climate change will not only differentially affect wheat cultivars in their geographic distribution and their growth, but also pathogens. Biotrophic pathogens such as powdery mildews, highly depend on the plant's health and its water and nitrogen status (Olesen et al. 2000). As plant disease development and spreading is influenced predominantly by increased atmospheric CO_2 levels, heavy rains, increased humidity, drought and warmer winter temperatures (Cannon 1998; Chakraborty et al. 2000; Pimentel et al. 2001; Berry et al. 2002; Anderson et al. 2004) we can expect that the lifecycle of some pathogens will be limited by increasing temperatures, while other pathogen species might respond positively to the same climatic changes. Changes in global minimum temperatures and rainfall patterns will presumably cause shifts in growing seasons of certain wheat cultivars and alter the land use of specific crops. This might then in turn lead to the occurrence of novel plant-pathogen interactions through the introduction of new host genotypes, new pathogens or both to a specific agro-ecosystem. In addition, temperature changes in critical periods of host infection might reduce the effectiveness of resistance genes, as it has been shown that some *R* genes against powdery mildew are known to be temperature-sensitive (Ge et al. 1998).

Clearly, based on the considerations described above, we can expect that climate change will have multiple, highly complex effects on plant disease epidemiology and the consequences on yield are difficult to predict. As today's agriculture primarily aims at crop yield improvement and breeding programs mainly focus on cultivars adapted to longer growth periods, drought and stress tolerance, it is of great importance to establish efficient disease screening methods which allow to monitor changing disease epidemics. This is because pathogens are not only important yield-reducing factors, but due to their short generation times also act as early indicators of environmental changes (Newton et al. 2011). Intensifying the research of climate change effects on plant-pathogen systems will certainly allow an improvement of the disease management practices necessary for a sustainable agriculture.

15.1.2 Wheat as the Host Plant for Blumeria graminis f.sp. tritici and Stagonospora nodorum

Bread wheat (*Triticum aestivum*, $2n = 6x = 42$, AABBDD) belongs to the four most important cereal crops in modern agriculture (http://www.FAOSTAT.org). The FAO estimates that 682.5 million t of wheat was harvested in the year 2011. Bread wheat accounts for approximately 20 % of the totally consumed human food calories and provides the major staple food for 40 % of the human population, predominantly

in Europe, North America and the western and northern parts of Asia (Peng et al. 2011). The origin of modern bread wheat lies in a region of the Near East known as the Fertile Crescent which covers parts of south-eastern Turkey, Israel, Syria, Iraq and Jordan. There, wild wheat progenitors such as Einkorn or Emmer (He et al. 2009), were among the first cereals subjected to human selection 10,000 years ago (Charmet 2011). Hexaploid bread wheat originated approximately 9,000 years ago from a hybridization event between the allotetraploid domesticated Emmer wheat (*T. turgidum spp. dicoccoides* (2n = 4x = AABB) and the diploid wild goatgrass *Ae. tauschii* (2n = 2x = DD). Bread wheat and its wild progenitors were selected by the first farmers for agriculturally advantageous traits in the specific agroecological system where domestication occurred. These traits also included disease resistance to fungal pathogens.

15.1.3 Characteristics of Powdery Mildew and Stagonospora Nodorum Blotch Diseases in Agricultural Systems

The powdery mildew pathogen of barley, *Blumeria graminis* f.sp. *hordei,* was found to have evolved on wild grasses in the Middle East (Koltin and Kenneth 1970; Wolfe 1984). Comparative genome analysis of wheat and barley powdery mildew revealed that these two *formae speciales* diverged about 10 million years ago, after divergence of their respective hosts (Oberhaensli et al. 2011). This suggests that the wheat powdery mildew, *Blumeria graminis* f.sp. *tritici*, originates from an ancestral pathogen which initially colonized ancestors of both wild wheat and barley. There is evidence that wheat powdery mildew originated and co-existed with wild wheat long before their domestication (Wicker et al. 2013).

Wheat yield losses caused by the two wheat fungal pathogens powdery mildew and *Stagonospora nodorum* are difficult to estimate. In controlled experimental environments, it is feasible to measure yield losses, but on farmer's fields, crop health and actual losses are significantly different from experimental calculations. Oerke et al. (1994) estimated that collectively all wheat diseases cause annual grain losses of about 12.4 %, including all developed and developing countries. Disease epidemics of the two described wheat pathogens of this chapter, powdery mildew and SNB, depend mostly on three factors: prevalence of inoculum, the genetic constitution of grown cultivars, and to a large extent on environmental conditions (Duveiller et al. 2007). The changes in agricultural practices during the last decades have led to changes at the microclimate level in wheat growing areas. In order to increase productivity, genetically uniform varieties are planted in dense stands. These genotypes contain often semi-dwarf varieties and have a high tillering density, thus increasing the humidity within the crop canopy. In addition, the regular application of nitrogen fertilizers and irrigation creates a microclimate which is highly favourable for the spreading of biotrophic fungal diseases (Sharma et al. 2004).

Systematic reports on SNB epidemics are lacking from most of the wheat growing areas. The most complete dataset is available from Rothamsted Broadbalk experiment archive (UK). There, wheat leaf samples have been collected for nearly 160 years (from 1844 to 2003) and used to estimate the epidemics of SNB and Septoria tritici blotch (STB) caused by *Mycosphaerella graminicola* (Bearchell et al. 2005; Shaw et al. 2008). The predominance of SNB was shifted towards *M. graminicola* after 1970. Bearchell et al. (2005) linked this shift in predominance of STB over SNB to the decrease in SO_2 emissions after 1970. Another suggested reason is that before 1970 the widely used cultivars had good partial resistance to STB, but not to SNB. Later, large efforts were made to introduce SNB resistance in newly released cultivars (Arraiano et al. 2009). Additionally, in Western Australia in regions where SNB dominates SO_2 pollution is very low. Therefore, the shift between SNB and STB epidemics is likely caused by a combination of factors (Oliver et al. 2012).

Breeding for resistant wheat varieties is the most effective strategy to counteract fungal diseases. Despite an overall good success of resistance breeding, changes in wheat genotypes as well as pathogen races are frequent, making breeding for disease resistance a continuous task. It is obvious that a better understanding about the molecular basis of disease resistance in wheat can contribute significantly to improve strategies in achieving resistance and to make resistance breeding faster and more efficient. Importantly, this can be achieved through the use of recently developed genomic tools such as high-throughput platforms for molecular marker analysis and genotyping in combination with classical breeding methods and increasing knowledge on the genomes of wheat and its relatives. Various genomics-assisted breeding approaches such as marker-assisted selection (MAS), association mapping, QTL identification and MAS for them, as well as genome -wide association studies have been successfully utilized in modern plant breeding for the development of improved crop varieties. The limited number of molecularly cloned resistance genes/QTL in wheat can be explained by the genetic complexity observed in this species. The large wheat genome size and the high amount of repetitive DNA (80 %) makes map-based cloning in wheat a challenging task. Nevertheless, some disease resistance genes (*Lr1*, *Lr10*, *Lr21, Lr34/Yr18/Pm38, Yr36*, *Pm3b and Tsn1*) have been cloned from hexaploid wheat (Feuillet et al. 2003; Huang et al. 2003; Yahiaoui et al. 2004; Cloutier et al. 2007; Qiu et al. 2007; Fu et al. 2009; Krattinger et al. 2009, Faris et al. 2010) using sub-genome chromosome walking techniques and comparative genomics.

In this chapter we will describe how genomic approaches for wheat resistance breeding against powdery mildew and Stagonospora nodorum leaf blotch have been used in the last years: the work discussed includes classical map-based cloning approaches but also new strategies such as allelemining and the use of transcriptomics and finally the new and exciting field of transgenic use of modified resistance genes.

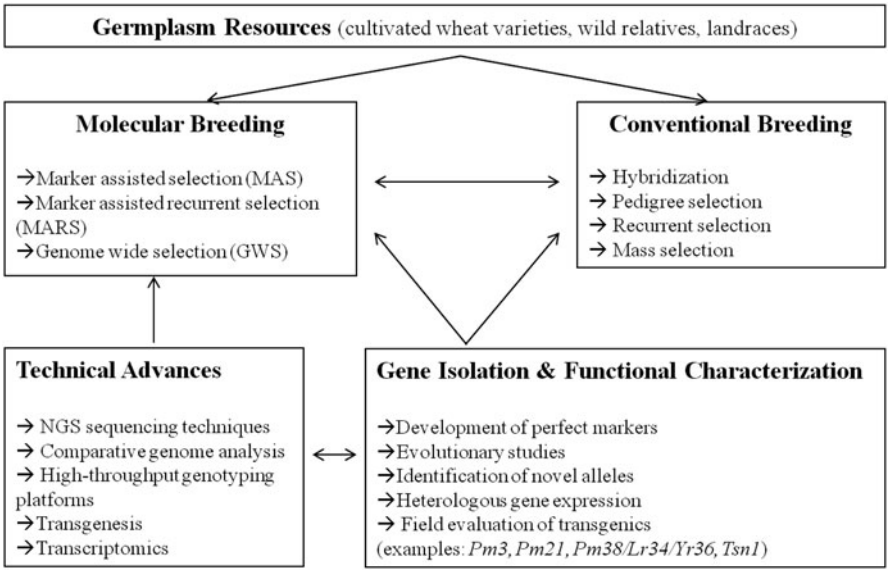

Fig. 15.1 Scheme of genomics-assisted breeding

15.2 Genomics-Assisted Breeding by Cloning of Major Resistance Genes

Major race-specific resistance genes can provide plants with a level of disease resistance which is close to immunity. However, fungal pathogens are fast evolving pathogens, which under selection pressure can rapidly adapt to overcome plant resistance mechanisms. Thus, there is a strong need to (i) identify new and durable sources of genetic resistance in order to avoid an erosion of the current pool of agriculturally important resistance genes and to find (ii) new and innovative ways to use the known resistance genes in a more durable way (Fig. 15.1).

15.2.1 Map-Based Cloning of Powdery Mildew Resistance Genes

15.2.1.1 Wheat Powdery Mildew: A Strictly Biotrophic Pathogen

Blumeria graminis f.sp. *tritici*, the causal agent of wheat powdery mildew is a highly specific pathogen which grows only on wheat species. It belongs to the obligate biotrophic pathogens which fully depend on the integrity of the invaded host plant cell (Horbach et al. 2011) to accomplish all important stages of pathogenesis such as attachment, host recognition, penetration and proliferation (Mendgen and Hahn 2002). The infection process starts when a spore lands on a leaf surface, germinates

and forms a primary and appressorial germ tube. The appressorium penetrates the cell wall using mechanical force and cell walldegrading enzymes, and invaginates the plant cell by forming a special feeding structure—the haustorium, which is surrounded by an extrahaustorial plasma membrane. The haustorium is not only required for nutrient supply, but is also important for signalling, communication and prevention of recognition by the host (Perfect and Green 2001; Horbach et al. 2011). Since it is essential for the pathogen to keep the host cell alive, biotrophic fungi suppress the programmed cell death induced at the infection site—a defense response known as hypersensitive reaction (HR). This defense suppression is possibly the result of the release of effector proteins during the penetration process (Panstruga 2003). Only very recently, the genome sequence of the barley powdery mildew became available (Spanu et al. 2010) and a genome sequence of the wheat powdery mildew can be expected in the near future (Wicker et al. 2013). This genomic sequence information, provides an extremely valuable tool to gain a better understanding of the biology of the powdery mildew pathogen, the factors required for biotrophy as well as virulence determinants. In addition, comparative genome analysis between the wheat and barley powdery mildews will allow an improvement of our understanding of host specialization in these diseases. Only by improved knowledge on all the components of the host-pathogen interaction, we will be able to develop rational resistance improvement based on molecular interactions in the future.

To date, 61 powdery mildew resistance genes including three recessive genes (*Pm5*, *Pm9* and *Pm26*) have been genetically described. They confer resistance against specific races of the pathogen and have been identified and mapped to 46 loci in the wheat genome (He et al. 2009; Hua et al. 2009; Luo et al. 2009; Huang et al. 2012). Out of these, Lillemo et al. (2008) identified two race non-specific genes, *Pm38* and *Pm39* which confer partial resistance. Among these 61 genes, only *Pm3b* (Yahiaoui et al. 2004), *Pm21* (Cao et al. 2011) and *Pm38* (Krattinger et al. 2009) have been cloned so far.

15.2.1.2 Wheat Genomics at Different Ploidy Levels Allows the Isolation of the *Pm3* Powdery Mildew Resistance Alleles in Wheat

A map-based cloning approach was used to isolate the *Pm3b* gene that controls powdery mildew resistance in the hexaploid wheat landrace Chul. The powdery mildew resistance gene was mapped genetically to the distal end of the short arm of chromosome 1A in 1,340 plants of a Chul x Frisal derived F2 population. Physical mapping was performed by using BAC libraries developed from the diploid wheat *T. monococcum* cv. DV92 and the tetraploid *T. durum* cv. Langdon wheat. By using these BAC libraries, Yahiaoui et al. (2004) proved the usefulness of exploiting wheat genomes with different ploidy levels, and combined sub-genome chromosome walking with haplotype analysis. The sub-genome chromosome walking between the three wheat species revealed dissimilarities in the haplotype structures at the *Pm3* locus. Haplotype similarity was found between the *durum* wheat cv. Langdon and the susceptible hexaploid parent Frisal, whereas only partial similarity between the haplotypes of

T. monococcum cv. DV92 and the resistant parent Chul was observed. This led to the isolation of the *Pm3b* gene from the hexaploid wheat donor line by deriving low copy probes from the conserved resistance-gene-like sequences in both genomes using long-range PCR (Yahiaoui et al. 2004). Validation of the candidate gene was done by γ- irradiated mutant analysis. Molecular analysis of 13 independent mutants showed six different deletion patterns. One mutant without any major deletion at the *Pm3* locus showed a single base pair deletion in the coding region of the candidate gene resulting in loss of expression as demonstrated by RT-PCR. Hence, the candidate gene for *Pm3b* could be confirmed.

The rapid resistance response occurring in leaf epidermal cells in the case of an incompatible interaction between wheat and powdery mildew leads to a termination of pre-haustorial fungal growth. This hypersensitive response provides the basis to study resistance gene function by a particle bombardment based, transient transformation of leaf epidermal cells. Using the GUS reporter gene (Schweizer et al. 1999; Douchkov et al. 2005), transformed cells undergoing an active defense response can be identified by co-bombarding the candidate gene with the GUS reporter plasmid. Thus, this functional assay does not require the time consuming procedure of generating stably transformed wheat plants. With this transient transformation assay *Pm3b* was functionally validated and assigned to the biggest class of R gene family, the CC-NBS-LRR proteins. It encodes a domain with 28 well conserved LRR domains and a protein of 1,415 amino acids. Subsequently, additional 7 *Pm3* alleles (*Pm3a-3 g*) were identified from other genetic backgrounds providing race-specific resistance to a different subsets of powdery mildew isolates, and used to develop functional allele-specific markers for germplasm screening (Tommasini et al. 2006) (Table 15.1).

15.2.1.3 Cloning of *Pm21*: Integration of Map-Based Cloning and Gene Expression Analysis to Isolate the *Pm21* Gene from a Non-Recombining Genetic Region

Gene expression analysis has complemented map-based cloning approaches and helped to identify a second powdery mildew resistance gene from wheat. Cao et al. (2011) used a high-throughput strategy of GeneChip microarray analysis in combination with genetic mapping to isolate *Pm21*, an important source of durable and broad spectrum resistance to wheat powdery mildew. *Pm21* was originally transferred from the short arm of chromosome 6V of the wild wheat relative *Haynaldia villosa* ($2n = 2x = 14$) to cultivated wheat by the development of a 6VS.6AL translocation line. Approaches to isolate *Pm21* by map-based cloning using this translocation line were unsuccessful due to the low chromosome pairing frequency and suppressed recombination between the 6VS chromosome from *H. villosa* and chromosome 6AS from wheat. A GeneChip approach was therefore applied to identify genes that are up-regulated upon *Bgt* infection in *H. villosa* compared to the mock control. Among the 196 differentially expressed genes, four resistance gene analogs (RGAs) were

Table 15.1 A success story of genomics-assisted breeding: the isolation and functional characterization of the wheat powdery mildew resistance gene *Pm3*

Step	Approach	Achievements	Reference
1	Identification of the genetic basis of resistance in a specific background	Chromosomal location of a powdery mildew resistance gene in wheat cv. Amigo	Heun et al. 1990
2	Identification of additional alleles of the resistance gene	Identification of other powdery mildew resistance genes/alleles at the *Pm3* locus in hexaploid wheat	Zeller et al. 1993
3	Development of molecular markers linked to the locus	Development of molecular markers for the different *Pm3* alleles	Hartl et al. 1993 Ma et al. 1994 Bougot et al. 2002
4	Gene isolation and functional validation	Map-based cloning of *Pm3b* from hexaploid wheat using genome analysis at different ploidy levels	Yahiaoui et al. 2004
5	Isolation of additional alleles	Identification of the *Pm3* allelic series in hexaploid wheat	Srichumpa et al. 2005
6	Development of functional allele-specific markers	Development of functional markers specific for the seven *Pm3* resistance alleles and their validation in the bread wheat gene pool	Tommasini et al. 2006
7	Evolutionary studies on the resistance genes/allelic sequences	Studies on the evolution of the *Pm3* alleles	Yahiaoui et al. 2006
		Comparative analysis on the evolution of the *Pm3* locus in three different wheat species and rice	Wicker et al. 2007
		Comparative analysis on the evolution of functional *Pm3* alleles from wild and cultivated wheat	Yahiaoui et al. 2009
8	Identification of novel alleles using additional genetic resources (wild relatives, landraces)	Identification of novel functional *Pm3* alleles using extended wheat genetic resources	Bhullar et al. 2009
		Large scale allele mining of wheat gene bank accessions for novel *Pm3* alleles	Bhullar et al. 2010
9	Functional gene characterization using transgenic technology	Intragenic allele pyramiding using the transgenic technology combines different specificities of wheat *Pm3* resistance alleles	Brunner et al. 2010
10	Field assessment of transgenic plants	Transgenic *Pm3b* wheat lines show resistance to powdery mildew in the field	Brunner et al. 2011
		Transgenic *Pm3* multilines of wheat show increased powdery mildew resistance in the field	Brunner et al. 2012

identified which were selected for further investigation. Using a series of alien deletion and translocation lines these genes were cytogenetically mapped by *in situ* hybridization (FISH). Only one RGA, a putative serine/threonine protein kinase (Stpk-V), was found to localize on chromosome 6VS of *H. villosa*, thus making this the best candidate gene for the *Pm21* resistance activity. Expression of Stpk-V was suggested to alter the function of target proteins by phosphorylation of serine or threonine residues. A significant decrease in the haustorial index was observed when epidermal cells were co-transformed with the *GUS* and the *Stpk-V* gene, in comparison to cells only transformed with the *GUS* gene. Also, transgenic plants expressing the *Stpk-V* gene showed an increased broad-spectrum powdery mildew resistance compared to the controls. Further validation of this gene was provided by virus-induced gene silencing (VIGS), where increased susceptibility was observed in *Stpk-V* silenced wheat and its wild relative. The isolation of *Pm21* sets a promising example for future efforts to identify potentially useful genetic sources from wild species by integration of cytogenetic, molecular and transcriptomic methods.

As discussed above for the *Pm21* gene, until recently high-throughput analysis of transcriptomes relied on the microarray technology (Varshney et al. 2009). Microarray based expression profiling has been successfully used to investigate and compare the transcript patterns in various cell types and organisms, however, tracking genetic diversity at the transcript level using the microarray technology has some limitations: Firstly, microarray technology is limited to already existing sequence information of genomes and their annotation. Thus, the gene content available on the array restricts the expression data which can be collected. Further, sensitivity and specificity can be low. The recent development of next generation sequencing (NGS) techniques allows sequencing of the entire transcriptome at a much higher coverage. Compared to the microarray technology, RNA sequencing also has the advantage of providing an unbiased representation of all transcripts. In addition, rare transcripts or alternative splice variants can be detected, as well as allele specific expression and expressed single nucleotide polymorphisms. Sequence variation at RNA levels is therefore more likely to be detected using next generation transcriptomics. Thus, NGS techniques combined with classical cloning methods serve as potentially useful tools to isolate additional disease resistance genes from wheat in the near future.

15.2.1.4 Allele Mining as a Strategy to Identify Additional and Novel Resistance Sources

The identification of genetic resistance sources in wheat and their combination and accumulation in particular cultivars has greatly contributed to the progress in resistance breeding. Nevertheless, we can presume that a huge portion of beneficial resistance genes in the wheat gene pool remains unexploited (Kumar et al. 2010). Several studies have found that resistance in cultivated wheat could be significantly improved by introducing novel alleles from wild relatives. It was further observed that expression of novel alleles or combinations thereof can vary tremendously depending on the genetic background (McCouch et al. 2007; Cao et al. 2011). Thus, a

great potential exists in finding new resistance sources by re-investigating the large germplasm material of wild progenitors or landraces and expressing the new genes in different genetic backgrounds. With the recent development of NGS technologies, sequence information from several crop species has greatly improved and made publicly available to the research community. Although this will presumably accelerate resistance gene discovery in wheat, our current knowledge about resistance genes is still very limited. Thus, it is even more important to use the existing knowledge on cloned resistance genes and exploit the genome information from germplasm resources in order to identify novel, potentially functional alleles. The cloning of the wheat *Pm3* gene and the molecular characterization of its alleles, together with the development of allele-specific markers, allowed an in-depth investigation of a large set of wheat landraces, aiming at the identification of new, potentially functional *Pm3* alleles (Kaur et al. 2008).

The dissection of naturally occurring variation at a known candidate gene locus is also referred to as "allele mining", a strategy taking advantage of an overall high sequence conservation at a specific locus (Kumar et al. 2010). Initial allele-mining studies focused on identification of sequence variation in coding sequences of important loci. However, with increasing evidence for non-coding regions having large effects on transcript and trait expression, mining for sequence variation in regulatory regions of resistance loci is relevant, too. In "promoter mining", promoter regions instead of gene coding sequences are investigated for sequence variation. Both allele and promoter mining have several important applications in resistance breeding. Superior and novel alleles can be identified, new markers can be developed to allow rapid identification of different haplotypes in marker-assisted selection, and evolutionary studies can be performed as well as expression studies. However, there are major considerations for a successful and efficient allele mining approach: besides the requirement of sufficient genome sequence information, there should be high-throughput techniques available to generate allelic data and efficient bioinformatic tools to identify nucleotide variation. Once novel alleles have been identified, a reliable and rapid system for functional validation of the novel alleles is desired. Besides these technical considerations, the foremost challenge in allele mining is the selection of a manageable and sensible number of genotypes capturing the highest possible sequence variation at a specific locus. One possible strategy is the Focused Identification of Germplasm Strategy (FIGS) which allows the identification of trait-specific sets of accessions with maximum diversity. Assuming that the frequency of the trait of interest is strongly influenced by the environment and is based on adaptive selection processes, accessions are selected according to eco-climatic parameters of their original collection sites (Endresen et al. 2011). To date, FIGS has been successfully used to identify new genetic diversity for resistance against abiotic and biotic stresses and specifically also in the case of *Pm3* based resistance (Bhullar et al. 2009). There, 1,320 accessions from 323 geographic sites with potentially high selection pressure for powdery mildew resistance were selected from a virtual collection of 16,089 accessions, and tested against different powdery mildew isolates (Kaur et al. 2008). Among them, 211 accessions which showed complete or intermediate resistance were further analyzed at the molecular level. 111 landraces which

were positive for a *Pm3* diagnostic fragment, but did not amplify specific markers for the known *Pm3a-Pm3g* alleles, were selected as candidates for potentially new functional *Pm3* alleles. Functional analysis of these 111 candidates used a combination of pathogenicity assays and virus-induced gene silencing (VIGS), and resulted in the identification of seven new functional alleles (*Pm3l-Pm3r*) in addition to previously described alleles. As the FIGS screening set contained accessions from a limited geographic area (with a strong focus on the Near East), a new set including accessions from more diverse locations was screened to investigate *Pm3* diversity in more depth. From a collection of an additional 733 wheat accessions eight new *Pm3* sequences were isolated. From these, two additional novel alleles, originating from Nepal *(Pm3s)* and China *(Pm3t)* respectively could be functionally validated (Bhullar et al. 2010). Thus, the large genebank collections comprising germplasm of wild wheat relatives and landraces provide a great potential to identify new resistance resources. In the case of the *Pm3* alleles, out of 30 different countries most of the functional alleles were isolated from accessions originating from Turkey, Afghanistan, Turkmenistan, China and Nepal (Bhullar et al. 2010). The germplasm derived specifically from these countries therefore has a great potential for further exploration specifically for powdery mildew resistance.

15.2.1.5 Field Assessment of Wheat Lines Carrying a Transgenic *Pm3* Resistance Gene

The molecular isolation of the two powdery mildew resistance genes *Pm3* and *Pm21* also provided the opportunity to modify their expression and investigate their efficiency under natural field conditions using transgenic approaches. This has been described in some detail for the *Pm3* resistance alleles. The question was if transgenic genes, under the control of a constitutive promotor would result in improved resistance, and if mixtures of genotypes with the same genetic background, but containing different *Pm3* alleles (so called multilines) would show enhanced resistance due to a mixture effect.

In order to test the transgenic use of race-specific *R* genes for their effectiveness in the field, transgenic wheat lines over-expressing *Pm3a*, *Pm3b*, *Pm3c*, *Pm3d*, *Pm3f* or *Pm3 g* were analyzed during one to three field seasons. All 12 tested transgenic lines were significantly more resistant than their respective non-transformed sister lines but the *Pm3* lines showed differences in the level of powdery mildew resistance. These differences were possibly caused by the differences in frequency of virulence to the particular *Pm3* allele in the powdery mildew population, *Pm3* expression levels and most likely also allele-specific properties. Half of the transgenic lines revealed additional phenotypes in the field, which were not visible under greenhouse conditions. Besides an increased powdery mildew resistance, three of four independent transgenic events carrying *Pm3b*, two *Pm3f* lines and a *Pm3g* line exhibited a leaf chlorosis phenotype, reduced fertility or a reduced plant height (Brunner et al. 2011). High *Pm3* gene expression levels or PM3 protein accumulation were the most likely,

but not all-embracing explanation for these phenotypes. This work showed the importance of field trials for assessment of agronomically relevant disease resistance. It further showed that the success of a transgenic use of *R*-genes in the field critically depends on optimization of expression levels, for example by using tissue and/or development-specific promoters.

To improve the durability of major *R* genes such as *Pm3*, the multiline strategy has been proven to be effective in small grain crops (Zhu et al. 2000; Mundt 2002). Multilines are seed mixtures of agronomically uniform lines that differ only in a specific trait, mostly disease resistance. Brunner et al. (2012) could show in a multiline field experiment that two-way seed mixtures between transgenic lines carrying *Pm3a*, *Pm3b* or *Pm3d* significantly increased the powdery mildew resistance when compared to the mean of the pure component lines alone. This demonstrates that diversity in a single *R* gene is sufficient to improve resistance levels when used in multilines, most probably through host-diversity effects.

15.2.1.6 The Use of Natural Variation to Make Artificial Resistance Genes with Broadened Specificity

The durability of major *R*-genes can possibly be improved by designing artificial resistance genes exhibiting broadened specificity. A successful example of this strategy was provided by Brunner et al. (2010) where they investigated in detail the powdery mildew isolate recognition spectra of different *Pm3* alleles and identified some alleles with enlarged resistance spectra compared to others. Sequence analysis of the natural variation occurring in the *Pm3* alleles exhibiting broad or narrow resistance spectra, allowed to propose hypotheses on the functional roles of individual protein subdomains. Domain-swap experiments revealed for example that the NB-ARC domain is also playing a role in resistance specificity, although pathogen recognition specificity is mostly determined by the LRR-domain. A chimeric, artificial PM3 protein combining different polymorphic residues of the functional alleles proved that intramolecular pyramiding of different *R*-gene recognition specificities is possible and a new resistance gene with a broader specificity can be made.

15.2.1.7 Molecular Analysis of Quantitative Resistance Against Wheat Powdery Mildew

Race-specific powdery mildew resistance genes based on a gene-for-gene interaction with the corresponding pathogen avirulence genes confer strong and effective resistance. Thus, there has been an extensive use of these race-specific *R* genes during the past decades. In the natural situation, the gene-for-gene relationship reflects a co-evolution between the pathogen and the host, where advantageous polymorphisms for either host resistance or pathogen virulence are balanced and stable. If the factors important for this balance are lost, as it is the case in modern agricultural systems, parasite evolution becomes instable and pathogens evolve at much higher

rates. Thus, host-pathogen dynamics resemble more an arms race and this type of resistance becomes of short duration only (Brown and Tellier 2011). It is therefore of great importance to reduce the opportunities for a pathogen to adapt to crop resistance, for example by increasing the genetic diversity of crops or by taking advantage of resistance genes interacting with costly pathogen avirulence genes. Most importantly, exploring durable or quantitative sources of resistance with a combination of several minor genes can greatly help to control powdery mildew diseases in a durable way. Quantitative resistance, also referred to as slow-mildewing or partial resistance is controlled by several genetic loci. It is also known as adult plant resistance (APR) due to the compatible interaction at all stages of growth combined with low infection frequency, prolonged latency period and reduced sporulation at adult plant stage. A series of studies has been conducted on the identification and mapping of quantitative loci involved in disease resistance in the past few years. The development of reliable selection tools has greatly helped to include APR genes in wheat breeding programs. APRs for powdery mildew have been mapped to all homeologous chromosomes of the wheat genome (Sharma et al. 2011). However, to date, there are very few success stories on the molecular isolation of quantitative resistance genes in plants, one being the isolation of the resistance gene, *Lr34/Yr18/Pm38* in wheat (Krattinger et al. 2009). *Lr34/Yr18/Pm38* presents one of the most important durable, race non-specific, adult plant resistance (APR) gene resources which was first identified in Canada by Dyck et al. (1966). Besides providing resistance to leaf rust, it also confers resistance against stripe rust (*Yr18*) (McIntosh 1992), powdery mildew (*Pm38*) (Spielmeyer et al. 2005; Lillemo et al. 2007), stem rust (Dyck 1987) and tolerance to barley yellow dwarf virus (*Bdv1*) (Ayala et al. 2002). Being an APR in nature, *Lr34/Yr18/Pm38* is most effective in the flag leaves of adult plants which also develop necrotic leaf tips, a morphological marker known as leaf tip necrosis (*Ltn*) associated with the presence of *Lr34/Yr18/Pm38* (Dyck 1991; Singh 1992).

The consensus genetic map of three *Lr34/Yr18/Pm38*– based high resolution mapping populations, marked the target interval of 0.15 cM for the *Lr34/Yr18/Pm38* locus. The complete sequencing of a 363 kb physical target interval from the *Lr34/Yr18/Pm38* containing Chinese Spring cultivar revealed eight open reading frames as candidate genes. These open reading frames shared homologies to a hexose carrier, an ATP-binding cassette (ABC) transporter, two cytochromes P450, two lectin receptor kinases, a cysteine proteinase and a glycosyl transferase (Krattinger et al. 2009). Sequence analysis of the candidate gene coding regions from the parental alleles as well as the *Lr34* mutants identified several sequence polymorphisms in the ABC transporter gene leading to either splice site mutations, amino acid changes, frame shift mutations or pre-mature stop codons, thus confirming the ABC transporter gene as the *Lr34/Yr18/Pm38* gene providing durable resistance against leaf rust (Krattinger et al. 2009). Thus, the *Pm38* gene is the first cloned quantitatively acting disease resistance gene against powdery mildew and was also reported in the cultivars Fukoho-Komugi and Saar from Japan and CIMMYT, respectively (Liang et al. 2006; Lillemo et al. 2008). As discussed above, there are many additional quantitative trait loci (QTL) involved in powdery mildew resistance. Keller et al. (1999) identified 18 QTLs against powdery mildew in a segregating wheat x spelt

(*Triticum spelta*) population explaining 77 % of the phenotypic variation. However, in most of the cases only 1–4 QTLs have major effects. The wheat cultivars Knox (Shaner 1973) and Massey (Griffey and Das 1994) are two cultivars showing effective powdery mildew APR, which presumably is governed by two to three genes only. Similarly, several other QTLs have been identified in different wheat cultivars originating from different countries such as RE 714, Festin, Courtot and RE 9001 from France (Chantret et al. 2001; Mingeot et al. 2002; Bougot et al. 2006), USG3209 from North America (Tucker et al. 2007), Oligoculm from Israel (Liang et al. 2006), Avocet from Australia (Lillemo et al. 2008), Suwon 92 from Korea (Xu et al. 2006) and Bainong64 originating from China (Lan et al. 2009). Once molecular markers for a number of QTL contributing additively to powdery mildew resistance are known, this will allow a very efficient breeding approach to combine such loci and obtain genotypes with sufficient field resistance efficiently.

15.2.2 Basis of Resistance to Stagonospora Nodorum Leaf Blotch in Wheat

Being a necrotrophic fungus, *Phaeosphaeria nodorum* (anamorph *Stagonospora nodorum)* infects and kills wheat leaf tissue and feeds from the organic compounds of the dead cells during its life cycle. To invade wheat leaves, *S. nodorum* produces proteinaceous Host Selective Toxins (HST). These HSTs interact with the plant host in a mirrored gene-for-gene interaction. In the following paragraphs, we will describe the current knowledge on toxin- mediated resistance to Stagonospora nodorum blotch (Oliver et al. 2012).

15.2.2.1 Interactions Between Fungal Toxins and Wheat Sensitivity Genes Cause Susceptibility

According to the classical gene-for-gene model developed by (Flor 1955), a pathogen is only able to invade the host successfully if the plant does not recognize the pathogens virulence factor by a corresponding *R* gene. In the mirrored gene-for gene interaction of Stagonospora nodorum leaf blotch, the infection will be successful only if the wheat cultivar has a corresponding susceptibility gene (Friesen et al. 2007). This type of interaction was identified as the cause of a few additional fungal diseases in different plant species (Table 15.2) (Mengiste 2012).

The recently sequenced genome of *S. nodorum* provided the opportunity to study the genetic basis of pathogenicity together with other features of the fungal lifestyle. The genome size was estimated to be 37.2 Mbp (Hane et al. 2007) and gene predictions and EST library analysis suggested that the genome contains at least 10,762 genes. Interestingly, a large number of identified genes were predicted to encode secreted proteins with no similarity to any known genes. Possibly, new host-selective toxins are among these genes. For instance, the host-selective toxin SnTox1 was identified by screening the whole *S. nodorum* genome for suitable candidates and then testing them in infection experiments (Liu et al. 2012).

Table 15.2 Cloned plant toxin-sensitivity genes which interact with fungal toxins resulting in susceptible disease response

Plant species	Fungal pathogen	Toxin	Susceptibility gene	References
Sorghum (*Sorghum bicolor*)	*Periconia circinata*	PC toxin	*Pc* (NBS-LRR)	Nagy et al. 2007
Arabidopsis thaliana	*Cochliobolus victoriae*	victorin	*LOV1* (NBS-LRR)	Lorang et al. 2007
Wheat (*Triticum aestivum*)	*Stagonospora nodorum*	ToxA	*Tsn1* (NBS-LRR)	Faris et al. 2010

Different strains of *S. nodorum* produce a range of HSTs. Five different toxins SnToxA, SnTox1, SnTox2, SnTox3 and SnTox4 have been identified until now (Liu et al. 2004a; Friesen et al. 2006; Friesen et al. 2007; Abeysekara et al. 2009). The susceptibility genes for all five toxins were mapped to different regions of the wheat genome: *Tsn1* interacts with ToxA and this interaction explains 77 % of the phenotypic variation in the population of cultivars 'BR34' and 'Grandin' (Liu et al. 2006) and 95 % of the phenotypic variation in the LD5B population of tetraploid wheat (Faris and Friesen 2009). The *Snn1* and SnTox1 interaction explains 58 % of variation in the ITMI population (Liu et al. 2004b) and *Snn2* – SnTox2, *Snn3* – SnTox3 and *Snn4* – SnTox4 are responsible for 47, 17 and 41 %, respectively, observed in segregating wheat populations derived from a cross between the hard red spring wheat line BR34 and cultivar Grandin for *Snn2* and *Snn3*, and a RIL population of Arina x Forno for *Snn4* (Abeysekara et al. 2009). Interestingly, each fungal toxin-wheat gene interaction is qualitative, but they contribute to the resistance response in a quantitative manner. For example, SnToxA-*Tsn1* and SnTox2-*Snn2* have additive effects during the infection (Oliver et al. 2012).

15.2.2.2 Quantitative Resistance to SNB

Classical genetic studies suggest that resistance to SNB is complex and in most cases polygenic (Scott et al. 1982; Fried and Meister 1987; Bostwick et al. 1993; Du et al. 1999). Monogenic resistance was also identified in some wheat varieties (Kleijer et al. 1977; Ma and Hughes 1995; Murphy et al. 2000). The resistance responses to SNB on leaves and glumes are genetically independent (Francki et al. 2011). Several QTL controlling partial resistance to Stagonospora nodorum blotch in seedlings were identified on chromosomes 2B, 3B, 5B and 5D using a double haploid population derived from a cross of winter wheat cultivars 'Liwilla' and 'Begra' (Czembor et al. 2003). However, their effect on adult plants was not tested. QTLs for resistance to SNB on the flag leaf might correspond to the loci associated with toxin insensitivity genes in the wheat genome: for example, Francki et al. (2011) discovered three QTLs using a cross of winter wheat 'P92201D5' and spring wheat 'EGA Blanko'. Two of them, located on chromosomes 1BS and 2AS respectively, did not correlate with any known toxin sensitivity genes. In contrast, the third QTL on chromosome 5BL was

associated with *Tsn1*-ToxA insensitivity. Independent genetic control of resistance to SNB in glumes and leaves combined with diverse resistance on different stages of plant growth suggests that the best strategy for breeding is to combine the different genetic loci and take advantage of their additive effects.

15.2.2.3 The SNB Susceptibility Gene Tsn1 Encodes an NBS-LRR Protein

The susceptibility genes have additive effects if multiple compatible interactions are acting at the same time. Therefore, disease resistance to Stagonospora nodorum leaf blotch depends on the presence of susceptibility genes and is quantitatively inherited (Abeysekara et al. 2009). *Tsn1* confers sensitivity to SnToxA and is located on the long arm of chromosome 5B. The *Tsn1* gene was recently cloned using a classical chromosome walking approach after establishing a physical contig of 350 kb containing the flanking markers (Faris et al. 2010). Bioinformatic analysis identified six genes cosegregating with *Tsn1*. An association study on 386 wheat accession narrowed the number of candidates down to four genes. Further validation revealed that *Tsn1* has a resistance gene-like structure consisting of a nucleotide-binding, leucine-rich repeat (NBS-LRR) and a serine/threonine protein kinase (S/TPK) domain. Mutagenesis experiments demonstrated that all three domains are required for disease susceptibility. The analysis of *Tsn1* suggests that the gene originated from a B-genome donor through a gene fusion. The exact mechanism of the HST-gene interaction still remains unknown. The presence of *Tsn1* is required for ToxA recognition, but yeast two-hybrid experiments suggest that the Tsn1 protein does not interact directly with ToxA. It was shown that *Tsn1* transcription is regulated by the circadian clock and light, indicating that the Tsn1-ToxA interactions are linked to photosynthesis processes. Faris et al. (2010) suggested that in the case of *Tsn1*- ToxA interaction, *S. nodorum* may have subverted a wheat defence mechanism based on an NBS-LRR immune receptor that was (and possibly still is) involved in resistance against a different pathogen species.

15.2.2.4 Genomics-Assisted Use of Genetic Resources for SNB Resistance Breeding Based on the Molecular Understanding of the Pathosystem

Based on the recent findings on host-specific toxins in the *S. nodorum*-wheat pathosystem, it is evident that the presence or absence of specific toxin receptors in the widely grown wheat cultivars will have a significant impact on disease prevalence. It was recently shown (McDonald et al. 2013) that there are significant differences between the frequencies of toxin presence in *S. nodorum* isolates originating from different geographical regions. This suggests that the presence/absence of sensitivity genes in the cultivars grown in particular regions has a strong effect: whenever a cultivar contains the sensitivity gene corresponding to a specific toxin, the presence of this toxin will be of selective advantage for the pathogen and races with the toxin will increase in frequency. On the other hand, if the sensitivity gene is absent, there will be no selective advantage for having the toxin and it is likely that the frequency of such races will decrease.

These findings immediately suggest that a breeding strategy which has the goal to eliminate as many relevant susceptibility genes as possible from the germplasm, might be effective (it remains to be determined which ones belong to this group in addition to *Tsn1*). This has not been tried yet but has considerable potential to reduce the problem of SNB based on diagnostic markers for a limited subset of toxin susceptibility genes. The markers would allow the elimination of all breeding material with active susceptibility genes. Clearly, this will only be possible if the molecular differences between susceptible and non-susceptible alleles will be known. At this stage, only the Tsn1 receptor is cloned and more map-based cloning projects are needed to molecularly isolate the other toxin receptor genes. Ideally such an effort to eliminate susceptible lines would be coordinated in large geographical areas to ensure success and reduce the frequency of toxin genes. Such a project is ongoing in Australia to eliminate the *Tsn1* gene from commercial germplasm (Oliver and Solomon 2010; Waters et al. 2011).

In conclusion, based on the molecular advancements in understanding the *S. nodorum*-wheat pathosystem, future resistance breeding efforts will possibly rely more on molecular markers for selecting against susceptibility (receptor) genes and not only depend on phenotyping under field conditions. It will be interesting to see if similar type of genes is responsible for resistance to Stagonospora nodorum glume blotch, the disease on the glume. As resistance in the glume is inherited independently from resistance in the leaf, other genetic factors must be involved (Schnurbusch et al. 2003).

15.2.2.5 Genomics Reveals an Interspecific Gene Transfer and Rapid Virulence Evolution in a Wheat Pathogen

It is assumed that rapid diversification of effectors in pathogens is closely linked to the avoidance of detection by the plant immune system (Dodds et al. 2006). Biotrophs, such as powdery mildew, are seeking for new ways to overcome the resistance genes and colonize the host. In contrast, necrotrophic pathogens benefit from the hypersensitive response and feed from the dead tissue. However, the diversification of effectors plays an important role for necrotrophs as well. It was suggested that the diversity of fungal toxins found in necrotrophs and in particular in *S. nodorum* can be explained by two hypotheses. The first hypothesis postulates that necrotrophs gain evolutionary benefits by tracking the appearance of new sensitivity alleles in the host (Stukenbrock and McDonald 2007). The second hypothesis suggests that the diversification of the toxins allows the pathogen to increase its fitness and aggressiveness (Tan et al. 2012). Effector diversity is the result of recombination and mutation events in the toxin genes, but also of non-vertical genetic exchanges (horizontal gene transfer) known to occur in filamentous fungi. It was found that some genes in the *S. nodorum* genome have no homology to any known genes in closely related fungi. The presence of those genes might indicate that they were acquired by horizontal transfer from another, more distantly related species (Oliver et al. 2012). Recently, Friesen et al. (2006) provided evidence for the gene transfer between the two fungal pathogens *S. nodorum* and *Pyrenophora tritici-repentis*. *P. tritici-repentis* produces the host-selective toxin ToxA. The ToxA gene was cloned

previously by Ciuffetti et al. (1997). Analysis of the sequenced *S. nodorum* genome (Hane et al. 2007) revealed the presence of a close homolog with a similar gene structure consisting of three exons and two introns and sharing 99.7 % identity with the *P. tritici-repentis* ToxA. The high similarity suggests a recent common ancestor gene. Several isolates of *S. nodorum* and *P. tritici-repentis* with different geographical origins were tested for their ToxA sequence diversity. Among 95 *S. nodorum* and 54 *P. tritici-repentis* ToxA amplicons only one haplotype was identified for *P. tritici-repentis* whereas 11 haplotypes were found in *S. nodorum*. This suggests that the ToxA gene was more ancient in the *S. nodorum* genome and was probably introduced only recently in the *P.tritici-repentis* genome. Further analysis of the 11 kb genomic region flanking the ToxA gene in both species revealed a high degree of conservation: 80–90 % in the distal parts and 98–100 % in the middle. Additionally, functional analysis of ToxA-disrupted mutants and their interaction with the wheat *Tsn1* gene indicated a role of ToxA in inducing a susceptible plant response for both *P. tritici-repentis* and *S. nodorum*. This hypothesis is also supported by the fact that tan spot in comparison with *S. nodorum* leaf blotch was described in wheat only. The first records about tan spot as an occasional pathogen of wheat date from 1928. However, only in 1942 the typical necrotic symptoms were described. In contrast, *S. nodorum* leaf blotch was known as an important wheat disease already since 1889. This strongly suggests that an interspecific gene transfer between *S. nodorum* and *P. triticii-repentis* indeed has occurred and it happened most likely around 1942. Analysis of the *S. nodorum* genome sequence shows that interspecific horizontal gene transfer is not a rare and exotic mechanism, but the significant contributor to the pathogen adaptation. Clearly, the application of genomic tools in pathogenomics has resulted in findings highly relevant for wheat resistance breeding.

15.3 Conclusions

Global food security strongly depends on a highly productive and sustainable agriculture. Fungal pathogens can cause severe yield losses in all major crops and are a serious threat for food security, especially in developing countries. Breeding for resistant wheat varieties is the most effective strategy to counteract these diseases, requiring however a better understanding of the molecular basis of disease resistance. The genetic complexity of wheat greatly complicates gene isolation and functional characterization, explaining the limited number of so far characterized resistance genes in wheat. Major race-specific resistance genes can provide plants with a high level of disease resistance. However, biotrophic fungi such as the powdery mildews are rapidly evolving pathogens which are able to overcome these resistance genes. Thus, new sources of genetic resistance have to be identified in order to avoid an erosion of the current pool of agriculturally important resistance genes.

Molecular isolation of the race-specific *Pm3* resistance gene provided highly valuable insights in the diversity and evolution of resistance genes. With the help of developed molecular markers and an established functional validation assay, the

allele mining strategy could be tested for its efficiency to explore genetic diversity and identify new resistance sources. Indeed, this strategy allowed the isolation of ten functional resistance alleles in addition to the seven genetically known *Pm3* alleles, demonstrating the importance of wild landraces and wheat progenitors as valuable genetic resources for resistance as well as the feasibility of the allele mining strategy.

The recent finding that in necrotrophic pathosystems such as *S. nodorum*, an interaction between a pathogen toxin and a susceptibility host component is required for a successful pathogen invasion, influenced research on the isolation of genes providing resistance to necrotrophic diseases and possibly explains the present limited knowledge thereof. Nevertheless, the awareness of susceptibility genes being required for pathogen establishment allows breeding for cultivars which lack these genes and thus provide higher resistance to necrotrophic fungi.

With the emergence of highly virulent pathogen strains which overcome previously effective resistance genes, disease resistance research is currently expanding towards the isolation of quantitative resistance. Although this type of resistance is often only partial, it was shown to be more durable (Kou and Wang 2010). Cloning of *Lr34/Yr18/Pm38* sets a successful example of isolation of quantitative, durable and broad spectrum disease resistance gene. With the isolation of *Pm21*, providing durable and broad spectrum resistance, it will be possible to gain additional insights into the molecular mechanisms of durable resistance, and also, similarly to *Pm3*, expand the variation of functional and durable *Pm21* alleles. In contrast, durable resistance to *S. nodorum* is still only poorly investigated and urgently needs further molecular analysis.

References

Abeysekara NS, Friesen TL, Keller B, Faris JD (2009) Identification and characterization of a novel host-toxin interaction in the wheat-*Stagonospora nodorum* pathosystem. Theor Appl Genet 120:117–126

Anderson PK, Cunningham AA, Patel NG et al (2004) Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. Trends Ecol Evol 19:535–544

Arraiano LS, Balaam N, Fenwick PM et al (2009) Contributions of disease resistance and escape to the control of septoria tritici blotch of wheat. Plant Pathol 58:910–922

Ayala L, Henry M, van Ginkel M et al (2002) Identification of QTLs for BYDV tolerance in bread wheat. Euphytica 128:249–259

Balter M (2007) Seeking agriculture's ancient roots. Science 316:1830–1835

Bearchell SJ, Fraaije BA, Shaw MW, Fitt BDL (2005) Wheat archive links long-term fungal pathogen population dynamics to air pollution. Proc Natl Acad Sci 102:5438–5442

Berry PM, Dawson TP, Harrison PA, Pearson RG (2002) Modelling potential impacts of climate change on the bioclimatic envelope of species in Britain and Ireland. Global Ecol Biogeogr 11:453–462

Bhullar NK, Street K, Mackay M et al (2009) Unlocking wheat genetic resources for the molecular identification of previously undescribed functional alleles at the *Pm3* resistance locus. Proc Natl Acad Sci 106:9519–9524

Bhullar NK, Zhang ZQ, Wicker T, Keller B (2010) Wheat gene bank accessions as a source of new alleles of the powdery mildew resistance gene *Pm3*: a large scale allele mining project. BMC Plant Biol 10:88

Bostwick DE, Ohm HW, Shaner G (1993) Inheritance of septoria-glume blotch resistance in wheat. Crop Sci 33:439–443
Bougot Y, Lemoine J, Pavoine MT et al (2002) Identification of microsatellite marker associated with *Pm3* resistance alleles to powdery in wheat. Plant Breed 121:325–329
Bougot Y, Lemoine J, Pavoine MT et al (2006) A major QTL effect controlling resistance to powdery mildew in winter wheat at the adult plant stage. Plant Breed 125:550–556
Brown JKM, Tellier A (2011) Plant-Parasite Coevolution: Bridging the Gap between Genetics and Ecology. Ann Rev Phytopathol 49:345–367
Brunner S, Hurni S, Streckeisen P et al (2010) Intragenic allele pyramiding combines different specificities of wheat *Pm3* resistance alleles. Plant J 64:433–445
Brunner S, Hurni S, Herren G et al (2011) Transgenic *Pm3b* wheat lines show resistance to powdery mildew in the field. Plant Biotech J 9:897–910
Brunner S, Stirnweis D, Quijano CD et al (2012) Transgenic *Pm3* multilines of wheat show increased powdery mildew resistance in the field. Plant Biotech J 10:398–409
Burger JC, Chapman MA, Burke JM (2008) Molecular insights into the evolution of crop plants. Am J Bot 95:13–122
Cannon RJC (1998) The implications of predicted climate change for insect pests in the UK, with emphasis on non-indigenous species. Glob Change Biol 4:785–796
Cao AH, Xing LP, Wang XY et al (2011) Serine/threonine kinase gene *Stpk-V*, a key member of powdery mildew resistance gene *Pm21*, confers powdery mildew resistance in wheat. Proc Natl Acad Sci 108:7727–7732
Chakraborty S, Tiedemann AV, Teng PS (2000) Climate change: potential impact on plant diseases. Environ Pollut 108:317–326
Chantret N, Mingeot D, Sourdille P et al (2001) A major QTL for powdery mildew resistance is stable over time and at two development stages in winter wheat. Theor Appl Genet 103:962–971
Charmet G (2011) Wheat domestication: Lessons for the future. C R Biol 334:212–220
Ciuffetti LM, Tuori RP, Gaventa JM (1997) A single gene encodes a selective toxin causal to the development of tan spot of wheat. Plant Cell 9:135–144
Cloutier S, McCallum BD, Loutre C et al (2007) Leaf rust resistance gene *Lr1*, isolated from bread wheat (*Triticum aestivum* L.) is a member of the large psr567 gene family. Plant Mol Biol 65:93–106
Czembor PC, Arseniuk E, Czaplicki A et al (2003) QTL mapping of partial resistance in winter wheat to Stagonospora nodorum blotch. Genome 46:546–554
Dodds PN, Lawrence GJ, Catanzariti AM et al (2006) Direct protein interaction underlies gene-for-gene specificity and coevolution of the flax resistance genes and flax rust avirulence genes. Proc Natl Acad Sci 103:8888–8893
Douchkov D, Nowara D, Zierold U, Schweizer P (2005) A high-throughput gene-silencing system for the functional assessment of defense-related genes in barley epidermal cells. Mol Plant Microbe Int 18:755–761
Du CC, Nelson LR, McDaniel ME (1999) Diallel analysis of gene effects conditioning resistance to *Stagonospora nodorum* (Berk.) in wheat. Crop Sci 39:686–690
Duveiller E, Singh RP, Nicol JM (2007) The challenges of maintaining wheat productivity: pests, diseases, and potential epidemics. Euphytica 157:417–430
Dyck PL (1987) The association of a gene for leaf rust resistance with the chromosome—7d suppressor of stem rust resistance in common wheat. Genome 29:467–469
Dyck PL (1991) Genetics of adult-plant leaf rust resistance in Chinese Spring and sturdy wheats. Crop Sci 31:309–311
Dyck PL, Samborskj D, Anderson RG (1966) Inheritance of adult-plant leaf rust resistance derived from common wheat varieties Exchange and Frontana. Can J Genet Cytol 8:665–671
Endresen DTF, Street K, Mackay M et al (2011) Predictive association between biotic stress traits and eco-geographic data for wheat and barley landraces. Crop Sci 51:2036–2055

Faris JD, Friesen TL (2009) Reevaluation of a tetraploid wheat population indicates that the *Tsn1*-ToxA interaction is the only factor governing Stagonospora nodorum blotch susceptibility. Phytopathol 99:906–912

Faris JD, Zhang Z, Lu H et al (2010) A unique wheat disease resistance-like gene governs effector-triggered susceptibility to necrotrophic pathogens. Proc Natl Acad Sci 107:13544–13549

Feuillet C, Travella S, Stein N et al (2003) Map-based isolation of the leaf rust disease resistance gene *Lr10* from the hexaploid wheat (*Triticum aestivum* L.) genome. Proc Natl Acad Sci 100:15253–15258

Flor HH (1955) Host-parasite interaction in flax rust—its genetics and other implications. Phytopathol 45:680–685

Francki MG, Shankar M, Walker E et al (2011) New quantitative trait loci in wheat for flag leaf resistance to Stagonospora nodorum blotch. Phytopathol 101:1278–1284

Fried PM, Meister E (1987) Inheritance of leaf and head resistance of winter-wheat to *Septoria-nodorum* in a diallel cross. Phytopathol 77:1371–1375

Friesen TL, Stukenbrock EH, Liu Z et al (2006) Emergence of a new disease as a result of interspecific virulence gene transfer. Nature Genet 38:953–956

Friesen TL, Meinhardt SW, Faris JD (2007) The *Stagonospora nodorum*-wheat pathosystem involves multiple proteinaceous host-selective toxins and corresponding host sensitivity genes that interact in an inverse gene-for-gene manner. Plant J 51:681–692

Fu DL, Uauy C, Distelfeld A et al (2009) A kinase-START gene confers temperature-dependent resistance to wheat stripe rust. Science 323:1357–1360

Ge YF, Johnson JW, Roberts JJ, Rajaram S (1998) Temperature and resistance gene interactions in the expression of resistance to *Blumeria graminis* f. sp. *tritici*. Euphytica 99:103–109

Griffey CA, Das MK (1994) Inheritance of adult-plant resistance to powdery mildew in Knox—62 and Massey winter wheats. Crop Sci 34:641–646

Hane JK, Lowe RGT, Solomon PS et al (2007) Dothideomycete-plant interactions illuminated by genome sequencing and EST analysis of the wheat pathogen *Stagonospora nodorum*. Plant Cell 19:3347–3368

Hartl L, Weiss H, Zeller FJ et al (1993) Use of RFLP markers for the identification of alleles of the *Pm3* locus conferring powdery mildew resistance in wheat (*Triticum aestivum* L.). Theor Appl Genet 86:959–963

He R, Chang Z, Yang Z et al (2009) Inheritance and mapping of powdery mildew resistance gene *Pm43* introgressed from *Thinopyrum intermedium* into wheat. Theor Appl Genet 118:1173–1180

Heun M, Friebe B, Bushuk W (1990) Chromosomal location of the powdery mildew resistance gene of Amigo wheat. Phytopathol 80:1129–1133

Horbach R, Navarro-Quesada AR, Knogge W, Deising HB (2011) When and how to kill a plant cell: Infection strategies of plant pathogenic fungi. J Plant Physiol 168:51–62

Hua W, Liu Z, Zhu J et al (2009) Identification and genetic mapping of *Pm42*, a new recessive wheat powdery mildew resistance gene derived from wild emmer (*Triticum turgidum* var. *dicoccoides*). Theor Appl Genet 119:223–230

Huang L, Brooks SA, Li WL et al (2003) Map-based cloning of leaf rust resistance gene *Lr21* from the large and polyploid genome of bread wheat. Genetics 164:655–664

Huang J, Zhao Z, Song F, Wang X et al (2012) Molecular detection of a gene effective against powdery mildew in the wheat cultivar Liangxing 66. Mol Breeding 30:1737–1745

Kaur N, Street K, Mackay M et al (2008) Molecular approaches for characterization and use of natural disease resistance in wheat. Europ J Plant Pathol 121:387–397

Keller M, Keller B, Schachermayr G et al (1999) Quantitative trait loci for resistance against powdery mildew in a segregating wheat x spelt population. Theor Appl Genet 98:903–912

Kleijer G, Bronnimann A, Fossati A (1977) Chromosomal location of a dominant gene for resistance at seedling stage to *Septoria-nodorum* berk in wheat variety Atlas—66. J Plant Breed 78:170–173

Koltin Y, Kenneth R (1970) Role of sexual stage in over-summering of *Erysiphe-graminis* dc fsp *hordei* marchal under semi-arid conditions. Ann Appl Biol 65:263–268

Kou Y, Wang S (2010) Broad-spectrum and durability: understanding of quantitative disease resistance. Curr Opin Plant Biol 13:181–185

Krattinger SG, Lagudah ES, Spielmeyer W et al (2009) A putative ABC transporter confers durable resistance to multiple fungal pathogens in wheat. Science 323:1360–1363

Kumar GR, Sakthivel K, Sundaram RM et al (2010) Allele mining in crops: Prospects and potentials. Biotech Adv 28:451–461

Lan C, Liang S, Wang Z et al (2009) Quantitative trait loci mapping for adult-plant resistance to powdery mildew in Chinese wheat cultivar Bainong 64. Phytopathol 99:1121–1126

Liang SS, Suenaga K, He ZH et al (2006) Quantitative trait loci mapping for adult-plant resistance to powdery mildew in bread wheat. Phytopathol 96:784–789

Lillemo M, Singh RP, Huerta-Espino J et al (2007) Leaf rust resistance gene *LR34* is involved in powdery mildew resistance of CIMMYT bread wheat line Saar. Wheat Production in Stressed Environments 12:97–102

Lillemo M, Asalf B, Singh RP et al (2008) The adult plant rust resistance loci *Lr34/Yr18* and *Lr46/Yr29* are important determinants of partial resistance to powdery mildew in bread wheat line Saar. Theor Appl Genet 116:1155–1166

Liu Z, Friesen TL, Ling H et al (2006) The *Tsn1*-ToxA interaction in the wheat-*Stagonospora nodorum* pathosystem parallels that of the wheat-tan spot system. Genome 49:1265–1273

Liu Z, Zhang Z, Faris JD et al (2012) The cysteine rich necrotrophic effector SnTox1 produced by *Stagonospora nodorum* triggers susceptibility of wheat lines harboring *Snn1*. PLoS Pathog 8:e1002467

Liu ZH, Faris JD, Meinhardt SW et al (2004a) Genetic and physical mapping of a gene conditioning sensitivity in wheat to a partially purified host-selective toxin produced by *Stagonospora nodorum*. Phytopathol 94:1056–1060

Liu ZH, Friesen TL, Rasmussen JB et al (2004b) Quantitative trait loci analysis and mapping of seedling resistance to Stagonospora nodorum leaf blotch in wheat. Phytopathol 94:1061–1067

Lorang JM, Sweat TA, Wolpert TJ (2007) Plant disease susceptibility conferred by a "resistance" gene. Proc Natl Acad Sci 104:14861–14866

Luck J, Spackman M, Freeman A et al (2011) Climate change and diseases of food crops. Plant Pathol 60:113–121

Luo PG, Luo HY, Chang ZJ et al (2009) Characterization and chromosomal location of *Pm40* in common wheat: a new gene for resistance to powdery mildew derived from *Elytrigia intermedium*. Theor Appl Genet 118:1059–1064

Ma H, Hughes GR (1995) Genetic-control and chromosomal location of *Triticum timopheevii*-derived resistance to *Septoria nodorum* blotch in durum-wheat. Genome 38:332–338

Ma ZQ, Sorrells ME, Tanksley SD (1994) RFLP markers linked to powdery mildew resistance gene *Pm1*, *Pm2*, *Pm3* and *Pm4* in wheat. Genome 37:871–875

McCouch SR, Sweeney M, Li JM et al (2007) Through the genetic bottleneck: *O. rufipogon* as a source of trait-enhancing alleles for *O. sativa*. Euphytica 154:317–339

McDonald MC, Oliver RP, Friesen TL, Brunner PC, McDonald BA (2013) Global diversity and distribution of three necrotrophic effectors in *Phaeosphaeria nodorum* and related species. New Phytol doi: 10.1111/nph.12257

McIntosh RA (1992) Close genetic-linkage of genes conferring adult-plant resistance to leaf rust and stripe rust in wheat. Plant Pathol 41:523–527

Mendgen K, Hahn M (2002) Plant infection and the establishment of fungal biotrophy. Trends Plant Sci 7:352–356

Mengiste T (2012) Plant immunity to necrotrophs. Ann Rev Phytopathol 50:267–294

Mingeot D, Chantret N, Baret PV et al (2002) Mapping QTL involved in adult plant resistance to powdery mildew in the winter wheat line RE714 in two susceptible genetic backgrounds. Plant Breed 121:133–140

Mundt CC (2002) Use of multiline cultivars and cultivar mixtures for disease management. Ann Rev Phytopathol 40:381–410

Murphy NEA, Loughman R, Wilson R et al (2000) Resistance to septoria nodorum blotch in the *Aegilops tauschii* accession RL5271 is controlled by a single gene. Euphytica 113:227–233
Nagy ED, Lee T-C, Ramakrishna W et al (2007) Fine mapping of the *Pc* locus of *Sorghum bicolor*, a gene controlling the reaction to a fungal pathogen and its host-selective toxin. Theor Appl Genet 114:961–970
Newton AC, Johnson SN, Gregory PJ (2011) Implications of climate change for diseases, crop yields and food security. Euphytica 179:3–18
Oberhaensli S, Parlange F, Buchmann JP et al (2011) Comparative sequence analysis of wheat and barley powdery mildew fungi reveals gene colinearity, dates divergence and indicates host-pathogen co-evolution. Fung Genet Biol 48:327–334
Oerke EC, Dehne HW, Schoenbeck F, Weber A (1994) Crop production and crop protection: Estimated losses in major food and cash crops. Elsevier Science Publishers, Amsterdam
Olesen JE, Mortensen JV, Jorgensen LN, Andersen MN (2000) Irrigation strategy, nitrogen application and fungicide control in winter wheat on a sandy soil. I. Yield, yield components and nitrogen uptake. J Agricult Sci 134:1–11
Oliver RP, Solomon PS (2010) New developments in pathogenicity and virulence of necrotrophs. Curr Opin Plant Biol 13:415–419
Oliver RP, Friesen TL, Faris JD, Solomon PS (2012) *Stagonospora nodorum*: From Pathology to Genomics and Host Resistance. Ann Rev Phytopathol 50:23–43
Panstruga R (2003) Establishing compatibility between plants and obligate biotrophic pathogens. Curr Opin Plant Biol 6:320–326
Peng JH, Sun D, Nevo E (2011) Domestication evolution, genetics and genomics in wheat. Molecular Breed 28:281–301
Perfect SE, Green JR (2001) Infection structures of biotrophic and hemibiotrophic fungal plant pathogens. Mol Plant Pathol 2:101–108
Pimentel D, McNair S, Janecka J et al (2001) Economic and environmental threats of alien plant, animal, and microbe invasions. Agr Ecosyst Environ 84:1–20
Qiu JW, Schurch AC, Yahiaoui N et al (2007) Physical mapping and identification of a candidate for the leaf rust resistance gene *Lr1* of wheat. Theor Appl Genet 115:159–168
Schnurbusch T, Paillard S, Fossati D et al (2003) Detection of QTLs for Stagonospora glume blotch resistance in Swiss winter wheat. Theor App Genet 107:1226–1234
Schweizer P, Pokorny J, Abderhalden O, Dudler R (1999) A transient assay system for the functional assessment of defense-related genes in wheat. Mol Plant Microbe Int 12:647–654
Scott PR, Benedikz PW, Cox CJ (1982) A genetic-study of the relationship between height, time of ear emergence and resistance to *Septoria-nodorum* in wheat. Plant Pathol 31:45–60
Shaner G (1973) Evaluation of slow-mildewing resistance of Knox wheat in field. Phytopathol 63:867–872
Sharma AK, Sharma RK, Babu KS (2004) Effect of planting options and irrigation schedules on development of powdery mildew and yield of wheat in the North Western plains of India. Crop Prot 23:249–253
Sharma S, Khan TA, Ashraf MS (2011) Studies on powdery mildew disease of mulberry (*Morus alba*): a new report from Uttar Pradesh, India. Archives of Phytopathology and Plant Protection 44:105–112
Shaw MW, Bearchell SJ, Fitt BDL, Fraaije BA (2008) Long-term relationships between environment and abundance in wheat of *Phaeosphaeria nodorum* and *Mycosphaerella graminicola*. New Phytol 177:229–238
Singh RP (1992) Association between gene *Lr34* for leaf rust resistance and leaf tip necrosis in wheat. Crop Sci 32:874–878
Spanu PD, Abbott JC, Amselem J et al (2010) Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. Science 330:1543–1546
Spielmeyer W, McIntosh RA, Kolmer J, Lagudah ES (2005) Powdery mildew resistance and *Lr34/Yr18* genes for durable resistance to leaf and stripe rust cosegregate at a locus on the short arm of chromosome 7D of wheat. Theor Appl Genet 111:731–735

Srichumpa P, Brunner S, Keller B, Yahiaoui N (2005) Allelic series of four powdery mildew resistance genes at the *Pm3* locus in hexaploid bread wheat. Plant Physiol 139:885–895

Stukenbrock EH, McDonald BA (2007) Geographical variation and positive diversifying selection in the host-specific toxin SnToxA. Mol Plant Pathol 8:321–332

Stukenbrock EH, Banke S, McDonald BA (2006) Global migration patterns in the fungal wheat pathogen *Phaeosphaeria nodorum*. Mol Ecol 15:2895–2904

Tan K-C, Ferguson-Hunt M, Rybak K et al (2012) Quantitative variation in effector activity of ToxA isoforms from *Stagonospora nodorum* and *Pyrenophora tritici-repentis*. Mol Plant Microbe Int 25:515–522

Tommasini L, Yahiaoui N, Srichumpa P, Keller B (2006) Development of functional markers specific for seven Pm3 resistance alleles and their validation in the bread wheat gene pool. Theor Appl Genet 114:165–175

Tucker DM, Griffey CA, Liu S et al (2007) Confirmation of three quantitative trait loci conferring adult plant resistance to powdery mildew in two winter wheat populations. Euphytica 155:1–13

Varshney RK, Nayak SN, May GD, Jackson SA (2009) Next-generation sequencing technologies and their implications for crop genetics and breeding. Trends Biotechnol 27:522–530

Waters ODC, Lichtenzveig J, Rybak K et al (2011) Prevalence and importance of sensitivity to the *Stagonospora nodorum* necrotrophic effector SnTox3 in current Western Australian wheat cultivars. Crop Pasture Science 62:556–562

Wicker T, Oberhaensli S, Parlange F et al (2013) The wheat powdery mildew genome shows the unique evolution of an obligate biotroph. Nat Genet 45:1092–1096

Wicker T, Yahiaoui N, Keller B (2007) Contrasting rates of evolution in *Pm3* loci from three wheat species and rice. Genetics 177:1207–1216

Wolfe MS (1984) Trying to understand and control powdery mildew. Plant Pathol (Oxford) 33:451–466

Xu W, Li C, Hu L, Wang H et al (2011) Identification and molecular mapping of *PmHNK54*: a novel powdery mildew resistance gene in common wheat. Plant Breed 130:603–607

Xu XY, Bai GH, Carver BF et al (2006) Molecular characterization of a powdery mildew resistance gene in wheat cultivar Suwon 92. Phytopathol 96:496–500

Yahiaoui N, Brunner S, Keller B (2006) Rapid generation of new powdery mildew resistance genes after wheat domestication. Plant J 47:85–98

Yahiaoui N, Kaur N, Keller B (2009) Independent evolution of functional *Pm3* resistance genes in wild tetraploid wheat and domesticated bread wheat. Plant J 57:846–856

Yahiaoui N, Srichumpa P, Dudler R, Keller B (2004) Genome analysis at different ploidy levels allows cloning of the powdery mildew resistance gene *Pm3b* from hexaploid wheat. Plant J 37:528–538

Zeller FJ, Lutz J, Stephan U (1993) Chromosome location of genes for resistance to powdery mildew in common wheat (*Triticum aestivum* L.) 1. *Mlk* and other alleles at the *Pm3* locus. Euphytica 68:223–229

Zhu YY, Chen HR, Fan JH et al (2000) Genetic diversity and disease control in rice. Nature 406:718–722

Part III
Genomics-Assisted Crop Improvement for Nutritional Quality

Chapter 16
Breeding for Apple (*Malus* × *domestica* Borkh.) Fruit Quality Traits in the Genomics Era

Satish Kumar, Richard K. Volz, David Chagné and Susan Gardiner

Abstract Apple is one of the most widely cultivated tree fruits in the temperate regions of the world. Development of new apple cultivars is generally based on eating quality, appearance, marketability, health/wellbeing and more recently, on environmental sustainability. A range of new technologies that will both reduce unit costs and increase production are being integrated in apple breeding programmes. Also, molecular breeding is opening an opportunity to significantly reduce the time to incorporate new traits from wild relatives into established crops; enhancing nutritional value, and enabling adaptation to new environmental conditions. The availability of genome sequence along with high throughput genotyping platforms is transforming the strategies for developing new cultivars. Now a range of genome-based selection strategies are providing opportunities to accelerate cultivar breeding. Genomic selection (GS) can be used to obtain genomic breeding values for choosing next-generation parents or selections for further testing as potential commercial cultivars. An empirical study in a New Zealand apple cultivar breeding programme showed that the selection response per unit time using GS compared with the conventional selection were very high (> 100 %) especially for low-heritability traits. The cost for highly paralleled targeted genotyping (e.g. single nucleotide polymorphism assays) is constantly decreasing, which would make implementation of GS affordable in near future.

Keywords Apple breeding · Genetic architecture · Quantitative trait loci · *Malus* × *domestica* Borkh · DNA sequence · Genomic selection

S. Kumar (✉) · R. K. Volz
The New Zealand Institute for Plant & Food Research Limited,
Private Bag 1401, Havelock North 4157, New Zealand
e-mail: Satish.Kumar@plantandfood.co.nz

D. Chagné · S. Gardiner
The New Zealand Institute for Plant & Food Research Limited,
Private Bag 11600, Palmerston North, New Zealand

R. Tuberosa et al. (eds.), *Genomics of Plant Genetic Resources,*
DOI 10.1007/978-94-007-7575-6_16,

16.1 Introduction

The cultivated apple, *Malus* × *domestica*, belongs to the Rosaceae family, and the genus *Malus* is reported to consists of 25–30 species, some of which are cultivated as ornamental trees (Brown 1975). Early studies based on morphological characteristics suggested that several *Malus* species, including *M. sylvestris* Miller, *M. prunifolia* (Willd.) Borkh., and *M. baccata* (L.) Borkh., were involved in the origin and/or domestication of the cultivated apple (Rehder 1940). As Asian *M.* × *asiatica, M. baccata, M. micromalus, M. orientalis, M. prunifolia* and *M. sieversii*, and European *M. sylvestris*, are the species taxonomically closest to *M.* × *domestica*, they are considered to have contributed, to differing extents, to the domestic gene pool (Robinson et al. 2001; Forsline et al. 2003). Velasco et al. (2010) surveyed molecular differences at 23 genes across the genus *Malus*, and supported the formation of the *M.* × *domestica* gene pool from *M. sieversii*. They even suggested that *M.* × *domestica* and *M. sieversii* are the same species, for which the more appropriate nomenclature of *M. pumila* Mill. could be adopted. However, when Harrison and Harrison (2011) re-analyzed some of the polymorphism data from Velasco et al. study, they concluded that gene flow from *M. sylvestris* to *M.* × *domestica* could not be ruled out, in line with previous suggestions (Harris et al. 2002; Coart et al. 2003, 2006). Hence, despite various taxonomic, biochemical, phylogenetic, and molecular studies, the jury is still out on the precise contribution of various *Malus* species to the cultivated apple.

Central Asia, which is the source of apple diversity, is believed to be the origin of apple (Harris et al. 2002). The history of domestication of the apple has been documented in various civilizations of mankind (Morgan and Richards 1993). Apple horticulture was initiated in Assyria (modern-day Iraq) 3,100 years ago, introducing techniques such as grafting, pruning, fruit storage and development of cultivars, which were later spread around various parts of the world by the Greeks and Romans as a result of their travel and invasions (Gardiner et al. 2007). The origin of controlled breeding of apples is attributed to Thomas Andrew Knight (1759–1838) who developed the first cultivars of known parentage (Janick et al. 1996). Apple is now one of the most widely cultivated tree fruits in the temperate regions of the world, and is the third most internationally traded fruit, behind only bananas and grapes.

The fruit of the domestic apple is an important source of nutrients, and is considered as one of the top functional foods—those foods that have an inherent health-promoting benefit beyond basic nutritional value. However, average per capita consumption for most major producing countries has remained static over the last decade (O'Rourke 2011). Even though the production of apples has grown by about 2 % per year in the last decade, world production of all other major fruits has grown even faster. Furthermore, there has also been a huge increase in the range and volume of manufactured snack options available for the consumer. As a result, competition from other fruits and food products for the consumer's preference is intense. In these circumstances, apple industries may benefit by providing consumers with distinctive new apple cultivars that have new and improved benefits, including attractiveness,

taste and texture as well as a reduced need for pesticide application during production. In addition, a further key factor to ensure the industry's future success will be to increase yields per hectare of high quality, marketable fruit. To achieve these goals, apple growers have increased the density of plantings, introduced improved rootstocks and more productive cultivars, and applied a wide array of new technologies that will both reduce unit costs and increase production.

The advent of affordable, high throughput DNA sequencing, coupled with improved bioinformatics and statistical analyses, is now bringing about major advances in the field of plant and animal breeding. Breeding programs are now in a position to investigate genome-wide variations in DNA sequences and link them to the inheritance of complex traits controlled by many genes. In plants, molecular breeding is opening an opportunity to help to ease world's food security dilemma by shortening the time it takes to domesticate new crops from semi-wild plants, and to incorporate new traits rapidly from wild relatives into established crops; or to tailor existing crops to meet new requirements, such as nutritional enhancement of the food product, or enabling adaptation to new environmental conditions caused by climate change. In apple, the recent publication of the genome sequence (Velasco et al. 2010) provides insight into the evolution of this important species, including a possible role of the genome-wide duplication. Apple breeders are now using genome sequence information to uncover clues to the genetic basis of fruit quality, pest and disease resistance as well as production-related traits, and are using this information to improve fruit yield, resistance to various pests and diseases and quality.

16.2 Breeding Systems and Breeding Objectives

16.2.1 Breeding Strategy

The majority of cultivated apples are functional diploids ($2n = 34$) and 17 bivalents form at meiosis (Lespinasse et al. 1976), although some cultivars, such as 'Jonagold', are triploid. Self incompatibility is a common feature in apple, although cases of cross-incompatibility are also reported. Breeding improves plants by concentrating favorable alleles at the expense of those less desirable. Most apple breeding programs worldwide are based on narrow genetic pools. Using historical pedigree records of commercial apple cultivars, Noiton and Alspach (1996) estimated the 'status number' (which is a measure of effective population size that is based on current relatedness only) of the top-50 mainstream cultivars to be 8. The traditional apple breeding strategy has involved crossing among a few top commercial cultivars and elite selections, and planting full-sib families in order to forward-select (based on phenotypic performance) individuals for further clonal testing. Selection of parents with complementary characteristics is essential in order to produce seedlings that have the desired attribute inherited from each parent. As breeders of woody perennial crops cannot afford the time to develop test-crosses to assess the ability of crossing combinations to achieve the breeding goals, there will be an aspect of chance

in selection of parents with a high specific combining ability with regard to quantitatively inherited traits (Bringhurst 1983). Crossing of two commercial cultivars produces progeny exhibiting a wide variation for any one quality trait, with only a very small proportion showing trait improvements over the parents. In apple, selected plants are used as potential cultivars and/or as breeding parents, so that frequencies of desirable alleles are increased in successive populations. The balance of these two strategies (i.e. breeding for cultivar selection *versus* population improvement) and the methods by which they are implemented are dictated by critical aspects of apple crop biology, breeding objectives, screening technologies (including genomic tools), selection thresholds and resources available.

16.2.2 Juvenility

The juvenile period in apple is long for *M*. x *domestica* populations on their own roots in the orchard. Flowering may begin as early as 4 years of age in some seedlings, with the majority fruiting after 6 years. However, juvenility can extend to over 10 years. This period can be reduced by promoting seedling growth in glasshouses and employing flower bud-inducing cultural practices. Substantial flowering in some seedling populations has occurred within 12–18 months from germination (Volz et al. 2009); however, such populations are generally derived from small-fruited *Malus* species (Zimmerman 1991). These practices have not been widely adopted, perhaps because of the requirement to produce fruit for phenotyping from trees similar to those used by the modern grower i.e. seedlings grafted as scions on dwarfing rootstocks in the orchard under "typical" external climatic environment. Kumar et al. (2011) showed that the genetic parameters, including genetic ranking, of many important fruit traits were different for seedlings on their own roots in the orchard than for those propagated onto 'M.9' rootstock.

Dwarfing rootstocks induce earlier flowering of seedlings than those grown on their own roots. However fruiting is somewhat delayed, simply because of the requirement to grow from the propagated single bud or 2–3 bud graft of the seedling, a structure on the rootstock capable of bearing fruit. In the Plant & Food Research cultivar breeding programme, scion-wood is taken from top (> 1.7 m) of two-year-old nursery-grown seedlings and grafted onto 'M.9' cuttings from stoolbeds. These are planted in the nursery for a further year, before planting in the orchard (Fig. 16.1). By using this method, first flowering of grafted plants can occur 3 years after germination (i.e. in the year of propagation); however, fruit phenotyping usually does not begin until 2 years following propagation, or 5 years after germination when 80–90 % of the seedlings are fruiting. Therefore, a combination of 'fast-growing' seedlings in the glasshouse, followed by propagation of the top part of the one-year-old seedlings onto dwarf rootstock in the nursery, can reduce the time to first flowering by 1 year (Fischer 1994; Flachowsky et al. 2011).

A substantial reduction in the juvenile period in apple has been achieved through genetic transformation and over-expression of the *BpMADS4* gene of silver birch

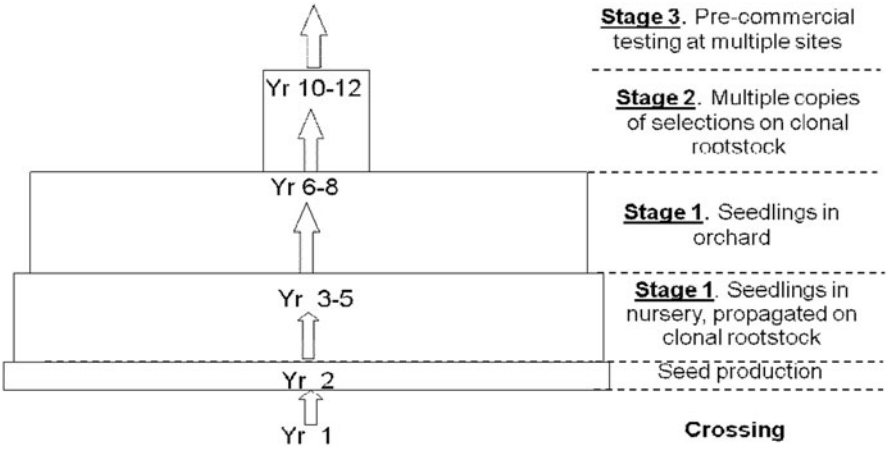

Fig. 16.1 A schematic presentation of the timeline for apple cultivar breeding in New Zealand conditions

(Flachowsky et al. 2007, 2009). Within several months of rooting, transgenic lines flowered and could be pollinated. Transgenic F_1 seedlings exhibited the early flowering trait and yielded seed within 12 months from sowing.

16.2.3 Seedling Selection

A large number of selection traits need to be evaluated before an apple seedling is considered for commercialization. The timing of phenotypic assessment for any one trait depends upon several factors. Some plant traits, such as resistance to apple scab, can be assessed in juvenile plants, with phenotypes similar to those observed on adult trees in the orchard, provided inoculum and environmental conditions are optimum. On the other hand, the evaluation of powdery mildew resistance based on one-year-old seedling assessments does not correlate with assessments made following infection of more mature trees in the orchard (Janse et al. 1994). Clearly fruit characteristics can only be assessed once trees are past the juvenile phase and appraisal of traits that influence yield, such as biennial bearing, requires several years of continuous evaluation (Guitton et al. 2012). For most fruit quality traits, successive measurements over 2–3 years are made in order to select seedlings for stage-2 testing. Generally, the restricted maximum likelihood (REML) estimates of genetic and environmental variance–covariance matrices are used in mixed model equations, and best linear unbiased prediction (BLUP) of breeding values are then obtained by solving these equations. Estimated breeding values are then used for identifying individuals as next-generation parents or as potential cultivars. Such analytical techniques are now commonly used in apple breeding (e.g. Durel et al. 1998; Kouassi et al. 2009; Kumar et al. 2011).

16.2.4 Breeding Objectives

16.2.4.1 Fruit Quality

Development of new food products in technologically advanced countries is today based on consumer drivers of pleasure, convenience and health/wellbeing (including food safety) and more recently, on environmental sustainability. The most important objective across most apple breeding programs has been to increase the marketability of the fruit by improving their eating quality and/or appearance (Laurens 1999). This covers a wide range of individual traits, which all must be considered at some stage during the breeding process.

In response to consumer-based research (e.g. Daillant-Spinnler et al. 1996; Jaeger et al. 1998), considerable effort is expended on breeding for desirable flesh textural attributes (especially high crispness and juiciness).Once picked, apple texture generally deteriorates with time, becoming soft, mealy and/or dry. While postharvest technologies during cold storage are used commercially to reduce this deterioration, they can induce other postharvest problems, such as flesh browning, and are not without cost. Increasingly, breeders are targeting genetic improvements that will assist in conserving flesh texture during long-term cold storage, with emphasis on slowing shelf life softening as well (Kouassi et al. 2009). This increases the marketing window of a new cultivar after harvest, giving much more marketing flexibility and increasing the likelihood that the consumer eating an apple has an enjoyable experience, so increasing its potential value. With the requirement for new cultivars to store for longer times, freedom from storage disorders such as bitter pit, superficial scald and rots has become increasingly important as a breeding objective.

Eating quality is also influenced by flavor. This can be deconstructed into desirable components of sweetness, acidity and aromatics. Breeding against non-desired astringency and bitterness, as well as bland low-flavor types, is also important. Although sweetness in apple is highly regarded by many consumers (Jaeger et al. 1998), there is a range of opinion amongst the public over what is a desired and highly valued flavor. Some general regional preferences are well known (e.g. higher acid for European *versus* lower acid for Japanese; Harker et al. 2003), and this may be reflected in a breeding program's objectives for specific regional taste preferences. However, strong differences in flavor preferences also exist amongst individuals within an ethnic or regional grouping, which we are only now beginning to understand (Harker et al. 2003). Perhaps from a breeding perspective, the consistency and stability of a particular flavor type during storage and shelf life is just as important as its type.

While fruit size preferences in individual countries vary considerably (e.g. consumers in the UK prefer smaller apples than the Japanese), generally apples need to weigh at least 150 g. Some exceptions can occur, such as the 50-mm apple 'PremA96', bred by Plant & Food Research and marketed as Rockit™. Fruit skin needs to be unblemished, with an absence of russet and other markings, together with an attractive and consistent color and shape.

16.2.4.2 Resistance to Pest and Diseases

Resistance to pests and diseases is an important objective in most apple breeding programs and has been for many years (Laurens 1999). The impact of a particular pathogen in any region will dictate the importance of resistance to it as a breeding objective. For instance, in the wet spring growing conditions of North Europe, most stations are breeding for resistance to apple scab (Laurens 1999), whereas in the hot dry climate of Washington this is not an objective (Evans 2012). Most programs with a resistance breeding objective are working on at least one of the major diseases/pests of apple such as, scab, powdery mildew, fireblight, European canker, rosy apple aphid, and woolly apple aphid (Laurens 1999; Lespinasse 2009; Bus et al. 2009).

Traditionally, chemical intervention has been the major method of control of most pests and diseases. However, pest and pathogen resistance to specific chemicals is an ongoing problem for apple industries, and over the last decade supermarkets are increasingly paying attention to consumer concerns about food safety and environmental sustainability. Therefore, plant resistance offers a cheaper and long-term solution to protecting trees and fruit from damage. Even now, suitable chemical control for diseases such as fireblight, which can devastate whole trees and orchards in a single season, does not exist in some regions. In such cases, development of resistant cultivars is the only means by which apples can be produced economically. Increasingly breeding is focused on pyramiding multiple resistances in new cultivars. As fungicide use declines through the introduction of cultivars as an example resistant to scab, other fungal pathogens not previously seen as important may become so.

16.2.4.3 Adaptability

Although high and consistent tree productivity is an important characteristic for new cultivars, few breeding programs in regions where apple is well adapted have this as their main objective. However, in some regions yield may be severely limited by adverse environmental conditions. Hence, we find that in areas with low winter chilling, breeding for high budbreak and therefore flower intensity is important (Labuschagné 2004), while in areas where severe autumn, winter or spring cold temperatures can occur, breeding for resistance to cold damage is a priority (Fischer and Fischer 2004). In some cases the objectives may concern fruit adaptability, including high red skin colouring under hot summer conditions (Batlle et al. 2009).

16.2.5 Genetic Control of Selection Traits

The accuracy of trait value for any one genotype predicted within a breeding programme depends upon the proportion of total phenotypic variation accounted for by genetic variation (heritability) with narrow-sense (additive) and broad sense heritability (additive + non-additive) important for parental and cultivar selection respectively

(Falconer and Mackay 1996). Simply inherited or quantitative traits with high heritability can be reliably evaluated based on single plant assessments. Conversely, traits more strongly influenced by the environment may require multiple assessments over numerous replicate clones over several sites, so that traits with a low heritability can be more accurately predicted.

Numerous major genes have been identified that confer resistance to apple scab, powdery mildew, rosy apple and wooly apple aphid (Gardiner et al. 2007; Bus et al. 2010). Accordingly, parental selection can be made and the estimated proportion of individuals with resistance to these pathogens determined relatively simply. In apple, modified backcrossing of donor material carrying major genes to recurrent cultivars or elite selections has been a common strategy for the introgression of major genes for disease and pest resistance, for many years. With rapid and effective screening for such resistances usually occurring at the juvenile plant stage, susceptible seedlings can be eliminated quickly from the programme in the glasshouse or the nursery. Complicating factors are the existence of pathogen strains that have overcome such resistances, and there is now a strong requirement to pyramid different resistances to the same pest, to ensure long-term durability of the resistance.

While major gene control has been recorded for a number of fruit and tree traits (Alston et al. 2000), relatively few of these have a strong connection with the current major selection traits related to fruit quality and tree productivity outlined above. Where they are important, there is an additional polygenic genetic component to the trait that may contribute significantly to the overall genetic variation (for instance, red skin colour coverage, russet, malic acid concentration (Brown and Harvey 1971). Other fruit quality traits such as fruit size, or texture components such as sensory and instrumental estimates of flesh firmness, crispness and juiciness and sweetness and acid taste, fruit susceptibility to storage disorders, as well as shape and yield components, including flower and fruit number, are quantitatively inherited. The limited heritability estimates for these traits that have been published indicate that most external and ripe eating quality traits (Durel et al. 1998; Kouassi et al. 2009; Kumar et al. 2010, 2011) have moderate to high (> 0.25) heritabilities. In contrast, heritabilities of some storage disorders, such as bitter pit, were found to be quite low in a broad-based breeding population (Volz et al. 2001). Taking into account the high cost of orchard planting and phenotyping, these data support the breeding strategy of selecting eating and appearance traits based on single-tree measurements, followed by cloning of the best selections for small-scale trials. Clonal replication enhances the reliability of phenotype, so allowing more robust selection to occur across a whole range of traits, including storage disorder susceptibility and yield. On the other hand, this does increase the time to cultivar development.

Yield is a complex trait that is mainly determined by fruit number on the tree, as well as fruit size. Fruit number is determined by the number of flower buds initiated on the tree and the fruit set from those flowers, and both these traits appear to be quantitatively inherited (Liebhard et al. 2003). Most breeding programs only start to assess and select for seedling yield in replicated trials, on the assumption that yield components have low heritability. Recently, Guitton et al. (2012) showed that the adverse biennial bearing character was strongly influenced by genotype relative to year in a segregating family.

16.3 Marker Assisted Selection (MAS)

16.3.1 Concept of Traditional MAS

Since the early 1990s, molecular markers have been considered as potential tools for cultivar development in various fruit crops. As most fruit crops have long juvenile periods, early genotypic selection with the aid of molecular markers promises major advantages over conventional phenotypic selection. However, molecular markers should not be considered as genes, because they often don't have any biological effect. Instead, markers should be thought of as constant landmarks in the genome, whose location on the genome is specified using a linkage map constructed using segregating full-sib families. Markers that are allelic variants of candidate genes for traits of interest can be located by testing for statistical associations with phenotype and are commonly regarded by breeders as the most effective for MAS, as the success of MAS is influenced by the relationship between the markers and the traits of interest. For these purposes, three types of observable polymorphic genetic loci can be distinguished (Andersson 2001; Dekkers 2004):

1. Direct markers: loci that code for the functional mutation. These types of markers are the most difficult to detect because causality is difficult to prove and, as a result, a limited number of examples are available, except for single-gene traits. Examples for apple are the markers that have been developed from the *Vfa2* gene conferring scab resistance, and the *MdMYB10* gene encoding the transcription factor that controls expression of 'Type 1' red flesh (Espley et al. 2009).
2. LD markers: loci that are in population-wide linkage disequilibrium (LD) with the functional mutation. Selection using these markers can be called LD-MAS. The LD markers are by necessity close to the functional mutation for sufficient population-wide LD between the marker and quantitative trait loci (QTL) to exist, which depends on population structure and history. The LD markers can be identified using candidate genes (Rothschild and Soller 1997) or fine-mapping approaches (Andersson 2001). In apple, LD markers for the two prominent fruit firmness/softening genes *ACS1* and *ACO1* have been screened on diverse apple cultivars and the best allelic combinations for selection for high fruit firmness have been identified (Oraguzie et al. 2004; Zhu and Barritt 2008; Cevik et al. 2009).
3. LE markers: loci that are in population-wide linkage equilibrium (LE) with the functional mutation in outbred populations. Selection using these markers can be called LE-MAS. The LE markers can be readily detected on a genome-wide basis by using full-sib families in outbred fruit crops. Such genome scans require only sparse marker maps to detect major loci, or most QTL of moderate to large effects. Examples of successful applications of MAS are discussed in the next section.

16.3.2 Applications of MAS in Apple Breeding Programs

The concept of using MAS to increase the efficiency of breeding in apple by increasing the precision of selection and enabling pyramiding of resistance genes, as well as reducing both the number of generations and the resources required for new cultivar development, has been well established for some time (Luby and Shaw 2001; Bassil and Lewers 2009; Bus et al. 2009; Korban and Tartarini 2009).

However, the application of MAS across apple breeding programs internationally is only now beginning. The earliest use of MAS in apple employed genetic markers flanking major gene resistance loci and involved the development and selection of parents with pyramided resistances to infection by a single pathogen, with the goal of eventually developing new cultivars with more durable resistance. These included dual resistances to apple scab such as *Rvi6* and *Rvi2* (Bus et al. 2009), or *Rvi2* and *Rvi4*, as well as the *Er1* and *Er3* resistances to woolly apple aphid infection (Bus et al. 2008; Bus et al. 2009) and powdery mildew resistances *Pl2* and *Pl1*, or *Pl1* and *Plmis* (Bus et al. unpublished). The next step involved stacking pyramided scab resistances with other resistances, for example *Rvi4* and *Rvi6* with the *FB-F7* resistance to fireblight (Baumgartner et al. 2012), or *Rvi6* and *Rvi2* with *Pl2* (Bus et al. unpublished). Furthermore, marker assisted selection of triple resistant seedlings combining a pyramided resistance to one pathogen with single resistances to two other pathogens has now been reported (Kellerhals et al. 2012). These include pyramided mildew resistances (*Pl1* and *Pl2*) with a single homozygous scab resistance (*Rvi6Rvi6*) and a heterozygous fireblight (*FB-F7*) resistance, as well as another set of seedlings carrying a pyramided resistance to scab (*Rvi4Rvi6*) and heterozygous resistances to mildew (*Pl2*) and fireblight (*FB-F7*). Newly identified breeding parents carrying homozygous *Rvi6* resistance to scab, as well as heterozygous mildew (*Pl2*) and fireblight (*FB-F7*) resistances, or a second source of fireblight resistance (*FB-E*) will enable development of breeding populations that will be 100 % heterozygous for *Rvi6*, in combination with other loci (Kellerhals et al. 2012). Selection of genetically elite seedlings carrying the transgene *BpMADS4* to enable early flowering for acceleration of breeding, combined with *FB-F7* as well as *Rvi2* and *Rvi4* scab resistances, or pyramided powdery mildew resistance (*Pl1* and *Pl2*), has recently been achieved (Flachowsky et al. 2011), as well as selection of seedlings carrying *BpMADS4* combined with *FB-E* (Le Roux et al. 2012).

Screening populations of several thousand seedlings has been facilitated by development of medium and high throughput technology for extracting apple DNA (Cook and Gardiner 2004; Frey et al. 2004) and the introduction of high resolution melting analysis as a routine tool for screening apple SNP markers (Chagné et al. 2008). Selection of seedlings for fruit traits such as red skin (Zhu et al. 2011), red flesh (Chagné et al. 2007) or reduced fruit softening (Zhu and Barritt 2008; Evans 2012), has also been reported. Some breeding programmes in the US and France, respectively employ MAS principally for parental selection for loci that include crispness, acidity, fruit color, *ACS*, *ACO*, *Vf* (Luby pers. comm.) and QTL resistances to apple scab (Laurens pers. comm.). The scion breeding programme at IASMA (Italy) will screen approximately 20,000 seedlings in 2012 with a total of 13 markers for resistances

to apple scab, fireblight, powdery mildew and for quality traits that include color, texture and storage, with each seedling being screened with two or three markers (Velasco pers. comm.). Both the United States Department of Agriculture (USDA) rootstock breeding programme at Cornell and that at Plant & Food Research (PFR) are making use of markers to select seedlings with resistance to woolly apple aphid, as well as dwarfing (Fazio, pers. comm., Malone pers. comm.). Additional selection for rooting characters will be performed in the USDA programme in the coming year. Furthermore, genetic markers are used to confirm genotype in advanced selections that have been developed using phenotypic evaluation. Examples include mildew and scab resistance in the New Zealand programme (Volz et al. unpublished), and fruit quality and storability in the Washington State University programme (Evans 2012).

The recently developed apple whole genome sequence (WGS) (Velasco et al. 2010) is proving itself as a resource for increasing the understanding of the genetic control of traits of agronomic importance, such as texture (Longhi et al. 2012), flavour (Wiedow et al. 2010; Dunemann et al. 2012), flesh colour (Espley et al. 2009), timing of budbreak (Celton et al. 2011), biennial bearing (Guitton et al. 2012), fruit weight (Devoghalaere et al. 2012), disease resistance (Gardiner et al. 2012), and fruit polyphenolic composition (Chagné et al. 2012a; Khan et al. 2012). The US-based international research programme 'RosBREED' ('Enabling marker assisted breeding in Rosaceae') is an outstanding example of a highly focused industry-researcher collaboration which has enabled the development of the International Rosaceae SNP Consortium (IRSC) 8K apple Infinium® SNP chip (Chagné et al. 2012a). The international availability of this single nucleotide polymorphism (SNP) array has opened the door to speedy construction of genetic maps in diverse bi-parental populations (Chagné et al. 2012a) (Troggio, pers. comm.) for subsequent QTL mapping and candidate gene/allele identification. The European programme 'FruitBreedomics', which focuses on bridging the gap between breeding and genomics, is developing an expanded apple array that will be based on the IRSC array and have functionality for association mapping.

16.3.3 Limitations of MAS

Impediments to the wider use of MAS include the relatively small number of traits for which genetic markers (flanking loci or candidate gene/allele markers) are available. Although the apple whole genome sequence and the derived SNP chip have the potential to increase the rate of marker discovery greatly in bi-parental populations, it is not possible to develop and phenotype the number of relatively large (> 200 progeny) bi-parental populations that would be needed to map genetically complex traits of agronomic importance to new cultivar development, across the breeding apple germplasm. Because of the low extent of LD in apple, flanking markers are needed for each locus, unless a candidate gene is obvious from inspection of the WGS in the QTL interval, thus doubling the number of markers needed for selecting at a locus and increasing costs. Furthermore, any specific marker identified in one

population is not heterozygous in all populations and hence cannot be assumed to segregate in any given breeding population (Micheletti et al. 2011). This issue of lack of transferability is particularly relevant to the SNP markers that are currently favoured because of their ease of screening, but also applies to single sequence repeat (SSR) markers (Patocchi et al. 2009). As well, the causative mutation may occur in a gene that is not easily recognized as a candidate, making development of gene/allele markers difficult. Because multiple genes may influence expression of complex traits (Kumar et al. 2012b), such as fruit firmness (Zhu and Barritt 2008) and skin colour (Zhu et al. 2011), correlation between a specific gene marker and a trait is often less than perfect across the germplasm.

At the moment, estimated breeding values (EBV)-based selection (which takes account of the collective effect of all genes) remains an important selection tool for breeding for quantitative (or complex) traits. Thus, for complex traits, the LD markers showing significant association with a trait can be incorporated along with a polygenic effect (to account for the genetic variance not captured by the markers) in the calculation of BLUP-BV (Fernando and Grossman 1989; Meuwissen and Goddard 1996). However, such approaches are not being used in apple breeding programs, largely because of the non-availability of large-effect LD markers for quantitative traits. The solution to the limitations of traditional MAS perhaps lies not in seeking single markers associated with single large effects, but in capitalizing on developing the capacity for identifying and scoring tens-of-thousands of markers at low cost (Jannink et al. 2010; Yang et al. 2010). The availability of genome sequences along with high throughput genotyping platforms is changing the nature of research undertaken to understand evolution of organisms, as well as transforming the strategies for genetic improvement. One such selection strategy, genomic selection (GS), has the potential to change the genetic improvement of animal and plant species.

16.4 Genome Sequencing and Genotyping Platforms

Deciphering genetic variations linked with agronomic traits requires large sets of markers that can be screened over large numbers of segregating individuals. What was a huge challenge in the 1980s, when only a handful of isoenzyme markers were available for plant species, is now history. The entire repertoire of DNA variations is now accessible for most major crops, including apple, thanks to enormous advances in high-throughput sequencing techniques and bioinformatics methods. Possession of a full repertoire of variants is a goldmine for marker development for both MAS and GS.

16.4.1 Genome Sequencing

The most efficient method for detecting SNPs is to search for them in sequence databases made from cDNA transcripts or whole genome. Bioinformatics tools and pipelines are developed almost on a weekly basis for the *in silico* detection of SNPs

and a review of these tools is not the purpose of this chapter, as they improve so fast. The first large-scale effort for sequencing significant chunks of apple genome was the sequencing of hundreds of thousands of expressed sequence tags (ESTs) (Newcomb et al. 2006; Gasic et al. 2009). Although this focused only on transcribed sequences, it was possible to detect sequence variation corresponding to putative SNPs (Chagné et al. 2008). On average, one SNP every 149 bp was detected in an apple EST set comprising mainly 'Royal Gala' cDNA sequences. The development of EST-based SSRs was also popular at the time these large EST data sets became available (Celton et al. 2009; Silfverberg-Dilworth et al. 2006).

The sequencing of the apple genome by an international consortium led by IASMA was performed on 'Golden Delicious' (GD) (Velasco et al. 2010), as a common founder of most breeding programs. The draft genome of GD is one of the early examples of the sequencing of a highly heterozygous genome. This heterozygosity is due to the self incompatibility locus in apple, which contributes to maintenance of a heterozygous state in most of the genome. The GD genome assembly that spanned a total of 598 Mb, anchored on 17 pseudo chromosomes, yielded more than 2 million SNPs. These SNPs provided a tool box of potential markers that are now available to be used by breeders for selection of seedlings carrying traits inherited from GD. The first set of apple multiplex SNP markers was based on the SNPlex™ Genotyping System, which is capable of screening 48 markers simultaneously (Pindo et al. 2008). The markers were derived from genomic electronic GD SNPs and 1489 were employed to anchor the genome sequence of GD to an integrated genetic map derived from six segregating populations.

16.4.2 Highly Parallel Genotyping

Modern genotyping techniques are capable of screening thousands of SNP markers simultaneously, in a density sufficient for multiple trait MAS and GS, as well as QTL mapping. The current cost for such an assay per data point is much lower than single-plex techniques such as SSRs and more amenable to screening a large number of individuals than whole genome re-sequencing. A subset of GD SNPs designed using the 48-plex SNPlex™ assay was transferred to a Golden Gate™ (Illumina) assay of 384 markers (Troggio, unpublished) that is useful for speedy genetic map construction. A second Golden Gate™ assay was developed independently from the apple genome sequence of GD (Khan et al. 2012), from SNPs detected in apple cDNA sequences. While this second assay represents a four-fold increase over the 384-plex assay, its design is based on a majority of ESTs derived from two cultivars: 'Royal Gala' and 'GoldRush'. The application of both SNP assays is therefore limited to genetic populations related to these three cultivars, since SNP transferability should be similar to the success rates that Micheletti et al. (2011) reported using the 48-plex assays derived from GD.

The strategy that improved the success of transferring SNPs across apple germplasm involved initial re-sequencing of a set of 27 apple accessions that represent most of the genomic variability in breeding programs worldwide. In total, 2,113,120 SNPs were detected and used by the IRSC to develop a new assay with 8,000 apple SNPs, based on the Infinium II® technology (Illumina) (Chagné et al. 2012a). The design of the IRSC assay had the goal of optimizing the SNP transfer success, without compromising the marker density. The design strategy consisted of choosing SNPs based on a range of SNP minor allelic frequencies (MAF) in the set of 27 accessions. The MAF ranged from frequent alleles (MAF ~ 0.5), to rarer accession-specific SNPs (MAF < 0.1). To increase the chance of identifying polymorphic loci further, the IRSC assay was designed to span clusters of up to 10 SNPs within a 100 kb genomic region centered on validated GD SNPs. While this strategy slightly increased the physical distance between markers, it guaranteed the success of there being at least one polymorphic SNP every centiMorgan, in any given accession. Preliminary SNP genotyping using the IRSC assay over a set of 1,600 individuals from Plant & Food Research, including F_1 populations and germplasm accessions with no direct relation to the 27 accessions used for sequencing, indicated that a large proportion of the 8,000 attempted SNPs are polymorphic ($\sim 70\,\%$).

Another important aspect in the design of the 8,000 IRSC assay system was SNP density and distribution in the apple genome. The IRSC assay was developed to locate at least one SNP marker every one centiMorgan on the apple genetic map, which corresponds to one SNP every 450 kb based on the GD genome assembly. This density is sufficient for high quality genetic map construction and for GS. Preliminary mapping analysis (Chagné, unpublished) suggests that the IRSC is highly efficient for dense genetic map construction and suitable for QTL analysis. Nevertheless, a significant number of SNPs (15 %) do not map genetically to their expected location in the GD genome assembly. This highlights the usefulness of the IRSC assay for developing future improved genome assembly versions for apple. In the near future, a new assay of 20,000 SNPs will be developed and this will include both the validated SNPs from GD and the IRSC polymorphic SNPs. Additional sequencing data from European breeding founders not used for the IRSC design will be added to span even more of the apple genomic variability.

16.4.3 The Future of Molecular Breeding: Genotyping by Sequencing

The cost per data point is constantly decreasing, not only for highly paralleled targeted genotyping (SNP assays) but also for sequencing. Human geneticists have started to use genotyping by sequencing (GBS) instead of large genotyping platforms, as it gives the complete set of variants common across individuals, as well as rare variants and structural variants such as copy number variants that are not identified in SNP detection panels. Some apple breeding programs have started to generate whole genome re-sequencing for some of their best breeding selections, or for germplasm accessions carrying unique traits, to add to the 27 apple accessions that were re-sequenced using low coverage re-sequencing for the development of the IRSC assay.

Hence, the era of utilizing GBS for breeding has indeed begun for apple. The next (big) step on this progression is to initiate the screening of large seedling populations using GBS. The current cost for sequencing and most importantly the limitations due to computing requirements for using such approach makes this prohibitive at the moment, however it is likely that GBS of breeding populations will be employed widely in the future.

16.5 Genomic Selection (GS)

16.5.1 Concept of GS

GS is a form of MAS that utilizes all available genome-wide markers simultaneously, calculating their association to the trait phenotypes, to estimate BV (the additive genetic component) or total genetic value (GV, which includes additive and non-additive genetic components). The difference between traditional MAS and GS is that MAS only utilizes the SNPs that are significant in a genome-wide marker-trait association study, whereas GS uses a genome-wide panel of dense markers so that all QTL are in population-wide LD with the markers. Since QTL positions for various traits of interests are unknown, equal spacing of markers across the whole genome is considered to be optimal. The main advantage of GS over MAS is that potentially all the genetic variance for a trait can be tracked by the marker panel, because the marker effect does not need to exceed the significance threshold to be used to predict breeding value. GS starts with a 'training population' of individuals with both genotypic and phenotypic data, to establish a prediction equation. Subsequently, this prediction equation is then applied to genotypic data from a 'selection population' of non-phenotyped individuals, and used to produce genomic estimated breeding values (GEBVs) (Meuwissen et al. 2001). As GS produces a single breeding value for each individual, it de-emphasizes the contribution of individual genes to a quantitative trait. The GEBVs are used as a selection tool to assist in choosing next-generation parents, or selections for further testing as potential commercial cultivars, similarly to the classical selection approach that takes account of the collective effect of all genes. Given that GS can be used to identify outstanding candidates at a very early stage, prior to extensive fruit-quality phenotyping, it has the potential to accelerate breeding efficiency significantly as a result of decreased generation interval or increased selection.

16.5.2 Genomic Selection Models

In a classical quantitative genetics framework, the phenotype is expressed as:

$$y = Z_1 b + Z_2 a + e \tag{16.1}$$

where $\boldsymbol{y}$ is a vector of observations on a trait, $\boldsymbol{b}$ is a vector of fixed effects (i.e., overall mean, replicate), $\boldsymbol{a}$ is a vector of random additive genetic effects of individual genotypes; $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$ are known incidence matrices relating the observations in $\boldsymbol{y}$ to effects in $\boldsymbol{b}$ and $\boldsymbol{a}$, respectively. The variances associated with the random effects $\boldsymbol{a}$ and $\boldsymbol{e}$ are σ_A^2 (additive variance) and σ_e^2 (residual variance) respectively. Genetic relationships among genotypes (i.e. random effects in $\boldsymbol{a}$) are taken into account via the numerator relationship matrix (also termed as additive genetic relationship matrix), in solving the mixed model equations (Henderson 1984).

The basic model for genomic selection can be written as:

$$y = Zb + Xg + e \tag{16.2}$$

where $\boldsymbol{y}$ is a vector of n phenotypes (or BLUP-BVs); $\boldsymbol{X}$ is a $(n \times m)$ design matrix allocating records to the m SNP effects, with element $X_{ij} = 0$, 1, or 2 if the genotype of seedling i at SNP_j is AA, AB, or BB, respectively; $\boldsymbol{g}$ is a $(m \times 1)$ vector of SNP effects. GEBV are then estimated for the validation population as: $\boldsymbol{Xg}$, using estimate of $\boldsymbol{g}$ from Eq. 2.

16.5.3 Factors Affecting Accuracy of Genomic Selection

The accuracy of a GS method is the correlation between the predicted values (GEBVs) and true breeding values (TBVs). A theoretical calculation of accuracy of GEBV, assuming the model used sufficiently dense markers (i.e., all SNPs have an effect from a normal distribution), has been given by Goddard (2009):

$$Accuracy = (1 - \lambda/(2T\sqrt{a}) * \log((1 + a + 2\sqrt{a})/(1 + a - 2\sqrt{a})))^{0.5} \tag{16.3}$$

where $a = 1 + 2\lambda/T$, T is the number of (Training) observations, $\lambda = qk/h^2$, q is the number of independent chromosome segments (effective number of loci) in the population, $k = 1/\log(2Ne)$, h^2 is the heritability of the phenotype used and N_e is the effective population size. Hayes et al. (2009a) recommend using $q \approx 2NeL$, where L is the genome length. When Hayes et al. (2009b) compared this formula with reported accuracies from real data calculated using validation samples for US and Australian Holstein and Jersey dairy cattle, they found that the theoretical formula agreed well with the reported accuracies, using estimated values of N_e from the recent past.

TBVs are unknown in practical situations, and thus in empirical studies the accuracy of GS is calculated using a cross-validation method. The method consists of splitting the data (phenotypic and marker) into a training data set and a validation data set. A key feature in cross-validation is the choice of the training and validation sets. The first choice is the size of each set, as there is a trade-off between precision of the model in the training set and over-fitting in the validation set. Usual recommendations are that the validation set should be one-fifth or one-tenth of the full data set (Legarra et al. 2008). The second and more critical choice is how to split the data

into training and validation sets. For some crops (including apple), a typical training population is generally composed of several full-sib families with varying degrees of genetic relationship among them. In such cases, there are basically three options for splitting the data. The first option is to sample whole families, i.e. allocate whole families randomly into either training or validation sets. The second option is to split each family randomly so that parts of each are in both training and validation sets. When training and validation sets are composed of different families, capturing marker-QTL LD will be more useful for prediction because of supposedly low relatedness between families. On the other hand, when training and validation sets are members of the same families, capturing the genetic relationship plays a bigger role in the accuracy of GS. The third option is to ignore the family structure and split the entire population randomly into training and validation sets. In this last option, all families would be represented in both datasets; however, their relative representation could vary if family sizes were quite different.

Pearson's correlation between the GEBV and observed EBV is commonly used to evaluate the performance (or accuracy) of the GS model because it is assumption free and comparable across models. The observed EBVs are regressed on the GEBV, where the regression coefficient reflects the degree of bias of the GEBV prediction. A regression coefficient of one indicates no bias. Various factors affecting accuracy of GS are discussed next.

16.5.3.1 Prediction Models

An important challenge in GS is building the predictive model of complex phenotypes, because such traits can be influenced by many loci. In their pioneering study, Meuwissen et al. (2001) showed that Bayesian estimation methods yielded somewhat higher accuracy of GEBVs than linear models (e.g. random regression (RR-BLUP) for simulated datasets. A number of GS models that vary in their basic assumptions about the distribution of SNP effects have been proposed (Table 16.1). Typically, the number and individual contributions of genes affecting a quantitative trait is unknown in practice. The influence of the genetic architecture on the prediction performance of alternative GS models has been reported in simulation studies. For example, Coster et al. (2010) reported a decrease in prediction accuracy for penalized and Bayesian models when the number of simulated QTLs increased. The accuracy decreased further when variation in the individual contributions of the QTLs increased, except for a method called Partial Least Square Estimation, which seemed robust to these perturbations (Coster et al. 2010). Daetwyler et al. (2010) reported that accuracy of RR-BLUP was not affected by the number of QTLs in the data, and in situations where few QTLs contribute to the trait, accuracies obtained with Bayesian methods are higher than accuracies obtained with RR-BLUP. However, at high number of QTLs these accuracies are identical, suggesting that Bayesian methods will always perform equally or better than RR-BLUP.

However, results from simulation studies are not always supported by those based on empirical data. In an empirical application of GS in the New Zealand dairy cattle

Table 16.1 Assumptions of some methods for deriving the single nucleotide polymorphism (SNP) prediction equation for calculation of genomic breeding values in genomic selection.

Name	Reference	Key assumptions and features
RR-BLUP	Meuwissen et al. (2001)	Normal distribution of SNP effects with a common marker effect variance; a very large number of QTL with small effect
BayesA	Meuwissen et al. (2001)	A *t*-distribution of SNP effects (B_j); most QTL effects are close to zero, but a few have very large effect
BayesB	Meuwissen et al. (2001)	Mixture distribution of zero effects (with probability $1-\pi$) and *t*-distribution of effects (with probability π); SNP with non-zero effects are sampled similar to that in BayesA
BayesC	Kizilkaya et al. (2010)	*A priori* where with probability π, the genes are normally distributed and with probability $(1-\pi)$ they are 0
BayesCπ	Lorenz et al. (2010)	Mixture distribution of zero effects and normal distribution of effects; assumes a common marker effect variance for all markers with nonzero effects, but rather than using a fixed π, it estimates π
Bayesian LASSO	Yi and Xu (2008)	Double exponential distribution of effects; very large proportion of SNP with effect close to zero, small proportion with moderate to large effect
Empirical Bayes	Xu (2007)	Similar to BayesA; instead of using MCMC to estimate the variance parameters, a maximization algorithm is used to reduce computation time
RKHS	Crossa et al. (2010)	Uses a kernel function to convert the marker dataset into a set of distances between pairs of observations. In a finite-dimensional setting, this amounts to modeling the vector of genetic values as multivariate normal
Machine Learning methods	Moser et al. (2009)	These are essentially nonparametric methods
PLSR	Moser et al. (2009)	Don't require a prior distribution of effects to be specified; computationally less demanding as it uses reduced dimensions of data

breeding programme, Harris et al. (2008) reported poor accuracy of a penalized regression method while Bayesian models slightly outperformed linear models. Also in dairy cattle, Moser et al. (2009) found similar degrees of accuracy for five penalized, Bayesian and semi-parametric predictive models. In agricultural crops, Lorenzana and Bernardo (2009) showed that RR-BLUP provided comparable or higher prediction accuracies than a Bayesian model. Similar conclusions have been drawn from studies on forest tree species (e.g. Resende et al. 2012a). Heslot et al. (2012), using eight wheat, barley and maize datasets, compared 11 GS models (including RR-BLUP, various Bayesian methods, and machine learning methods) and concluded that a similar degree of accuracy was observed for many models, but the computation time and the distribution of marker effect estimates varied. An excellent review of comparative performance of GS models across various plant species can be found in Lorenz et al. (2011).

16.5.3.2 Training Population

Meuwissen et al. (2001) showed that when the number of individuals was reduced from 2,200 to 500, the correlation between true breeding values and GEBV was reduced by 17–61 % for various GS models. Coster et al. (2010) reported that design parameters, such as the number of individuals in the training population, the number of SNP markers, and trait heritability could have variable effects on accuracies of different methods in calculating GEBVs. In dairy cattle, the size of the training population and trait heritability affected accuracy (0 to 1 scale) of GS for a number of traits with a range from 0.63 to 0.84 (Harris et al. 2008; VanRaden et al. 2009; Hayes et al. 2009c). Habier et al. (2010) suggested that a lower accuracy with decreasing training set size could be the result of a lower number of relatives in training. Meuwissen (2009) reported that the genetic relationships between training individuals and selection candidates would provide higher GEBV accuracy in practical applications, than where there was no such relatedness.

When the training data for GS consist of individuals from reproductively isolated ancestral populations, estimates of marker effects may be biased because of population stratification and admixture. Habier et al. (2007) reported that the presence of complex pedigrees in training data can yield population stratification and may reduce prediction accuracy because of confounding of relatedness with SNP effects. However, provided high density SNPs are used and analyzed simultaneously, as in GS, admixed populations can be used to develop reliable GS prediction equations even if pedigree and breed (or population) origin has not been explicitly modeled to avoid spurious signals (Toosi et al. 2009).

16.5.3.3 Marker Density and Linkage Disequilibrium

Increasing SNP density has been shown to yield more accurate predictions of BVs for all traits in dairy cattle (VanRaden et al. 2009). When using a lower SNP density (say,

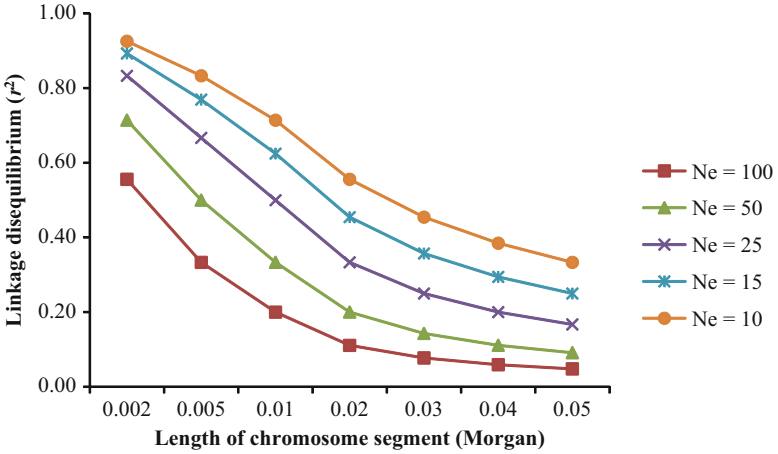

Fig. 16.2 The extent of LD (as measured by r^2) for increasing chromosome segment length, for varying effective population sizes (*Ne*)

only one SNP per cM), Calus and Veerkamp (2007) suggested including polygenic effects in the GS models, especially for lower heritability traits. Accuracies were improved by including a polygenic effect in the model for the lower heritability traits, when the LD-measure r^2 between adjacent markers became smaller than 0.10. GEBV accuracy was higher than the accuracy of standard pedigree-based BLUP-EBV only if there was LD between SNPs and QTLs. Without LD, the accuracy is expected to decline fast in the generations following the estimation of marker effects. In case of linkage equilibrium, the accuracy of GEBVs is not necessarily zero, but will approach the accuracy of pedigree-based BLUP-EBV as the number of SNPs fitted in the model increases, because SNPs capture additive-genetic relationships (Habier et al. 2007).

To understand better the optimal SNP densities for GS in fruit crops, we first need to understand the structure of LD. With the availability of huge amount of genome-wide SNPs, researchers are now beginning to unravel the patterns of LD in various fruit crops (e.g. Aranzana et al. 2010; Kumar et al. 2012b). Assuming that the finite sample size of the founders of the current breeding populations could be the key driver of LD, it is possible to derive a simple expectation for the amount of LD (measured by r^2) for a given size of chromosome segment. This expectation is:

$$E(r^2) = 1/(4N_e c + 1) \tag{16.4}$$

where N_e is the effective population size, and c is the length of the chromosome segment in Morgans (Sved 1971). This equation predicts rapid decline in LD as genetic distance increases, and this decrease will be larger with large effective population sizes (Fig. 16.2).

Table 16.2 Ranges of accuracies for genomic estimated breeding values (GEBVs) across traits as estimated in different countries for various sizes of the reference population in dairy cattle (Calus 2010)

Reference	Size reference population	Number of SNPs included	Range accuracy of GEBVs
Harris et al. (2008)	2,490	44,146	0.63 to 0.82
Berry et al. (2009)	596	42,598	0.56 to 0.71
De Roos et al. (2009)	3,600	48,000	0.52 to 0.82
Lund and Su (2009)	1,238	38,055	0.55 to 0.85
Schenkel et al. (2009)	3,966–4,127	38,416	0.36 to 0.77
VanRaden et al. (2009)	5,335	38,416	0.44 to 0.79

SNP single nucleotide polymorphism

16.5.4 GS Studies in Commercial Species of Animals and Plants

Among various commercial animal species, only in dairy cattle has application of GS been widely studied. The accuracies derived in the various studies (Table 16.2) were obtained using different methods, apply for different groups of animals and may be for GEBVs, or GEBVs blended with traditional proofs (or parental averages). Although inclusion of many different traits and use of different methods to assess the accuracy in different studies make comparison of the obtained accuracies difficult, the reported maximum accuracy of each study was for a milk production trait. Those maximum values per study show that having 600 reference bulls resulted in lower accuracies (0.71) than when using at least 1200 reference bulls (accuracy > 0.77). Despite the limited number of studies in Table 16.2, and differences among the studies, the results suggest that there is a trend of increasing accuracy of GEBVs with increasing numbers of bulls in the reference population. Therefore, it seems advisable to include > 1000 bulls in the reference population to obtain GEBVs for juvenile selection candidates with accuracies that are higher than those for pedigree indexes for all traits (Calus 2010). A review of studies from various dairy cattle GS experiments also showed that GS methods predicted breeding values better than pedigree information alone (Hayes et al. 2009c).

Studies in crop plants indicate that GS could be an extremely useful tool for plant breeding (reviewed by Heffner et al. 2009). Simulation (Wong and Bernardo 2008; Zhong et al. 2009) and empirical studies (Lorenzana and Bernardo 2009) demonstrated that in plant populations GS would lead to greater gains per unit time than phenotypic selection. Furthermore, studies comparing GS with current MAS approaches support the superiority of GS for improving gain per unit time. A study comparing GS to MAS in a simulated maize breeding programme found GS to have increased response from selection, especially for traits of low heritability (Bernardo and Yu 2007). Similarly, a simulation in oil palm found GS to have higher gains than MAS even at population sizes feasible for tree species (Wong and Bernardo 2008). A brief overview of study designs and GS accuracies in more recent studies is shown in Table 16.3. Crossa et al. (2010) evaluated GS in 599 historical wheat

lines and 284 maize inbreds from the International Maize and Wheat Improvement Center (CIMMYT). Using multiple GS models and environments, accuracies of GS for wheat grain yield ranged from 0.36 to 0.61, for maize flowering time ranged from 0.46 to 0.79, and for maize grain yield ranged from 0.42 to 0.53 (Crossa et al. 2010). Results from these studies strongly support the utility of GS in plant breeding because deterministic simulation has shown that if GS accuracy for net merit (i.e., overall performance) exceeds 0.50, GS could greatly outperform conventional MAS in terms of gain per unit time and cost (Heffner et al. 2010).

Breeding cycles of most forest tree species are much longer, largely because of the extended period necessary for reliable phenotypic expression of selection traits. Incorporation of early selection based on GS prediction models can increase genetic gain per generation through early selection. A few studies on the potential of GS in forest tree species have been reported recently. Using a training population of 800 loblolly pine (*Pinus taeda*) trees genotyped with about 5,000 SNPs, Resende et al. (2012b) reported accuracies of GS ranging from 0.65 to 0.75 for diameter, and 0.63 to 0.74 for height in different environments. In two separate training populations (size 738 and 920) of *Eucalyptus* species, predicted accuracies of GS ranged from 0.55 to 0.88 for various economic traits (Resende et al. 2012c). This latter study also showed that GS models had no appreciable accuracy when applied to an unrelated selection population. Superiority of GS over phenotypic selection was also demonstrated in a simulation study in *Cryptomeria japonica* (Iwata et al. 2011).

16.6 Implementation of GS in Apple Breeding

In New Zealand conditions, it generally takes about 7 years from apple seed before outstanding individuals can be identified for further use as a parent, or as a potential stage 2 cultivar (Fig. 16.1). GS could be used for selecting next-generation parents and/or selecting potential future cultivars. In the former case, one would want to predict the BV (only the additive effects), while in the latter case, prediction of total genetic value (additive, and non-additive) is required. BV can be predicted by fitting only the additive effects at each SNP locus; however, the accuracy of predicting the total genetic value could potentially be improved by including non-additive effects, depending on the proportion of total genetic variance explained by these factors. Kumar et al. (2012a) outlined a strategy for implementing GS in apple that would help to identify outstanding seedlings for cultivar evaluation much earlier, thus increasing the efficiency of the breeding system (Fig. 16.3).

Kumar et al. (2012b) conducted an empirical study for evaluating accuracy of GS in the PFR apple cultivar breeding programme. The training dataset used for this purpose was obtained from an existing population generated in 2004 using a factorial mating design involving four female parents with two male parents. The number of offspring per full-sib family varied from 40 to 353, with a total sample size of about 1,200. These seedlings were assessed for a range of fruit quality traits (e.g. fruit firmness, soluble solids, russet, astringency, titratable acidity) using instrumental, sensory, or visual assessment methods (Kumar et al. 2012b). The training population was genotyped using the IRSC apple 8 K SNP array v1 (Chagné et al. 2012a).

Table 16.3 Ranges of accuracies for genomic estimated breeding values (GEBVs) across traits as estimated in different populations for various sizes of the reference population in agricultural crops

Crop (Reference)	Training population size	Number of SNPs included	Range accuracy of GEBVs
Wheat (Heslot et al. 2012)	551	319	0.53–0.64
Maize (Heslot et al. 2012)	332–370	319–355	0.52–0.75
Barley (Heslot et al. 2012)	761–911	338–2146	0.53–0.57
Maize (Albrecht et al. 2011)	1,380	1152	0.72–0.79
Wheat (Crossa et al. 2010)	599	1,279[a]	0.36–0.61
Maize (Crossa et al. 2010)	284	1,148	0.42–0.79

[a]DArT, diversity array technology markers (Triticarte Pty. Ltd. Canberra, Australia)
SNP single nucleotide polymorphism

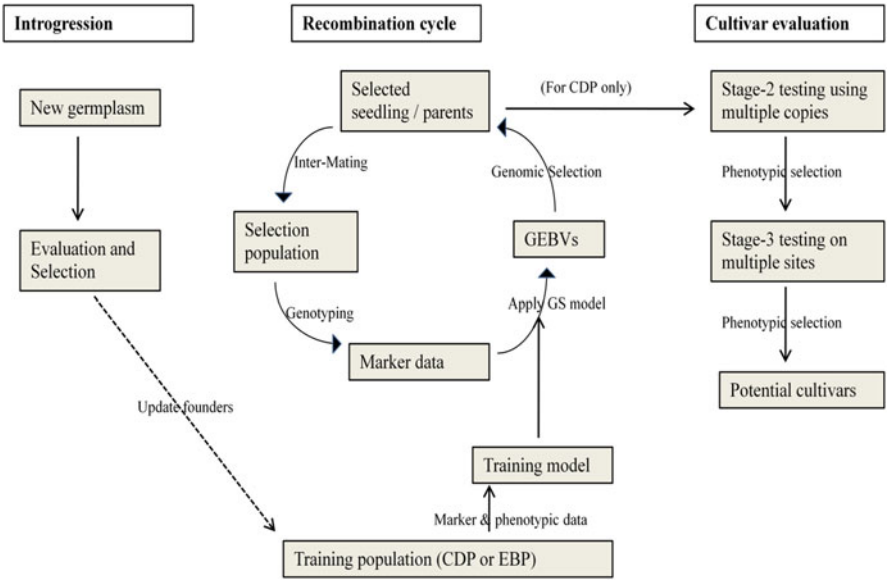

Fig. 16.3 A schematic representation of a recurrent genomic selection scheme for a typical apple breeding programme. *CDP*: cultivar development population; *EBP*: elite breeding population; *GS*: genomic selection; *GEBV*: genomic breeding value

Accuracies were very similar for the two contrasting GS models (RR-BLUP and Bayesian LASSO), and varied from 0.70 to 0.90 for various fruit quality traits (Kumar et al. 2012b). The selection response per unit time using GS compared with the traditional BLUP-based selection was very high (> 100 %), especially for low-heritability traits. These results demonstrated that genomic selection is a credible alternative to conventional selection for fruit quality traits. GS showed strong potential to accelerate the breeding cycle by making selections prior to extensive fruit quality phenotyping. Realized accuracies of operational GS in PFR's apple breeding programme will be available in the next few years, which should help to boost confidence in this revolutionary selection technique further.

16.7 Challenges for GS

GS offers opportunities for faster introgression of new traits into new commercial apple cultivars with added health and adaptability benefits. However, development of robust GS models will require assembling large training populations catering for genes of small effects, genetic diversity of breeding populations, and the magnitude of genotype-by-environment interaction. The literature on GS in animal and other plant species will guide fruit breeders in designing appropriate reference populations. Depending on the extent of LD, training populations will need to be genotyped with high-density SNP arrays, which will be a substantial cost for any apple breeding programme. Now that gene technologies are advancing rapidly, a huge challenge for apple breeders is the high-throughput phenotyping of large populations in order to further our understanding of which genes influence which physical traits. Lack of such technology would constitute a 'phenotyping bottleneck' that would limit our ability to capitalize on substantial investments made in genome sequencing. Near-infrared spectroscopy on agricultural harvesters and spectral reflectance of plant canopies are among some novel phenotyping platforms (Montes et al. 2007).

Simulation and empirical studies show that GS can accelerate short-term gain; however, there are still several unknowns for long-term gains. Marker and QTL alleles will recombine, and their frequencies will shift, changing linkage disequilibrium (LD) between them and therefore the predictive ability of the markers. Hence a valid question concerns how often marker effects will have to be re-estimated and new markers identified. A practical approach to re-estimate SNP effects would be to add continuously (to the existing training dataset) phenotypic information that becomes available for candidates that were selected as juveniles based on their GEBVs. In the time required to obtain this additional information, one or two generations of selection might be carried out based on the GS model developed from the current training dataset (Calus 2010). As new germplasm (such as wild accessions) is frequently introduced for introgression of new traits in breeding programs, SNP arrays will need to be updated accordingly to avoid ascertainment bias. As GS uses only markers that have been estimated to have an effect on the target trait, some of the favourable alleles that are in low LD with any marker will inevitably be lost. A solution to this problem would be to use an index selection, whereby a weight assigned to a marker would vary according to its frequency, so that markers for which the favorable allele has a low frequency receive more weight in the index (Goddard 2009; Jannink 2010).

Although the genotyping cost per data point is constantly decreasing, still genotyping with high density SNP arrays will be a substantial cost for any apple breeding programme. So, the use of GS in practical breeding programme will depend on a cost-benefit analysis of recording of DNA samples against expected additional economic gains.

References

Albrecht A, Wimmer V, Auinger H-J et al (2011) Genome based prediction of testcross values in maize. Theor Appl Genet 123:339–350

Alston FH, Phillips KL, Evans KM (2000) A *Malus* gene list. Acta Hort 538:561–570

Andersson L (2001) Genetic dissection of phenotypic diversity in farm animals. Nat Rev Genet 2:130–138

Aranzana MJ, Abbassi EK, Howad W, Arús P (2010) Genetic variation, population structure and linkage disequilibrium in peach commercial varieties. BMC Genet 11:69

Basil N, Lewers K (2009) Genomics opportunities, new crops and new products. In: Folta KM, Gardiner SE (eds) Genetics and genomics of the Rosaceae. Springer, New York, pp 55–70

Batlle I, Lozano L, Iglesias I et al (2009) The IRTA-HortResearch apple scion breeding programme: aiming for high fruit quality under warm growing conditions. Acta Hort 814:209–214

Baumgartner IO, Leumann LR, Frey JE et al (2012) Breeding apples to withstand infection pressure by fireblight and other diseases. In: 15th International Conference on Organic Fruit-Growing, Hohenheim, Germany, 20th–22nd February. www.ecofruit.net

Bernardo R, Yu J (2007) Prospects for genome-wide selection for quantitative traits in maize. Crop Sci 47:1082–1090

Berry DP, Kearney F, Harris BL (2009) Genomic Selection in Ireland. Proceedings of the interbull international workshop—genomic information in genetic evaluations, Uppsala, Sweden, Bulletin no. 39

Bringhurst RS (1983) Breeding strategy. In: Janick J, Moore JN (eds). Methods in fruit breeding. Purdue University Press, West Lafayette, pp 147–153

Brown AG (1975) Apples. In: Janick J, Moore JN (eds) Advances in fruit breeding. Purdue University Press, West Lafayette, pp 3–37

Brown AG, Harvey DM (1971) The nature and inheritance of sweetness and acidity in the cultivated apple. Euphytica 20:68–80

Bus VGM, Chagne D, Bassett HCM et al (2008) Genome mapping of three major resistance genes to woolly apple aphid (*Eriosoma lanigerum* Hausm.). Tree Genet Genomes 4:223–236

Bus VGM, Esmenjaud D, Buck E, Laurens F (2009) Application of genetic markers in rosaceous crops. In: Folta KM, Gardiner SE (eds) Genetics and genomics of the rosaceae. Springer, New York, pp 563–600

Bus VGM, Bassett H, Bowatte D et al (2010) Genome mapping of an apple scab, a powdery mildew and a woolly apple aphid resistance gene from open-pollinated mildew immune selection. Tree Genet Genomes 6:477–487

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

Calus M, Veerkamp R (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 124:362–368

Celton JM, Tustin DS, Chagne D, Gardiner SE (2009) Construction of a dense genetic linkage map for apple rootstocks using SSRs developed from *Malus* ESTs and *Pyrus* genomic sequences. Tree Genet Genomes 5:93–107

Celton JM, Martinez S, Jammes MJ et al (2011) Deciphering the genetic determinism of bud phenology in apple progenies: a new insight into chilling and heat requirement effects on flowering dates and positional candidate genes. New Phytol 192:378–392

Cevik V, Ryder CD, Popovich A et al (2009) A FRUITFULL-like gene is associated with genetic variation for fruit flesh firmness in apple (*Malus domestica* Borkh.). Tree Genet Genomes 6:271–279

Chagné D, Carlisle C, Blond C et al (2007) Mapping a candidate gene (*MdMYB10*) for red flesh and foliage colour in apple. BMC Genomics 8:212

Chagné D, Gasic K, Crowhurst RN et al (2008) Development of a set of SNP markers present in expressed genes of the apple. Genomics 92:353–358

Chagné D, Crowhurst RN, Troggio M et al (2012a) Genome-Wide SNP Detection, Validation, and Development of an 8 K SNP Array for Apple. PLoS ONE 7(2):e31745

Coart E, Vekemans X, Smulders MJM et al (2003) Genetic variation in the endangered wild apple (*Malus sylvestris* (L.) Mill.) in Belgium as revealed by amplified fragment length polymorphism and microsatellite markers. Mol Ecol 12:845–857

Coart E, Van Glabeke S, De Loose M et al (2006) Chloroplast diversity in the genus *Malus*: new insights into the relationship between the European wild apple (*Malus sylvestris* (L.) Mill.) and the domesticated apple (*Malus domestica* Borkh.). Mol Ecol 15:2171–2182

Cook M, Gardiner S (2004) Development of a fully automated system to extract DNA from difficult plant tissues for genomics research. In: Plant & Animal Genome XII Conference, San Diego, CA (Abstract P180)

Coster A, Bastiaansen JWM, Calus MPL et al (2010) Sensitivity of methods for estimating breeding values using genetic markers to the number of QTL and distribution of QTL variance. Genet Sel Evol 42:9

Crossa J, de los CG, Pérez P et al (2010) Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. Genetics 186:713–724

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

Daillant-Spinnler B, MacFie HJH, Beyts PK, Hedderley D (1996) Relationships between perceived sensory properties and major preference directions of 12 varieties of apples from the Southern Hemisphere. Food Qual Pref 7:113–126

De Roos APW, Schrooten C, Mullaart E et al (2009) Genomic selection at CRV. Proceedings of the interbull international workshop—genomic information in genetic evaluations. Uppsala, Sweden, Bulletin no. 39

Dekkers JCM (2004) Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. J Anim Sci 82:E313–E328

Devoghalaere F, Doucen T, Guitton B et al (2012) A genomics approach to understanding the role of auxin in apple (*Malus* x *domestica*) fruit size control. BMC Plant Biol 12. DOI 10.1186/1471–2229-12–7

Dunemann F, Ulrich D, Malysheva-Otto L et al (2012) Functional allelic diversity of the apple alcohol acyl-transferase gene *MdAAT1* associated with fruit ester volatile contents in apple cultivars. Mol Breeding 29:609–625

Durel C-E, Laurens F, Fouillet A, Lespinasse Y (1998) Utilisation of pedigree information to estimate genetic parameters from large, unbalanced data sets in apple. Theor Appl Genet 96:1077–1085

Espley RV, Brendolise C, Chagne D et al (2009) Multiple repeats of a promoter segment causes transcription factor autoregulation in red apples. Plant Cell 21:168–183

Evans K (2012) Apple breeding in the pacific northwest. Acta Hort (in press)

Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics, 4th ed. Longman, Harlow

Fernando R, Grossman M (1989) Marker assisted selection using best linear unbiased prediction. Genet Select E 21:467–477

Fischer C (1994) Shortening of the juvenile period in apple breeding. In: Schmidt H, Kellerhals M (eds) Progress in temperate fruit breeding. Kluwer, Dordrecht, pp 161–164

Fischer M, Fischer C (2004) 75 years of tradition in classical pillnitz fruit breeding—aims, results. Acta Hort 663:699–706

Flachowsky H, Peil A, Sopanen T et al (2007) Overexpression of BpMADS4 from silver birch (*Betula pendula* Roth.) induces early flowering in apple (*Malus x domestica* Borkh.). Plant Breed 126:137–145

Flachowsky H, Hanke M-V, Peil A et al (2009) A review on transgenic approaches to accelerate breeding of woody plants. Plant Breed 128:217–226

Flachowsky H, Le Roux P-M, Peil A et al (2011) Application of a high-speed breeding technology to apple (*Malus x domestica*) based on transgenic early flowering plants and marker-assisted selection. New Phytol 192:364–377

Forsline PL, Aldwinckle HS, Dickson EE et al (2003) Collection, maintenance, characterization, and utilization of wild apples of central Asia. Hortic Rev 29:1–61
Frey JE, Frey B, Sauer C, Kellerhals M (2004) Efficient low-cost DNA extraction and multiplex fluorescent PCR method for marker-assisted selection in breeding. Plant Breed 123:554–557
Gardiner S, Norelli J, de Silva N et al (2012) Candidate gene markers associated with quantitative trait loci for fireblight resistance in *Malus* 'Robusta 5' accessions. BMC Genet 13:25. doi: 10.1186/1471–2156-13–25
Gardiner SE, Bus VGM, Rusholme RL et al (2007) Apple. In: Kole C (ed) Genome mapping and molecular breeding in plants, Vol. 4, fruits and nuts. Springer, Heidelberg, pp 1–62
Gasic K, Gonzalez DO, Thimmapuram J et al (2009) Comparative analysis and functional annotation of a large expressed sequence tag collection of apple. Plant Genome 2:23–38
Goddard ME (2009) Genomic selection: prediction of accuracy and maximisation of long term response. Genetica 136:245–257
Guitton B, Kelner J-J, Velasco R et al (2012) Genetic control of biennial bearing in apple. J Exp Bot 63:131–149
Habier D, Fernando RL, Dekkers JC (2007) The impact of genetic relationship information on genome-assisted breeding values. Genetics 177:2389–2397
Habier D, Tetens J, Seefried F-R et al (2010) The impact of genetic relationship information on genomic breeding values in German Holstein cattle. Genet Sel Evol 42:5
Harker FR, Gunson FA, Jaeger SR (2003) The case for fruit quality: an interpretive review of consumer attitudes and preferences. Postharvest Biol Tec 28:333–347
Harris BL, Johnson DL, Spelman RJ (2008) Genomic selection in New Zealand and the implications for national genetic evaluation. Proceedings of the 36th ICAR Session, Niagara Falls (USA), pp 325–331
Harris SA, Robinson JP, Juniper BE (2002) Genetic clues to the origin of the apple. Trends Genet 18:426–430
Harrison N, Harrison R (2011) On the evolutionary history of the domesticated apple. Nat Genet 43:1043–1044
Hayes BJ, Visscher PM, Goddard ME (2009a) Increased accuracy of artificial selection by using the realized relationship matrix. Genet Res 91:47–60
Hayes BJ, Daetwyler HD, Bowman P et al (2009b) Accuracy of genomic selection: comparing theory and results. Proc Assoc Advmt Ani Breed Genet 18:34–37
Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME (2009c) Invited review: genomic selection in dairy cattle: progress and challenges. J Dairy Sci 92:433–443
Heffner EL, Lorenz AJ, Jannink JL, Sorrells ME (2010) Plant breeding with genomic selection: gain per unit time and cost. Crop Sci 50:1681–1690
Heffner EL, Sorrells ME, Jannink JL (2009) Genomic selection for crop improvement. Crop Sci 49:1–12
Henderson CR (1984) Application of linear models in animal breeding. University of Guelph
Heslot N, Yang H-P, Sorrells ME, Jannink J-L (2012) Genomic selection in plant breeding: a comparison of models. Crop Sci 52:146–160
Iwata H, Hayashi T, Tsumura Y (2011) Prospects for genomic selection in conifer breeding: a simulation study of *Cryptomeria japonica*. Tree Genet Genomes 7:1–12
Jaeger SR, Andani Z, Wakeling IN, MacFie HJH (1998) Consumer preferences for fresh and aged apples: a crosscultural comparison. Food Qual Pref 9:355–366
Janick J, Cummins JN, Brown SK, Hemmat M (1996) Apples. In: Janick J, Moore JN (eds) Fruit breeding. Vol I. Tree and tropical fruits. John Wiley & Sons Inc., New York, pp 1–77
Jannink JL (2010) Dynamics of long-term genomic selection. Genet Sel Evol 42:35
Jannink JL, Lorenz AJ, Iwata H (2010) Genomic selection in plant breeding: from theory to practice. Brief Funct Genomics 9:166–177
Janse J, Verhaegh JJ, den Nijs APM (1994) Early selection of partial resistance to powdery mildew, *Podosphaera leucotricha* (Ell. et EV.) Salm. in apple progenies. Euphytica 77:7–9

Kellerhals M, Bamgartner I, Leumann L et al (2012) Progress in pyramiding disease resistances in apple breeding. Acta Hort (in press)
Khan SA, Chibon P-Y, de Vos RCH et al (2012) Genetic analysis of metabolites in apple fruits indicates an mQTL hotspot for phenolic compounds on linkage group 16. J Expt Bot 8:895–2908
Kizilkaya K, Fernando RL, Garrick DJ (2010) Genomic prediction of simulated multibreed and purebred performance using observed fifty thousand single nucleotide polymorphism genotypes. J Anim Sci 88:544–551.
Korban SS, Tartarini S (2009) Apple structural genomics. In: Folta KM, Gardiner SE (eds) Genetics and genomics of the rosaceae. Springer, New York, pp 85–119
Kouassi AB, Durel C-E, Costa F et al (2009) Estimation of genetic parameters and prediction of breeding values for apple fruit-quality traits using pedigreed plant material in Europe. Tree Genet Genomes 5:659–672
Kumar S, Volz RK, Alspach PA, Bus VGM (2010) Development of a recurrent apple breeding programme in New Zealand: a synthesis of results, and a proposed revised breeding strategy. Euphytica 173:207–222
Kumar S, Volz RK, Weskett R (2011) Genetic architecture of fruit quality traits in *Malus* x *domestica* (Borkh.) compared between own-rooted seedlings and vegetative propagules on 'M. 9' rootstock. Tree Genet Genomes 7:1079–1088
Kumar S, Bink MCAM, Volz RK et al (2012a) Towards genomic selection in apple (*Malus* × *domestica* Borkh.) breeding programs: prospects, challenges and strategies. Tree Genet Genomes 8:1–14
Kumar S, Chagné D, Bink MCAM et al (2012b) Genomic selection for fruit quality traits in apple (*Malus* × *domestica* Borkh.). PLoS ONE 7(5):e36674
Labuschagné IF (2004) Budbreak number as selection criteria for breeding apples adapted to mild winter climatic conditions: a review. Acta Hort 663:775–781
Laurens F (1999) Review of the current apple breeding programs in the world: objectives for scion cultivar improvement. Acta Hort 484:163–170
Le Roux P-M, Flachowsky H, Hanke M-V et al (2012) Development of apple pre-breeding genotypes highly resistant to fireblight by early flowering In: BiotechFruit 2012, Nelson, New Zealand, 25th-29th March, pp 133. www.biotechfruit.com
Legarra A, Robert-Granie C, Manfredi E, Elsen J-M (2008) Performance of genomic selection in mice. Genetics 180:611–618
Lespinasse Y (2009) Review of pome fruit breeding in Europe: which strategies for the near future? Acta Hort 824:865–871
Lespinasse Y, Alston FH, Watkins R (1976) Cytological techniques for use in apple breeding. Ann Appl Biol 82:349–353
Liebhard R, Kellerhals M, Pfammatter W et al (2003) Mapping quantitative physiological traits in apple (*Malus* × *domestica* Borkh.). Plant Mol Biol 52:511–526
Longhi S, Moretto M, Viola R et al (2012) Comprehensive QTL mapping survey dissects the complex fruit texture physiology in apple (*Malus* × *domestica* Borkh.). J Expt Bot 63:1107–1121
Lorenz AJ, Hamblin MT, Jannink J-L (2010) Performance of single nucleotide polymorphisms versus haplotypes for genome-wide association analysis in barley. PLoS ONE 5: e14079.
Lorenz AJ, Chao S, Asoro FG et al (2011) Genomic selection in plant breeding: knowledge and prospects. Adv Agron 110:77–123
Lorenzana RE, Bernardo R (2009) Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. Theor Appl Genet 120:151–161
Luby JJ, Shaw DV (2001) Does marker-assisted selection make dollars and sense in a fruit breeding program? HortScience 36:872–879
Lund M, Su G (2009) Genomic selection in the Nordic countries. Proceedings of the interbull international workshop—genomic information in genetic evaluations. Uppsala, Sweden, Bulletin no. 39
Meuwissen THE (2009) Accuracy of breeding values of 'unrelated' individuals predicted by dense SNP genotyping. Genet Sel Evol 41:35–43

Meuwissen THE, Goddard ME (1996) The use of marker haplotypes in animal breeding schemes. Genet Sel Evol 28:161–176
Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829
Micheletti D, Troggio M, Zharkikh A et al (2011) Genetic diversity of the genus *Malus* and implications for linkage mapping with SNPs. Tree Genet Genomes 7:11
Montes JM, Melchinger AE, Reif JC (2007) Novel throughput phenotyping platforms in plant genetic studies. Trends Plant Sci 12:433–436
Morgan J, Richards A (1993) The book of apples. Ebury Press, London
Moser G, Tier B, Crump R et al (2009) A comparison of five methods to predict genomic breeding values of dairy bulls from genome-wide SNP markers. Genet Sel Evol 41:56
Newcomb RD, Crowhurst RN, Gleave AP et al (2006) Analyses of expressed sequence tags from apple. Plant Physiol 141:147–166
Noiton DAM, Alspach PA (1996) Founding clones, inbreeding, coancestry, and status number of modern apple cultivars. J Am Soc Hortic Sci 121:773–782.
O'Rourke D (2011) World apple review. Belrose Inc, Pullman, Washington
Oraguzie NC, Iwanami H, Soejima J et al (2004) Inheritance of the *Md-ACS1* gene and its relationship to fruit softening in apple (*Malus* × *domestica* Borkh.). Theor Appl Genet 108:1526–1533
Patocchi A, Frei A, Frey JE, Kellerhals M (2009) Towards improvement of marker assisted selection of apple scab resistant cultivars: *Venturia inaequalis* virulence surveys and standardization of molecular marker alleles associated with resistance genes. Mol Breeding 24:337–347
Pindo M, Vezzulli S, Coppola G et al (2008) SNP high-throughput screening in grapevine using the SNPlexTM genotyping system. BMC Plant Biol 8:12
Rehder A (1940) Manual of cultivated trees and shrubs. MacMillan, New York, NY, pp 389–399
Resende MDV, Resende MFRJ, Sansaloni CP et al (2012c) Genomic selection for growth and wood quality in *Eucalyptus*: capturing the missing heritability and accelerating breeding for complex traits in forest trees. New Phytol 194:116–128
Resende MFRJ, Muñoz P, Resende MDV et al (2012a) Accuracy of genomic selection methods in a standard dataset of loblolly pine (*Pinus taeda* L.). Genetics 190:1503–1510
Resende MFRJ, Muñoz P, Acosta JJ et al (2012b) Accelerating the domestication of trees using genomic selection: accuracy of prediction models across ages and environments. New Phytol 193:617–624
Robinson JP, Harris SA, Juniper BE (2001) Taxonomy of the genus *Malus* Mill. (Rosaceae) with emphasis on the cultivated apple, *Malus x domestica* Borkh. Plant Syst Evol 226:35–58
Rothschild MF, Soller M (1997) Candidate gene analysis to detect genes controlling traits of economic importance in domestic livestock. Probe 8:13–20
Schenkel FS, Sargolzaei M, Kistemaker G et al (2009) Reliability of genomic evaluation of Holstein cattle in Canada. Proceedings of the interbull international workshop—genomic information in genetic evaluations. Uppsala, Sweden, Bulletin no. 39
Silfverberg-Dilworth E, Matasci CL, Van de Weg WE et al (2006) Microsatellite markers spanning the apple (*Malus* x *domestica* Borkh.) genome. Tree Genet Genomes 2:202–224
Sved JA (1971) Linkage disequilibrium and homozygosity of chromosome segments in finite populations. Theor Popul Biol 2:125–141
Toosi A, Fernando RL, Dekkers JCM (2009) Genomic selection in admixed and crossbred populations. J Anim Sci 88:32–46
VanRaden PM, Van Tassell CP, Wiggans GR et al (2009) Invited review: reliability of genomic predictions for North American Holstein bulls. J Dairy Sci 92:16–24
Velasco R, Zharkikh A, Affourtit J et al (2010) The genome of the domesticated apple (*Malus* x *domestica* Borkh.). Nat Genet 42:833–839
Volz RK, Alspach PA, White AG, Ferguson IB (2001) Genetic variability in apple fruit storage disorders. Acta Hort 553:241–244

Volz RK, Rikkerink E, Austin P et al (2009) "Fast Breeding" in apple: a strategy to accelerate introgression of new traits into elite germplasm. Acta Hort 814:163–168

Wiedow C, Chagné D, Souleyre E et al (2010) QTL and candidate gene mapping of 'ripe apple' aroma. In: Plant & Animal Genome XVIII Conference, San Diego, CA, January 9–13

Wong C, Bernardo R (2008) Genomewide selection in oil palm: increasing selection gain per unit time and cost with small populations. Theor Appl Genet 116:815–824

Xu S (2007) An empirical Bayes method for estimating epistatic effects of quantitative trait loci. Biometrics 63:513–521.

Yang JA, Benyamin B, McEvoy BP et al (2010) Common SNPs explain a large proportion of the heritability for human height. Nat Genet 42:565–131

Yi N, Xu S (2008) Bayesian LASSO for quantitative trait loci mapping. Genetics 179:1045–1055.

Zhong S, Dekkers JCM, Fernando RL, Jannink J-L (2009) Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a barley case study. Genetics 182:355–364

Zhu Y, Barritt BH (2008) *Md-ACS1* and *Md-ACO1* genotyping of apple (*Malus* x *domestica* Borkh.) breeding parents and suitability for marker-assisted selection. Tree Genet Genomes 4:555–562

Zhu Y, Evans K, Peace C (2011) Utility testing of an apple skin color MdMYB1 marker in two progenies. Mol Breeding 27:525–532

Zimmerman RH (1991) Flowering in crabapple seedlings: methods of shortening the juvenile phase. J Am Soc Hort Sci 96:404–411

Chapter 17
Enhancing Nutritional Quality in Crops Via Genomics Approaches

Meike S. Andersson, Wolfgang H. Pfeiffer and Joe Tohme

Abstract Micronutrient malnutrition—also known as hidden hunger—is a growing public health concern that affects especially women and children in the developing world. Worldwide, at least 2 billion people suffer from vitamin A, iron, and zinc deficiencies. Here we review recent advances in the application of genomic approaches for biofortification of staple crops to enhance their nutritional quality and thus reduce 'hidden hunger'. The application of genomic tools such as marker-assisted selection in conventional breeding or genetic modification offers sustainable and cost-effective ways to provide essential micronutrients (here provitamin A or iron) to people in developing countries. To maximize the benefits of genomic approaches for biofortification, we need to extend our understanding of the genetic control mechanisms and relative contribution from different rate-limiting steps for both provitamin A and iron accumulation in edible plant parts.

In the past decades, modern agriculture has had reasonable success in increasing productivity and meeting the energy needs of poor populations in developing countries. However, micronutrient malnutrition—also known as hidden hunger—has been increasing. Worldwide, at least 2 billion people, mostly women and children, suffer from micronutrient malnutrition, caused by a lack of critical micronutrients such as vitamin A, iron (Fe), and zinc (Zn) in the diet (Kennedy et al. 2003; Micronutrient Initiative 2009). In particular, zinc and iron deficiencies are a growing public health concern, especially in the developing world (Fig. 17.1). Hidden hunger leads to impaired mental and physical development of children and adolescents, and can result in lower IQ, stunting, and blindness, as well as increased risk of diseases and mortality (WHO and FAO 2006; Walker et al. 2007; Black et al. 2008). Therefore, agriculture must focus not only on producing more food, but on delivering more nutritious food as well.

J. Tohme (✉)
CIAT, 6713 Cali, Colombia
e-mail: j.tohme@cgiar.org

M. S. Andersson · W. H. Pfeiffer
HarvestPlus, c/o CIAT, 6713 Cali, Colombia

R. Tuberosa et al. (eds.), *Genomics of Plant Genetic Resources,*
DOI 10.1007/978-94-007-7575-6_17,

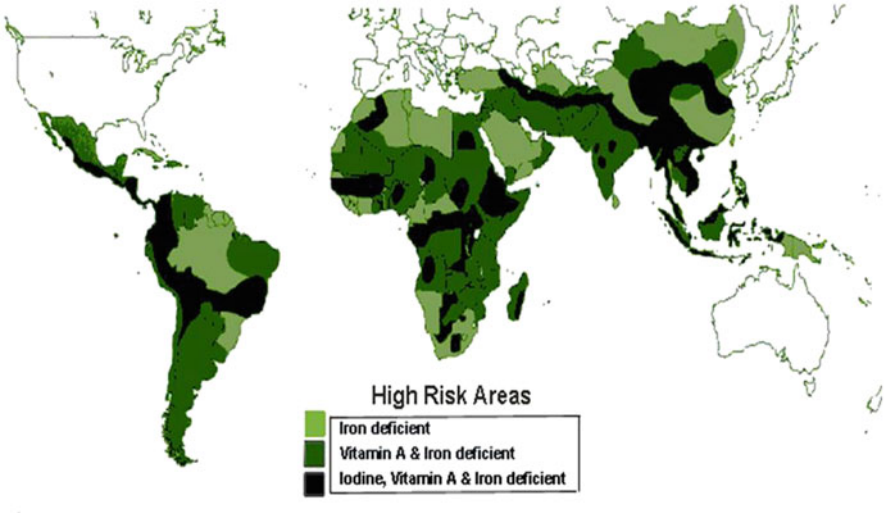

Fig. 17.1 Global map showing high risk areas of micronutrient deficiencies. (Source: USAID)

Biofortification is a new approach that bridges the gap between agriculture and health, and complements other interventions such as dietary diversification, fortification, and supplementation programs (Gómez-Galera et al. 2010; Zhao and Shewry 2011). Biofortification involves improving the nutritional content of food staples eaten widely by the poor, using the best traditional breeding practices and modern biotechnology approaches. The new, biofortified varieties are not only rich in micronutrients, but also agronomically competitive with commercial, non-biofortified varieties in terms of productivity, pest and disease resistance, and marketability traits such as color, size, or taste.

Biofortification is a comparatively inexpensive, cost-effective and sustainable means for combating micronutrient deficiencies among the poor in developing countries (Meenakshi et al. 2010; Bouis et al. 2011). HarvestPlus, an interdisciplinary alliance of research institutions and implementing agencies in more than 40 countries (www.harvestplus.org), leads a global effort to develop and deliver biofortified varieties of major staple food crops with high concentrations of one or more of the three most limiting nutrients: vitamin A, zinc, and iron (Table 17.1). Activities are classified into three phases—discovery, development, and dissemination—that follow a pathway of impact (Fig. 17.2).

Genomic approaches are an integral part of HarvestPlus breeding activities, in particular during crop improvement. Using novel genomic, genetic, and molecular techniques to study the biochemical pathways in plants allows us to identify and better understand the basic mechanisms underlying the synthesis and accumulation of essential vitamins and minerals—specifically zinc, iron, and vitamin A (DellaPenna 1999). This knowledge can then be used to make the breeding process more efficient through (a) marker-assisted selection for conventional breeding or (b) the development of transgenic lines with increased micronutrient levels.

Table 17.1 Schedule of HarvestPlus product releases. (Bouis et al. 2011)

Crop	Nutrient	Countries of first release	Agronomic trait	Release year[a]
Sweet potato	Provitamin A	Uganda, Mozambique	Disease resistance, drought tolerance, acid soil tolerance	2007
Bean	Iron, zinc	Rwanda, Democratic Republic of Congo	Virus resistance, heat and drought tolerance	2012
Pearl millet	Iron, zinc	India	Mildew resistance, drought tolerance	2012
Cassava	Provitamin A	Nigeria, Democratic Republic of Congo	Disease resistance	2011
Maize	Provitamin A	Zambia	Disease resistance, drought tolerance	2012
Rice	Zinc, iron	Bangladesh, India	Disease and pest resistance, cold and submergence tolerance	2013
Wheat	Zinc, iron	India, Pakistan	Disease and lodging resistance	2013

[a]Approved for release by national governments after intensive multilocation testing for agronomic and micronutrient performance.

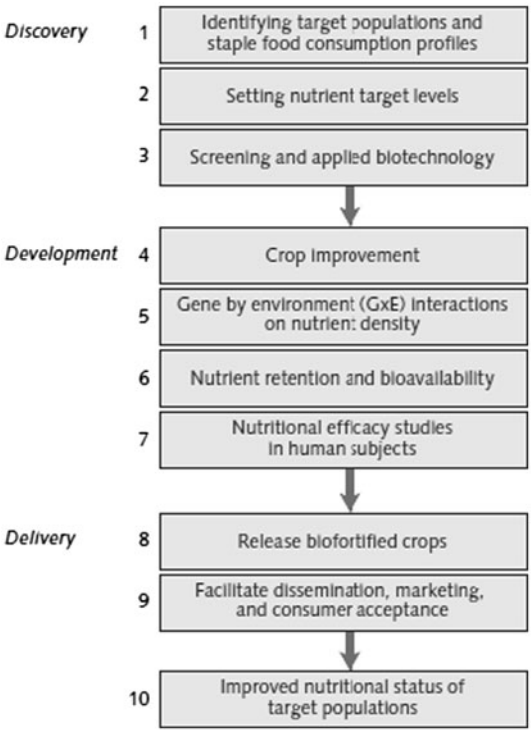

Fig. 17.2 Biofortification activities along the HarvestPlus Impact Pathway. (Bouis et al. 2011)

In the following, we present two contrasting examples of the use of applied genomics for biofortification: (1) marker-assisted selection for high-provitamin A cassava and maize, and (2) transgenic approaches for high-iron rice.

17.1 Molecular Markers to Enhance the Breeding Efficiency for High Provitamin A Cassava and Maize

Carotenoids are produced by the plant itself in the carotenoid biosynthetic pathway and their biosynthesis is well understood (Cuttriss et al. 2011) with control by relatively few genes, and the trait heritability is high (Egesel et al. 2003; Menkir and Maziya-Dixon 2004; Grüneberg et al. 2005; Morillo Coronado 2009). Provitamin A accumulation in crops can be enhanced by favorably altering the biochemical pathway and maximizing pathway flux for carotenoid synthesis.

Cassava (*Manihot esculenta* Crantz) plays an important role in the tropics where it is the main source of calories and income for about 800 million of the world's poorest people. Even though the starchy roots of cassava are a valuable source of low-cost energy, their nutritional value is limited as they are poor in protein and essential micronutrients such as iron, zinc, or provitamin A (Ceballos et al. 2007). The roots of nearly all commercial cassava cultivars and breeding populations are white, however natural genetic variation exists in the crop and some yellow-rooted cultivars and landraces have been identified that produce provitamin A carotenoids (Chávez et al. 2005; Esuma et al. 2012).

Through conventional breeding methods based on mass-recurrent phenotypic selection, scientists were able to increase the provitamin A levels in cassava several fold (Ceballos et al. 2012a, b). As the result of a collaborative breeding effort between national and international research centers coordinated by HarvestPlus, cassava varieties with yellow-fleshed roots have been officially released in Brazil in 2005 and 2009, in DR Congo in 2008, and in Nigeria in 2011. However, long breeding cycles and the complex genetic nature of the crop—cassava is highly heterozygous—render varietal recovery extremely difficult. The breeding process could be considerably accelerated and would become much more efficient if molecular markers could be identified. Access to molecular markers would allow the selection of favorable alleles more efficiently and inexpensively, thus enabling developing-country breeders to produce cassava roots with higher provitamin A levels in locally adapted varieties with high consumer preference.

Initially, molecular marker development in cassava has relied upon simple sequence repeats (SSRs), including microsatellite sequences (Roa et al. 2000; Mba et al. 2001; Raji et al. 2009) and single-nucleotide polymorphisms (Lopez et al. 2005; Tangphatsornruang et al. 2008; Ferguson et al. 2012). However, these SSR and SNP markers were sparsely distributed across the cassava genome and insufficient for fine-mapping and inexpensive, large-scale assays. Recent advances in molecular marker screening through *"next generation sequencing"* (NGS) are more promising, and with the increasing availability of genomic resources in cassava research—including

the first full cassava genome sequence (Prochnik et al. 2012)—the application of NGS in molecular breeding becomes a reality. NGS makes high-throughput genotyping for orphan crops like cassava possible, and test assays for the latter are already underway.

For maize (*Zea mays* L.), the use of marker-assisted breeding for enhancing provitamin A levels has already become a reality. This was possible after association mapping research that revealed polymorphisms in two crucial genes further down the CBP, namely lycopene epsilon-cyclase (*LycE*; Harjes et al. 2008) and carotene beta-hydroxylase 1 (*CrtR-B1*; Yan et al. 2010). PCR-based markers were subsequently developed to enable selection for favorable alleles of these two genes, thus boosting provitamin A levels in new orange maize generations and accelerating breeding by one season (Dwiwedi et al. 2012; Zhang et al. 2012). A large-scale allele mining effort is now underway at CIMMYT, and has already yielded a number of promising lines with favorable allele constitution at either or both loci in diverse tropical/sub-tropical germplasm backgrounds (Pixley et al. 2011; Menkir et al. 2012).

Cassava is not far away from making similar progress and breakthroughs in germplasm screening. A first step in this direction is the discovery by Welsch et al. (2010), that the underlying causes of higher provitamin A levels in yellow-fleshed cassava cultivars could be traced back to a single amino acid change in phytoene synthase (*psy*); an enzyme that catalyzes the initial steps of the Carotenoid Biosynthesis Pathway (CBP). Welsch and his team were able to show that a single polymorphism in this enzyme leads to markedly increased carotenoid formation and accumulation in cassava storage roots. These findings provide means to increase the levels of provitamin A carotenoids in cassava roots through both breeding (using marker-assisted selection) or genetic modification (employing root-specific promoters).

Recently, CIAT's cassava program has developed a high provitamin A/ high dry-matter content segregating population (LA Becerra, personal communication 2012). This population has shown one of the highest β-carotene levels ever recorded in cassava (25–30 ppm, fresh root weight) and is linked with exceptionally high dry-matter content (approx. 42 %) prerequisite for a high value cassava root. The molecular characterization of this segregating population by high-density SNP-based fine mapping and five genes of the cassava CBP, coupled with high-resolution biochemical phenotyping, is currently underway. This approach is another important step towards unraveling the genetic control mechanisms for provitamin A accumulation in cassava roots.

17.2 Transgenic Approaches to Enhance Iron Levels in Rice

Rice (*Oryza sativa* L.) is the primary staple food for more than half of the world's population, but it is also a poor source of essential nutrients such as Fe and Zn (Mayer et al. 2008; Bhullar and Gruissem 2012). Iron concentrations in the polished rice grain are very low (3 ppm) compared to the 14.5 ppm Fe required to meet the daily Fe needs

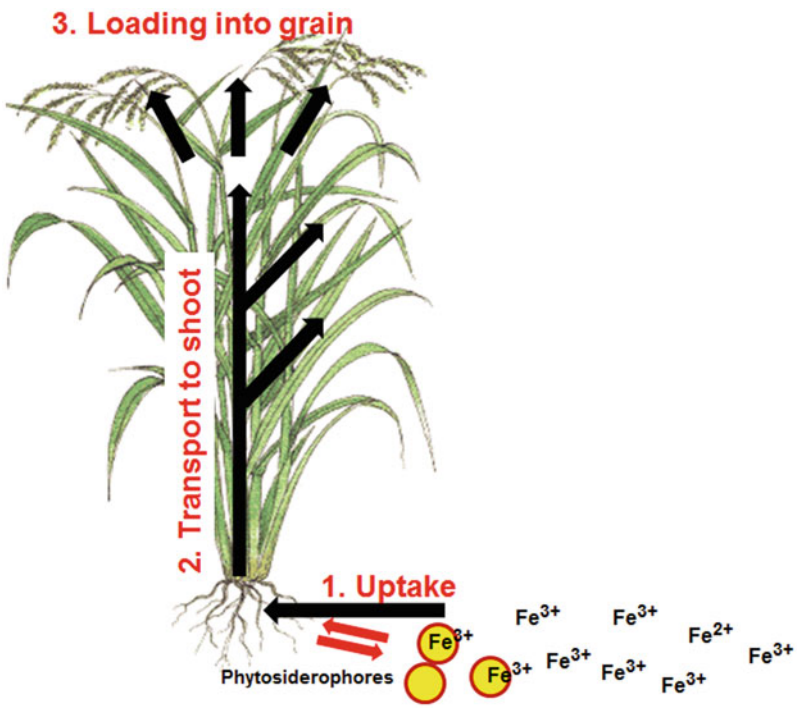

Fig. 17.3 Fe pathway from soil to seed. (Johnson 2010)

with a rice-based diet (Lucca et al. 2002; Hotz and Clafferty 2007; Bouis and Welch 2010). Genetic variation for Fe is low in the rice gene pool, and so far conventional breeding has failed to increase Fe levels beyond 8 ppm in polished grain (Pfeiffer and McClafferty 2007; Khush et al. 2012). Strategies for increasing Fe levels in rice therefore need to explore options for manipulating the genetic control mechanisms that regulate Fe uptake, translocation, and deposition in the grain (Fig. 17.3).

While provitamins are synthesized by the plant itself and have relatively high heritability, essential minerals such as Fe and Zn are taken up by plants from the soil, transported to stems and leaves, and subsequently translocated to seeds or other edible parts (Welch 1995; Mori 1999; Schurr 1999). The uptake and translocation of Fe from the roots is controlled by complex physiological and homeostatic mechanisms that regulate the absorption, transport, and allocation to seed and non-seed parts (Grusak et al. 1999; Hell and Stephan 2003; Borg et al. 2009; Cakmak et al. 2010). The mechanisms involved in Fe homeostasis are controlled by several relevant genes with intermediate heritability (Gross et al. 2003; Long et al. 2004; Cichy et al. 2005; Briat et al. 2007; Pfeiffer and McClafferty 2007; Courtney et al. 2008).

Staple cereals such as rice, wheat and maize—and other graminaceous plants—acquire Fe(III) by releasing phytosiderophores from their roots (Strategy II) to solubilize and take up soil Fe (Takagi et al. 1984; Römheld and Marschner 1986).

Rice is exceptionally Fe-efficient because in addition it possesses a direct Fe(II)-transport system (Strategy I) which is usually employed by non-graminaceous species only (Marschner et al. 1986; Ishimaru et al. 2006).

Transgenic approaches have been applied to increase Fe (and Zn) levels in rice by:

1. *Increasing Fe uptake*—via modifying the Strategy I or II mechanisms,
2. *Increasing Fe transport*—via increasing nicotianamine production or content, and
3. *Increasing Fe loading to the grain*—via increasing ferritin protein levels.

1. *Increasing Fe uptake*

The uptake of Fe can be increased by modifying either of the two strategies employed by plants to mobilize Fe from the soil (Curie and Briat 2003; Briat et al. 2007).

Takahashi et al. (2001) managed to increase phytosiderophore secretion (Strategy II) in rice by introducing nicotianamine aminotransferase (NAAT) genes from barley. Phytosiderophores are iron chelators synthesized by graminaceous plants and released into the soil to solubilize iron (Fe(III) chelation; Strategy II). Nicotianamine (NA) is a precursor of phytosiderophores and a plant metal chelator. NA is found in all higher plants and plays a critical role in metal homeostasis because it is (a) an intermediate for the biosynthesis of phytosiderophores (Higuchi et al. 2009), and (b) facilitates the transport of Fe and other metals inside plants. The levels of NA in plants are controlled by NAAT and NA synthase (Mori 1999). Takahashi et al. (2001) found that transgenic rice expressing two NAAT genes from barley secreted larger amounts of phytosiderophores than non-transformants, and that Fe levels in the plant tissue were significantly increased.

Ishimaru et al. (2007) achieved a similar effect by enhancing the reduction of Fe(III) to Fe(II) for direct uptake (Strategy I). They introduced a yeast Fe(III) chelate-reductase gene (*refre1/372*) to rice plants. Transgenic rice expressing the *refre1/372* gene showed higher Fe(III) chelate-reductase activity and a higher Fe-uptake rate than vector controls under Fe-deficient conditions.

However, in neither of the two cases could an increase in seed Fe content be observed, indicating that Fe uptake from the soil is not a rate-limiting step for increasing seed Fe content in rice (Guerinot 2007; Sperotto et al. 2012).

2. *Increasing Fe transport*

Different groups have studied approaches to increase Fe transport and translocation in rice via increasing nicotianamine (NA) production or content. NA can form stable complexes with metals in plants and is important for the translocation of Fe and other metals from the roots to the aerial parts of the plant (Pich et al. 1994; von Wiren et al. 1999; Douchkov et al. 2005). More specifically, it facilitates phloem transport and cytoplasmic distribution of metals through transporters encoded by *YSL* (=Yellow Stripe Like) genes (Koike et al. 2004; Curie et al. 2009; Morrissey and Guerinot 2009).

Cheng et al. (2007) observed that the disruption of the NAAT enzyme—which uses NA as substrate—in mutant rice lines results in NA accumulation and subsequent

increased uptake and translocation of Fe. Iron concentrations were 2.7-fold higher in shoots and 3.8-fold higher in polished grains (4.6 ppm) of the mutant than in the wild-type counterpart (1.2 ppm grain Fe), with no obvious effect on agronomic traits. The authors speculate that the accumulation of NA in the mutant rice may mimic Fe starvation. This in turn may have led to signals stimulating the increased expression of genes that encode the up-regulation of Fe(II) (*NAS*) and NA-metal transporters (*OsYSL2*).

Similarly, the overexpression of NA synthase genes in transgenic rice plants results in increased tolerance to Fe and Zn deficiencies and more Fe in both leaves and seeds. Overexpressing rice *OsNAS* genes, different research groups obtained significantly more Fe in brown (up to 2.9-fold) and white (up to 4.0-fold) transgenic rice than in the wild-type (Lee et al. 2009; Zheng et al. 2010; Johnson et al. 2011). Introduction of a barley NA synthase gene (*HvNAS1*) to rice resulted in up to 4.0-fold greater amounts of NA (Higuchi et al. 2001; Usuda et al. 2009) and more than 3-fold higher Fe concentrations in polished seeds of the transgenic lines (Masuda et al. 2009).

Fe (and Zn) transport to seed via NA is thus a rate-limiting step. Modifying the production of NA has proven to be an effective approach for increasing Fe levels in brown and white rice, by enhancing the translocation of Fe from vegetative tissue to the seed. For the first time, rice lines (*OsNAS*) have been reported with Fe concentrations that reach the nutrition target of 14.5 ppm Fe for effective rice biofortification (Johnson et al. 2011).

3. *Increasing Fe loading to the grain*

The loading of Fe to the grain is controlled by a combination of Fe supply and sink strength (Grusak 1994, 2002). Ferritins are Fe storage proteins that take up Fe, store it in a nontoxic form, and release it when needed for metabolic functions (Theil 2004). They are known to be involved in the remobilization of Fe from leaves to stems and then to seeds (Theil 1987). In rice, only 4 % of shoot Fe is transported to the seeds (Marr et al. 1995). Hence, increasing the levels of storage proteins by overexpressing ferritin specifically in seeds is another approach that has been used to enhance Fe content in rice grains.

Plant ferritin genes have been identified in common bean, cowpea, lentil, maize, pea, and soybean (Nirupa and Prasad 2008). When overexpressing soybean or common bean ferritins in transgenic rice, increased levels of grain Fe were observed in both brown and white rice seeds (Goto et al. 1999; Lucca et al. 2001; Vasconcelos et al. 2003; Qu et al. 2005; Khalekuzzaman et al. 2006). However, the grain Fe content in these transgenic rice lines increased only 2- to 3-fold, whereas ferritin protein levels increased 13-fold. This suggests that sink strength is indeed a rate-limiting step, but only until Fe transport to the grain (i.e. Fe supply) becomes limiting.

Wirth et al. (2009) addressed both sink strength and Fe transport as limiting factors, and obtained more than 6-fold greater Fe levels (and no yield penalty) in the seed endosperm of transgenic rice plants expressing an Arabidopsis NAS gene (*AtNAS*) and common bean ferritin genes (*Pvferritin*).

Several quantitative trait loci (QTL) for grain Fe in rice have been identified and mapped, but the fine-mapping of these loci is still to be done (reviewed in Sperotto

et al. 2012). Furthermore, none of the individual QTL accounts for more the 30 % of the Fe variation. Sperotto et al. (2010) have identified a putative "list" of target genes in rice to manipulate Fe and Zn concentrations in the grain. Candidates belong to the YSL, NRAMP (natural resistance-associated macrophage protein), NAS, NAC, and FRO (ferric reductase oxidase) gene families, and their expression levels in flag leaves were significantly correlated with Fe and/or Zn concentrations in the seeds. Recently, *in-silico* studies by Chandel et al. (2011) and Anuradha et al. (2012) have revealed a promising way to identify candidate genes involved in uptake, transport and accumulation of micronutrients in rice. Different combinations of these genes could be expressed to study their effect on Fe levels in the grains of rice or other model crops. Furthermore, new generation candidate gene based markers should be used for the fine mapping of QTL to establish effective marker-assisted selection (MAS) programs for grain nutritive traits.

In summary, Fe uptake from the soil is not a limiting factor for loading of Fe into the rice grain, but Fe transport throughout the plant is. Sink strength is moderately limiting and may be best addressed in conjunction with transport limitations. Strategies for future development of high-Fe (and high-Zn) rice should therefore focus on NAS in combination with ferritin or other genes involved in Fe transport and loading to the grain.

17.3 Conclusions

Using biofortification, staple foods can be bred with provitamin A, Fe, and Zn levels that have a measurable effect on the nutritional status of the poor in developing countries, thus reducing micronutrient deficiencies in a cost-effective and sustainable way.

Biofortification has several advantages as compared to other interventions: (a) it targets the poor whose diets consist of high amounts of staple foods, (b) it is rural-based and hence complements fortification and supplementation, (c) it is cost-effective because research can be multiplied across countries and time, and (d) it is sustainable once it becomes part of mainstream breeding.

Biotechnology and applied genomics provide tools for biofortification breeding and transgenic approaches, that can enable the more efficient and rapid development of nutritionally improved food crops. In particular, the lack of natural genetic variation in some crop/micronutrient combinations can be addressed using biotechnology (e.g. Fe content in rice). Furthermore, biotechnological strategies that work for rice are very likely transferrable to other cereals such as wheat or maize.

References

Anuradha K, Agarwal SY, Rao YV et al (2012) Mapping QTLs and candidate genes for iron and zinc concentrations in unpolished rice of Madhukar × Swarna RILs. Gene 508:233–240

Bhullar NK, Gruissem W (2012) Nutritional enhancement of rice for human health: the contribution of biotechnology. Biotechnol Adv. doi.org/10.1016/j.biotechadv.2012.02.001

Black RE, Allen LH, Bhutta ZA et al, for the Maternal and Child Undernutrition Study Group (2008) Maternal and child undernutrition: global and regional exposures and health consequences. Lancet 371:243–260

Borg S, Brinch-Pedersen H, Tauris B, Holm PB (2009) Iron transport, deposition and bioavailability in the wheat and barley grain. Plant Soil 325:15–24

Bouis HE, Hotz C, McClafferty B et al (2011) Biofortification: a new tool to reduce micronutrient malnutrition. Food Nutr Bull 32(Suppl 1):31S–40S

Bouis HE, Welch RM (2010) Biofortification—a sustainable agricultural strategy for reducing micronutrient malnutrition in the global south. Crop Sci 50:S1–S13

Briat JF, Curie C, Gaymard F (2007) Iron utilization and metabolism in plants. Curr Opin Plant Biol 10:276–282

Cakmak I, Pfeiffer WH, McClafferty B (2010) Review: biofortification of durum wheat with zinc and iron. Cereal Chem 87:10–20

Ceballos H, Fregene M, Pérez JC et al (2007) Cassava genetic improvement. In: Kang MS, Priyadarshan PM (eds) Breeding major food staples. Blackwell, Ames, pp 365–391

Ceballos H, Hershey C, Becerra López-LavalleLA (2012a) New approaches to cassava breeding. In: Janick J (ed) Plant Breeding Reviews, vol 36, chapter 6. Wiley, New York, pp 427–504

Ceballos H, Morante N, Sanchez T et al (2012b) Progress increasing carotenoids content in cassava roots through fast recurrent selection. In: Global Cassava Partnership Second Scientific Conference GCP21-II. Cassava: Overcoming challenges of global climatic change. National Crops Resources Research Institute. Namulonge, Uganda. June 18–22, 2012

Chandel G, Samuel P, Dubey M, Meena R (2011) In silico expression analysis of QTL specific candidate genes for grain micronutrient (Fe/Zn) content using ESTs and MPSS signature analysis in rice (*Oryza sativa* L.). J Plant Genet Transgenics 2:11–22

Chávez AL, Sánchez T, Jaramillo G et al (2005) Variation of quality traits in cassava roots evaluated in landraces and improved clones. Euphytica 143:125–133

Cheng L, Wang F, Shou H et al (2007) Mutation in nicotianamine aminotransferase stimulated the Fe(II) acquisition system and led to iron accumulation in rice. Plant Physiol 145:1647–1657

Cichy KA, Forster S, Graft KF, Hosfield GL (2005) Inheritance of seed zinc accumulation in navy bean. Crop Sci 45:864–870

Courtney M, McHaro M, Bonte D, Grüneberg W (2008) Heritability estimates for micronutrient composition of sweetpotato storage roots. HortScience 43:1382–1384

Curie C, Briat JF (2003) Iron transport and signaling in plants. Annu Rev Plant Biol 54:183–206

Curie C, Cassin G, Couch D et al (2009) Metal movement within the plant: contribution of nicotianamine and yellow stripe 1-like transporters. Ann Bot 103:1–11

Cuttriss AJ, Cazzonelli CI, Wurtzel ET, Pogson BJ (2011) Carotenoids. In: Rébeillé F, Douce R (eds) Adv Bot Res 58:1–36

DellaPenna D (1999) Nutritional genomics: manipulating plant micronutrients to improve human health. Science 285:375–379

Douchkov D, Gryczka C, Stephan UW et al (2005) Ectopic expression of nicotianamine synthase genes results in improved iron accumulation and increased nickel tolerance in transgenic tobacco. Plant Cell Environ 28:365–374

Egesel CO, Wong JC, Lambert RJ, Rocheford TR (2003) Combining ability of maize inbreds for carotenoids and tocopherols. Crop Sci 43:818–823

Esuma W, Rubaihayo P, Pariyo A et al (2012) Genetic diversity of provitamin A cassava in Uganda. J Plant Stud 1:60–71

Ferguson ME, Hearne SJ, Close TJ et al (2012) Identification, validation and high-throughput genotyping of transcribed gene SNPs in cassava. Theor Appl Genet 124:685–695

Gómez-Galera S, Rojas E, Sudhakar D et al (2010) Critical evaluation of strategies for mineral fortification of staple food crops. Transgenic Res 19:165–180

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

Gross J, Stein RJ, Fett-Neto AG, Fett JP (2003) Iron homeostasis related genes in rice. Genetics Mol Biol 26:477–497

Grüneberg WJ, Manrique K, Zhang D, Hermann M (2005) Genotype × environment interactions for a diverse set of sweetpotato clones evaluated across varying ecogeographic conditions in Peru. Crop Sci 45:2160–2171

Grusak MA (1994) Iron transport to developing ovules of *Pisum sativum*. I. Seed import characteristics and phloem iron-loading capacity of source regions. Plant Physiol 104:649–655

Grusak MA (2002) Enhancing mineral content and bioavailability in plant food products. J Am Coll Nutr 21:178S–183S

Grusak MA, Pearson JN, Marentes E (1999) The physiology of micronutrient homeostasis in field crops. Field Crops Res 60:41–56

Guerinot ML (2007) It's elementary: enhancing Fe3 + reduction improves rice yields. Proc Natl Acad Sci USA 104:7311–7312

Harjes CE, Rocheford TR, Bai L et al (2008) Natural genetic variation in lycopene epsilon cyclase tapped for maize biofortification. Science 319:330–333

Hell R, Stephan UW (2003) Iron uptake, trafficking and homeostasis in plants. Planta 216:541–551

Higuchi K, Takahashi M, Nakanishi H et al (2001) Analysis of transgenic rice containing barley nicotianamine synthase gene. Soil SciPlant Nutr 47:315–322

Higuchi K, Suzuki K, Nakanishi H et al (2009) Cloning of nicotianamine synthase genes, novel genes involved in the biosynthesis of phytosiderophores. Plant Physiol 119:471–479

Hotz C, McClafferty B (2007) From harvest to health: challenges for developing biofortified staple foods and determining their impact on micronutrient status. Food Nutr Bull 28:S271–S279

Ishimaru Y, Suzuki M, Tsukamoto T et al (2006) Rice plants take up iron as an Fe^{3+} -phytosiderophore and as Fe^{2+}. Plant J 45:335–346

Ishimaru Y, Kim S, Tsukamoto T et al (2007) Mutational reconstructed ferric chelate reductase confers enhanced tolerance in rice to iron deficiency in calcareous soil. Proc Natl Acad Sci US A 104:7373–7378

Johnson AAT (2010) Strategies for increasing micronutrient mineral levels in cereal crops. Online Proceedings First Global Conference on Biofortification, November 9–11, 2010, Washington, DC. http://biofortconf.ifpri.info/conference-agenda/symposia-november-10/

Johnson AAT, Kyriacou B, Callahan DL et al (2011) Constitutive overexpression of the OsNAS gene family reveals single-gene strategies for effective Iron- and Zinc-biofortification of rice endosperm. PLoS ONE 6(9):e24476. doi:10.1371/journal.pone.0024476

Kennedy G, Nantel G, Shetty P (2003) The scourge of "hidden hunger": global dimensions of micronnutrient deficiencies. Food Nutr Agric 32:8–16

Khalekuzzaman M, Datta K, Oliva N et al (2006) Stable integration, expression and inheritance of the ferritin gene in transgenic elite indica rice cultivar BR29 with enhanced iron level in the endosperm. Indian J Biotechnol 5:26–31

Khush G, Lee S, Cho JI, Jeon JS (2012) Biofortification of crops for reducing malnutrition. Plant Biotechnol Rep 6:195–202

Koike S, Inoue H, Mizuno D et al (2004) OsYSL2 is a rice metal-nicotianamine transporter that is regulated by iron and expressed in the phloem. Plant J 39:415–424

Lee S, Jeon US, Lee SJ et al (2009) Iron fortification of rice seeds through activation of the nicotianamine synthase gene. Proc Natl Acad Sci US A 106:22014–22019

Long JK, Bänziger M, Smith ME (2004) Diallel analysis of grain iron and zinc density in Southern African-adapted maize inbreds. Crop Sci 44:2019–2026

Lopez C, Piégu B, Cooke R et al (2005) Using cDNA and genomic sequences as tools to develop SNP strategies in cassava (*Manihot sculenta* Crantz). Theor Appl Genet 110:425–431
Lucca P, Hurrell R, Potrykus I (2001) Approaches to improving the bioavailability and level of iron in rice seeds. J Sci Food Agric 81:828–834
Lucca P, Hurrell R, Potrykus I (2002) Fighting iron deficiency anemia with iron-rich rice. J Am Coll Nutr 21:184S–190S
Marr KM, Batten GD, Blakeney AB (1995) Relations between minerals in Australian brown rice. J Sci Food Agric 68:285–291
Marschner H, Römheld V, Kissel M (1986) Different strategies in higher plants in mobilization and uptake of iron. J Plant Nutr 9:695–713
Masuda H, Usuda K, Kobayashi T et al (2009) Overexpression of the barley nicotianamine synthase gene HvNAS1 increases iron and zinc concentrations in rice grains. Rice 2:155–166
Mayer JE, Pfeiffer WH, Beyer P (2008) Biofortified crops to alleviate micronutrient malnutrition. Curr Opin Plant Biol 11:166–170
Mba REC, Stephenson P, Edwards K et al (2001) Single sequence repeat (SSR) marker survey of the cassava (*Manihot esculenta* Crantz) genome: towards an SSR-based molecular genetic map of cassava. Theor Appl Genet 1002:21–31
Meenakshi JV, Johnson NL, Manyong VM et al (2010) How cost-effective is biofortification in combating micronutrient malnutrition? An ex ante assessment. World Dev 38:64–75
Menkir A, Maziya-Dixon B (2004) Influence of genotype and environment on β-carotene content of tropical yellow-endosperm maize genotypes. Maydica 49:313–318
Menkir A, Pixley K, Maziya-Dixon B, Gedil M (2012) Recent advances in breeding maize for enhanced pro-vitamin A content. In: Worku M, Twumasi-Afriyie S, Wolde L et al (eds) Meeting the challenges of global climate change and food security through innovative maize research. Proceedings of the Third National Maize Workshop of Ethiopia. Addis Ababa, Ethiopia, pp 66–73
Micronutrient Initiative (2009) Investing in the future: a united call to action on vitamin and mineral deficiencies, Global Report 2009. Micronutrient Initiative, Ontario, Canada
Mori S (1999) Iron acquisition by plants. Curr Opin Plant Biol 2:250–253
Morillo Coronado Y (2009) Herencia del contenido de carotenos en raíces de yuca (*Manihot esculenta* Crantz). Dissertation, National University of Colombia
Morrissey J, Guerinot ML (2009) Iron uptake and transport in plants: the good, the bad, and the ionome. Chem Rev 109:4553–4567
Nirupa N, Prasad MNV (2008) Iron bioavailability, homeostasis through phytoferritins and fortification strategies: implications for human health and nutrition. In: Prasad MNV (ed) Trace elements as contaminants and nutrients: consequences in ecosystems and human health. Wiley, New York, pp 233–265
Pfeiffer WH, McClafferty B (2007) HarvestPlus: breeding crops for better nutrition. Crop Sci 47(Suppl 3):S88–S105
Pich A, Scholz G, Stephan UW (1994) Iron-dependent changes of heavy metals, nicotianamine, and citrate in different plant organs in the xylem exudate of two tomato genotypes. Nicotianamine as possible copper translocator. Plant Soil 165:189–196
Pixley K, Palacios N, Babu R, Menkir A (2011) Maize harvestplus: biofortifying maize with provitamin A carotenoids. In: Zaidi PH, Babu R, Cairns J et al (eds) Addressing climate change effects and meeting maize demand for Asia. Book of extended summaries of the 11th Asian Maze Conference, Nanning, China, 7–11 November 2011. CIMMYT, Mexico, pp 317–319
Prochnik S, Marri PR, Desany B et al (2012) The cassava genome: current progress, future directions. Trop Plant Biol 5:88–94
Qu LQ, Yoshihara T, Ooyama A et al (2005) Iron accumulation does not parallel the high expression level of ferritin in transgenic rice seeds. Planta 222:225–233
Raji AA, Anderson JV, Kolade OA et al (2009) Gene-based microsatellites for cassava (*Manihot esculenta* Crantz): prevalence, polymorphisms, and cross-taxa utility. BMC Plant Biol 9:118

Roa AC, Chavarriaga-Aguirre P, Duque MC et al (2000) Cross-species amplification of cassava (*Manihot esculenta*) (Euphorbiaceae) microsatellites: allelic polymorphism and degree of relationship. Am J Bot 87:1647–1655

Römheld V, Marschner H (1986) Evidence for a specific uptake system for iron phytosiderophores in roots of grasses. Plant Physiol 80:175–180

Schurr U (1999) Dynamics of nutrient transport from the root to the shoot. Prog Bot 60:234–253

Sperotto RA, Boff T, Duarte GL et al (2010) Identification of putative target genes to manipulate Fe and Zn concentrations in rice grains. J Plant Physiol 167:1500–1506

Sperotto RA, Ricachenevsky FK, Waldow VdA, Fett JP (2012) Iron biofortification in rice: it's a long way to the top. Plant Sci 190:24–39

Takagi S, Nomoto K, Takemoto T (1984) Physiological aspect of mugineic acid, a possible phytosiderophore of graminaceous plant. J Plant Nutr 7:469–477

Takahashi M, Nakanishi H, Kawasaki S et al (2001) Enhanced tolerance of rice to low iron availability in alkaline soils using barley nicotianamine aminotransferase genes. Nat Biotechnol 19:466–469

Tangphatsornruang S, Sraphet S, Singh R et al (2008) Development of polymorphic markers from expressed sequence tags of *Manihot esculenta* Crantz. Mol Ecol Resour 8:682–685

Theil EC (1987) Ferritin: structure, gene regulation, and cellular function in animals, plants, and microorganisms. Annu Rev Biochem 56:289–315

Theil EC (2004) Iron, ferritin, and nutrition. Annu Rev Nutr 24:327–343

Usuda K, Wada Y, Ishimaru Y et al (2009) Genetically engineered rice containing larger amounts of nicotianamine to enhance the antihypertensive effect. Plant Biotechnol J 71:87–95

Vasconcelos M, Datta K, Oliva N et al (2003) Enhanced iron and zinc accumulation in transgenic rice with the ferritin gene. Plant Sci 164:371–378

von Wiren N, Klair S, Bansal S et al (1999) Nicotianamine chelates both FeIII and FeII. Implications for metal transport in plants. Plant Physiol 119:1107–1114

Walker SP, Wachs TD, Gardner JM et al (2007) Child development: risk factors for adverse outcomes in developing countries. Lancet 369:145–157

Welch RM (1995) Micronutrient nutrition of plants. Crit Rev Plant Sci 14:49–82

Welsch R, Arango J, Bär C et al (2010) Provitamin A-accumulation in cassava (*Manihot esculenta*) roots driven by a single nucleotide polymorphism in a phytoene synthase gene. Plant Cell 22:3348–3356

WHO and FAO (2006) Guidelines on food fortification with micronutrients. World Health Organization, Geneva, Switzerland

Wirth J, Poletti S, Aeschlimann B et al (2009) Rice endosperm iron biofortification by targeted and synergistic action of nicotianamine synthase and ferritin. Plant Biotechnol J 7:631–644

Yan JB, Kandianis CB, Harjes CE et al (2010) Rare genetic variation at *Zea mays* crtRB1 increases beta-carotene in maize grain. ure Genet 42:322–327

Zhang X, Pfeiffer W, Palacios-Rojas N et al (2012) Probability of success of breeding strategies for improving pro-vitamin A content in maize. Theor Appl Genet 125:235–246

Zhao FJ, Shewry PR (2011) Recent developments in modifying crops and agronomic practice to improve human health. Food Pol 36(Suppl 1):S94–S101

Zheng L, Cheng Z, Ai C et al (2010) Nicotianamine, a novel enhancer of rice iron bioavailability to humans. PLoS One 5(4):e10190. doi:10.1371/journal.pone.0010190

Chapter 18
Genomics of Mineral Nutrient Biofortification: Calcium, Iron and Zinc

Owen A. Hoekenga

Abstract Dietary deficiencies affect nearly half of the people on the planet, who simply do not receive sufficient nutrition from the food they buy or grow. Inadequate calcium, iron, and zinc consumption create short and long term health problems, which in turn can magnify and stagnate national development. Dietary diversity, use of industrially fortified foods, and medical interventions are all effective solutions to this suite of related problems. However, each of these solutions requires infrastructure, economic support, and either education or access to markets, and thus are more suitable for the urban than rural poor. Biofortification, or the nutritional enhancement of staple and specialty crops, represents a low cost, sustainable, and potentially effective solution to addressing dietary deficiency and malnutrition in the rural poor. Recent progress on calcium, iron, and zinc biofortification using quantitative genetics, mutational genetics, and genetic engineering technologies will be discussed.

18.1 Introduction

18.1.1 Scope of Problem

A sizable fraction of the world's population does not receive adequate nutrition from the food they grow or purchase. Elemental dietary deficiencies affect approximately 3 billion people, where multiple deficiencies are possible in the same person or population (Welch and Graham 2004). Approximately one-third of the world's population suffers from iron deficiency or anemia (Boccio and Iyengar 2003). Zinc deficiency affects a similar number of people as iron deficiency, although without an inexpensive and effective clinical assay this problem is perhaps harder to track (Welch and Graham 2004). Average calcium intake in the developing world is less than half that seen in the developed world, representing approximately one-third the US RDA for adults

O. A. Hoekenga (✉)
Robert W. Holley Center for Agriculture and Health,
United States Department of Agriculture, Agricultural Research Service,
Ithaca, NY 14853, USA
e-mail: Owen.Hoekenga@ars.usda.gov

R. Tuberosa et al. (eds.), *Genomics of Plant Genetic Resources,*
DOI 10.1007/978-94-007-7575-6_18,

and less than one-quarter that required for adolescents (FAO/WHO 2004a). Strictly speaking, calcium is not a micronutrient but a macronutrient, but given the shared issues of low concentration and bioavailability from plant-based foods, calcium will be considered in this discussion. Micronutrient malnutrition reduces growth potential in children, increases morbidity and mortality, and reduces national productivity (Bouis 2003). This problem is largely due to poverty, as low incomes limit food choice and access among the rural and urban poor. These problems have solutions—dietary diversity, food product fortification, and supplementation are all effective mechanisms to reduce micronutrient malnutrition (Bouis 2003). The urban poor and rural poor have somewhat different problems, as the urban poor have greater access to markets (with processed food products) and may have access to governmental agencies or NGOs (for supplementation programs). A majority of the world's poor households are in rural areas, where the principal occupation is agriculture (Pinstrup-Andersen 2002). The rural poor often have much more limited cash reserves, less access to markets, and can be beyond the reach of supplementation/intervention style programs. The rural poor therefore require crop-based solutions to best serve their needs, by enhancing the value and quality of locally produced foodstuffs (Tontisirin et al. 2002). Biofortification, the improvement of nutritional quality in crops, is a cost-effective, efficient and sustainable strategy to reduce malnutrition (Bouis 2003; Bouis and Welch 2010). However, to achieve these biofortification objectives, clear goals for breeding or transgenic modification must exist to guide plant improvement efforts.

18.1.2 Brief Introduction to Nutritional Quality

One must consider both concentration and bioavailability of the nutrients to fully understand nutritional quality. It is not enough to know how much calcium, iron or zinc is present in a foodstuff, but also to know what fraction will be absorbed by the consumer and available for metabolism (Hunt 2003). Dietary diversity is promoted to reduce micronutrient malnutrition, to provide the consumer with a nutritionally balanced and complete diet (Tontisirin et al. 2002). Two sample diets, a rice-based diet and a maize-based diet, are shown as examples (Fig. 18.1) (FAO/WHO 2004b). At first glance, it may appear that there are excessive amounts of iron and zinc, as both diets provide more than 100 % of the US recommended daily allowance (RDA) for both mineral nutrients. However, the bioavailability of iron and zinc- the fraction within a foodstuff that is potentially digestible, absorbable, and thus useful to the consumer – is quite low from plant-based foods. HarvestPlus, the international consortium working to biofortify staple crops, takes as given that iron bioavailability from maize is approximately 5 %, while 10 % from rice (http://www.harvestplus.org, verified August 8, 2011; Bouis and Welch 2010). There is some disagreement in the field, as other estimates for iron bioavailability would place this value lower for plant-based foods (Hunt 2003) or higher (Hurrell and Egli 2010). In a similar fashion, HarvestPlus takes as given that zinc has approximately 25 % bioavailability

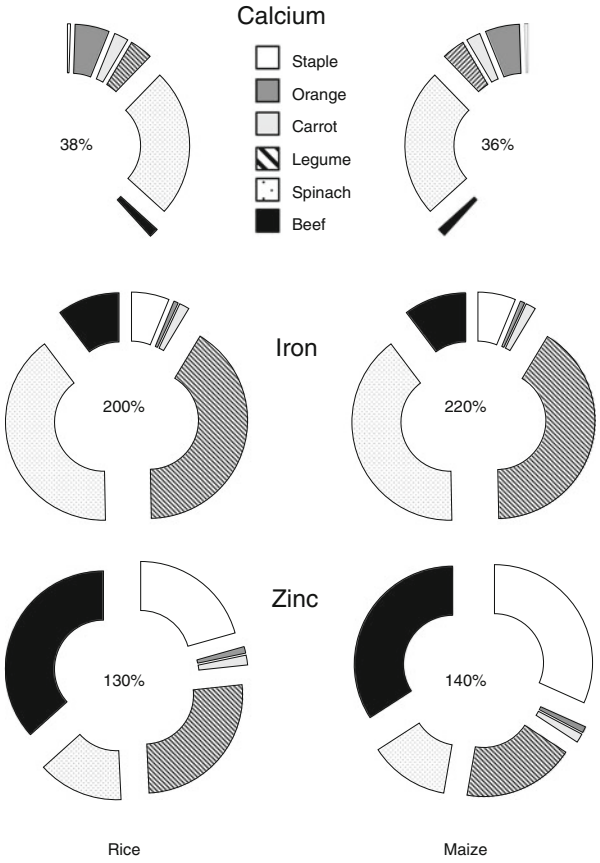

Fig. 18.1 Nutrient concentrations in sample balanced diets. Two sample diets are shown, a rice-based diet at left and a maize-based diet at right. Each of the ring graphs denotes the fractional contribution of each diet component relative to calcium, iron, and zinc content for the entire diet. The number at the center of the ring graphs indicates the percentage of the US Recommended Daily Allowance for calcium, iron, and zinc. While both diets have more than 100 % of the US RDA for iron and zinc, given the reduced bioavailability of iron and zinc from plant foodstuffs, these diets represent sufficient rather than excessive levels of nutrition. These diets were originally suggested by the World Health Organization and contain 500 g of the starchy staple and one serving of each of the other components. A culturally appropriate legume appears in each diet—lentils for the rice diet and black beans for the maize diet. (FAO/WHO 2004b)

from plant-based foods. Animal based foods often have higher intrinsic bioavailability for iron and zinc and contain compounds that promote bioavailability in a meal (Hirschi 2009; Hunt 2003). The summary of the rice-based diet suggests that the meat contribution to total iron is small and an orange provides little or no iron itself. However, the existence of promoters in these foods, such as ascorbate, cysteine rich proteins, and other as-yet undefined enhancers, dramatically enhances the bioavailability of the overall diet (Hurrell and Egli 2010). It is the presence of these

promoters in these high cost foods that allow the sample diet to reach a sufficient level of iron. While zinc bioavailability is generally thought to be greater than that of iron, zinc is generally found at lower concentration than iron. This combination of higher bioavailability with lower concentration or content brings the bioavailable level of zinc in the sample diets back to sufficient. This issue of concentration is exacerbated in the case of calcium, as the dietary requirements are more than 100 × higher on a per mass basis than either iron or zinc, while calcium concentrations in these foods are only 25 × higher. Neither of these sample diets provide more than a fraction of the US RDA for calcium, although these values are quite typical for consumers in developing countries (FAO/WHO 2004a).

The promotion of dietary diversity to address micronutrient malnutrition is an admirable goal. It does presume that the consumer has sufficient economic power and access to markets to purchase these or similarly complementary foods. For a subsistence farmer, this sort of diet assumes either economic power or sufficient flexibility and education to produce a collection of complementary plant and animal products (Tontisirin et al. 2002). Seasonality is also a major problem for micronutrient nutrition, which is to say that the combination of crops in the dry season may have a far different nutritional profile than those grown in the wet season (see FAO (2000) for a discussion with examples). Improvement of the nutritional quality of staple crops, as they often make up a majority of the bulk of a diet and are available year-round, represents a major opportunity to address and reduce micronutrient malnutrition (Arsenault et al. 2010).

This review will address some of the key issues, tools, and recent progress for calcium, iron, and zinc biofortification in plants, including a brief discussion of phytate. Developments since the excellent reviews by both Hirschi and the team of White and Broadley will be emphasized (Hirschi 2009; White and Broadley 2005, 2009). For the sake of simplicity, comparative mapping figures that summarize results within species or families will consider calcium, iron, and zinc quantitative trait locus mapping experiments simultaneously, rather than using nutrient-specific figures.

18.2 Phytate

Phytate, or inositol-hexakisphosphate, is the major storage form for phosphate in seeds and an excellent chelate for cations such as calcium, iron and zinc (Raboy et al. 2000). Quantitative trait locus (QTL) mapping has been conducted in numerous plants to dissect the genetic basis of phytate accumulation (Bentsink et al. 2003; Stangoulis et al. 2007; Zhao et al. 2008). This information can be used to estimate the bioavailable fraction of calcium, iron, and zinc to make potential selections in a breeding program. Germplasm surveys have identified accessions within and between species that have significantly lower phytate levels (Thavarajah et al. 2009; Zhao et al. 2007). *Low phytic acid* (*lpa*) mutants have been isolated in a variety of plants, including barley, maize, rice, soybean, and Arabidopsis (Kim et al. 2008; Kim and Tai 2011; Larson and Raboy 1999; Liu et al. 2007; Panzeri et al. 2011).

Animal feeding studies with *lpa* maize strains, using both pig and chicken animal models, have demonstrated enhanced nutrient utilization correlated with lower levels of phytate in the ration (Li et al. 2000; Veum et al. 2001). Phytate plays a number of positive roles within the plant, including acting in signaling pathways including disease resistance, guard cells, and abiotic stress tolerance (Georges et al. 2009; Murphy et al. 2008; Nagy et al. 2009). Phytate acts in human cells as well, also as a signaling molecule and has been implicated in cell proliferation and cancer prevention (Vucenik and Shamsuddin 2003, 2006). This duality of positive and negative effects has given some members of the biofortification community pause as regards reducing phytate as a general mechanism to enhance calcium, iron, and zinc bioavailability (Welch 2002).

18.3 Calcium

18.3.1 Calcium Biology and Human Nutrition

Most of the calcium in the human body is associated with the skeleton. However, the remaining 1 % is essential for neuromuscular function or as an enzyme co-factor (FAO/WHO 2004a). Sodium, protein, oxalate, phytate all have negative correlations with calcium nutrition, homeostasis, or bioavailability (FAO/WHO 2004a). Sodium competes with calcium for absorption in the kidney, such that high sodium levels lead to increased calcium requirement. Reducing dietary sodium reduces calcium requirement. Increasing protein content in the diet, especially animal protein, increases the loss of calcium from the body. There is a fortuitous correlation between animal protein consumption rates and calcium intakes, as those with lowest animal protein are also likely to have lowest calcium intake. Consuming animal protein at this lower level leads to less loss of calcium. Oxalate and phytate are both good ligands for calcium, such that foods that are rich in either will have low bioavailable calcium, as the calcium will too tightly bound to easily exchange and be absorbed (Gibson et al. 2010). To maximize effective calcium intake then requires reducing sodium, protein, oxalate, and phytate intake while consuming foods rich in calcium. This is often very difficult to accomplish from plant-based foods, such that even in the US and UK calcium deficiency is not uncommon (Broadley and White 2010).

18.3.2 Conventional Genetic Approaches

18.3.2.1 Quantitative Trait Locus Mapping

Ionomics, or the systems biology approach to elemental analysis, is a highly cost effective phenotyping technology (Baxter et al. 2007; Lahner et al. 2003). Ionomic analysis of recombinant inbred (RI) or other mapping populations returns data on

as many as 16 elements in a single analytical chemical assay; this also enables a network or system-wide view of plant/mineral relations (Baxter 2009; Baxter et al. 2007). The majority of quantitative trait locus (QTL) mapping experiments that have examined calcium concentration in either leaves or seeds therefore also collected data on other elements. As mentioned previously, while the interaction of concentration and bioavailability is important to understanding nutritional quality, calcium as a biofortification goal requires perhaps greater emphasis on concentration due to deficit in many plant foods relative to human nutritional demand. While QTL mapping is a highly useful tool for the discovery of markers to use in breeding programs, in the post genomic era it is the integration of results across multiple studies and even species where value is now added and the pace of discovery accelerates. Thus, studies in model systems such as Arabidopsis add power and perspective to mapping work in economically important Brassicaceae, such as *B. oleraceae* and *B. rapa*.

Multiple investigators have chosen the *Arabidopsis thaliana* Landsberg *erecta* × Cape Verde Islands RI population for ionomic profiling (Buescher et al. 2010; Vreugdenhil et al. 2004; Waters and Grusak 2008). This level of replication for an experiment offers special insight into the role of experimental and environmental conditions for the detection and reproducibility of QTL. Two of the studies focused on ionomic profiling of seeds, using greenhouse grown plants on soilless potting mixes (Vreugdenhil et al. 2004; Waters and Grusak 2008). For example, one of the QTL for calcium content in the seed was detected by both groups, landing on the same confidence interval (Fig. 18.2). This is remarkable, given that calcium loading into the developing fruit can be quite influenced by transpiration; one imagines that greenhouses in Wageningen and Houston are quite different during the course of a growing season (Pomper and Grusak 2004). From the perspective of a model system, understanding the seed ionome of Arabidopsis has value. However, the leafy Brassicas are the nearest economically important relatives of Arabidopsis that are consumed as whole food products. With this in mind, Buescher et al. examined the leaf ionome of two Arabidopsis RI populations, both Landsberg *erecta* × Cape Verde Islands and Bay × Shahdara (Buescher et al. 2010). It is perhaps not surprising that a QTL for leaf calcium was not found in the same interval bounded by *phyA* and *nga63* by all three teams (Fig. 18.2). Instead, a QTL for leaf calcium was detected at a more proximal location on chromosome 1 in the Bay × Shahdara population (Buescher et al. 2010). Broadley et al., working with a broad cross in *B. oleraceae*(A12DHd, Chinese cabbage × GDDH33, Italian broccoli) detected many QTL for leaf calcium and magnesium in both field and greenhouse grown plants (Broadley et al. 2008). One of the Brassica leaf calcium QTL on chromosome C7 was detected in the syntenic position to the leaf calcium QTL on Arabidopsis chromosome 1, although to make the comparison clear several physical and genetic maps were required to compensate for the dearth of shared markers (Alonso-Blanco et al. 1998; Lister and Dean 1993; Swarbreck et al. 2008). If this relationship is more than just coincidental, one expects that the wealth of genomic resources in Arabidopsis will leverage gene discovery in Brassica (Atwell et al. 2010).

Fig. 18.2 Comparative mapping of calcium and iron QTL within the Brassicaceae. A subset of QTL mapping results from four studies were compared and combined into the single analysis presented here (Broadley et al. 2008; Buescher et al. 2010; Vreugdenhil et al. 2004; Waters and Grusak 2008). Two additional genetic maps and a sequence-based map were required to join the studies together (Alonso-Blanco et al. 1998; Lister and Dean 1993; Swarbreck et al. 2008). Long open bars depict genetic maps for chromosome 1 from *Arabidopsis thaliana* or chromosome C7 from *Brassica oleraceae*. The long gray bar depicts the pseudomolecule for *Arabidopsis thaliana* chromosome 1, with marker positions indicated by black bars (Mb). Shorter gray bars indicate QTL confidence intervals, with light gray corresponding to iron and medium gray corresponding to calcium. Lines connect shared markers between maps, where genetic distances are constant within all maps. Scale bars indicate genetic and physical distances

18.3.2.2 Germplasm Evaluation and Development

Evaluating diverse germplasm for novel traits is a key component of any breeding program. In the post genomic era, these germplasm surveys can also be used for genetic analyses beyond the initial descriptions of new phenotypes. As the cost of genotyping rapidly falls, information density per experimental event is increasing st a similar rate (Elshire et al. 2011). Phenotypic surveys can be used both to identify new donors to breeding programs, as well as fuel association mapping that can dissect the bases of complex traits (Elshire et al. 2011; Hamblin et al. 2011). Association mapping takes advantage of ancestral recombination that has accumulated through evolutionary time, rather than recent recombination created by experimenters in F_2 or RI populations (Hamblin et al. 2011). Greater effective recombination rates between markers and traits leads to identifying smaller chromosomal segments (or even single genes, or portions of genes) if the analysis has sufficient statistical power to succeed. Association mapping can be used either with candidate genes (Harjes et al. 2008; Krill et al. 2010) or with genome-wide markers (Huang et al. 2010). In a practical sense, there will be great value in revisiting older datasets with high quality phenotypes and applying genotyping-by-sequencing methods to characterize the germplasm collections from which they were generated (see a summary of germplasm screening studies as tabulated in White and Broadley (2009)). Alternatively, phenotyping a characterized germplasm collection such as the Maize Diversity Panel (282 inbred lines) or the Rice Diversity Panel (517 landraces) using a high impact method, such as ionomic profiling, would also leverage existing

genomics resources for rapid progress in a particular breeding program (Gore et al. 2009; Hoekenga et al. 2010; Huang et al. 2010).

One of the main limitations to association mapping is inadequate statistical power—not having enough accessions in the collection to resolve the effect of the QTL of interest. This may be due to minor alleles, which may be present at too low frequency, or small phenotypic effects that are difficult to detect reliably. This issue is likely why a recent genome-wide association study of Arabidopsis failed to identify many ionomic QTL, as only 93 accessions were included in this portion of their work plan (Atwell et al. 2010). The study of leaf calcium in *Brassica oleraceae* described by Broadley et al. included a survey of 376 accessions from the Warwick HRI Genetic Resources Unit, a deliberate subsample from more than 4,300 accessions (Broadley et al. 2008). If and when this subsample of the Warwick HRI core collection is genotyped by sequencing, association mapping could clarify which genes are important regulators for leaf calcium and which alleles would be most prized for breeding.

18.3.2.3 Mutant Analysis

Mutants have been used to modify calcium nutritional quality, largely focusing on reducing phytate and oxalate concentrations, to examine whether calcium bioavailability would be increased while calcium concentrations were held constant. The maize *lpa1* mutation was evaluated using human feeding studies, where subjects consumed tortillas prepared from both mutant and wild-type sibling isolines (Hambidge et al. 2005). The *lpa1* maize represented a ~ 60 % reduction in seed phytate; an increase of calcium absorption of similar magnitude was observed in the human subjects following the *lpa1* meal. This indicated that phytate concentration in the maize seed was a major determinant for calcium bioavailability, and thus manipulating phytate levels could have a large impact on calcium nutritional quality (Hambidge et al. 2005). A similar removal of an anti-nutrient was observed in *Medicago truncatula* mutant that does not accumulate oxalate crystals in leaves (Morris et al. 2007). In this study, mice were fed *cod5* mutant leaves or the isogenic wild-type sibling; more than 90 % of leaf oxalate found in the wild-type was missing in the mutant while calcium levels were unchanged. Mice fed the *cod5* mutant absorbed ~ 23 % more calcium than the wild-type sibling (Morris et al. 2007). While this is an obvious improvement in calcium bioavailability, the lack of congruence between the reduction in oxalate and the increase in calcium bioavailability suggest that other factors are in play. Additionally, the calcium oxalate crystals serve as major feeding deterrent to herbivores, such that the *cod5* mutant exhibits greater defoliation than wild-type plants when under insect pressure (Korth et al. 2006). Mitigation of unintended effects will be crucial to the success of a biofortified variety, if either phytate or oxalate reduction is implemented as a general strategy for enhancing calcium bioavailability. It is worth remembering that the intended end-user is not a mechanized farmer in the Global North with easy access to irrigation and agro-chemicals, but rather a subsistence farmer in the Global South, where attenuated drought tolerance or herbivory resistance could dramatically reduce the chances of adoption and widespread use.

An unexpectedly neutral effect on calcium bioavailability was observed in a soybean mutant, selected for enhanced aroma in soymilk products (Martino et al. 2008). The cultivar UFV-116 was developed to reduce lipoxygenase activity and volatile compound production, which give a disagreeable flavor to soy products. It was observed during evaluation of the new variety that it contained 20 % more calcium, but also 30 % more phytate and nearly double the oxalate levels found in the progenitor. This begged the question of whether this cultivar would have unacceptable reductions in calcium bioavailability (Martino et al. 2008). Following a mouse protocol very similar to that used in the *cod5* evaluation in Medicago, mice were fed either UFV-116 or a standard commercial variety. No differences in calcium bioavailability were observed, in spite of the increased concentrations of inhibitors. Martino et al. hypothesized that the increased concentration of calcium may have compensated; it is also possible that additional unintended effects to food composition or quality took place during derivation of the UFV-116 cultivar. These results suggest that more detailed analyses of calcium bioavailability are required, using animal subjects or other proxies, to better describe the enhancers and inhibitors of calcium nutritional quality. This information will enhance the accuracy of phenotyping and accelerate the rate of selection.

18.3.3 Transgenic Approaches

A highly effective strategy to increase calcium concentration in plant tissues was developed a number of years ago by Hirschi et al. A calcium/proton antiporter from Arabidopsis, when over-expressed, was capable of increasing calcium concentrations by approximately 50 % in carrot (Park et al. 2004). This technology was subsequently applied to lettuce, tomato, potato, and rice (Kim et al. 2005; Park et al. 2005a, 2005b, 2009). Importantly, feeding studies with both mice and humans demonstrated that the increased calcium concentration in carrots translated into increased bioavailable calcium and thus nutritional quality (Morris et al. 2007, 2008). Expression of the *sCAX1* recombinant calcium/proton antiporter did not alter taste characteristics of transgenic lettuce, suggesting there were no unintended effects that would limit consumer acceptance due to flavor as result of the transgene (Park et al. 2009). Expression of the *sCAX1* antiporter in tomato extended shelf life, much as exogenously applied calcium does in many horticultural crops, providing an added benefit to producers and consumers beyond the enhanced nutritional quality (Park et al. 2005a). However, plants expressing *sCAX1* appeared to have an increased calcium requirement, as additional calcium was required to restore normal growth and alleviate deficiency symptoms. A related antiporter, *CAX4*, was also transformed into tomato; while calcium levels were elevated 50 % in fruit and shelf life was extended by a smaller degree than with *sCAX1*, there were no alterations to gross growth or morphology of the plant nor were calcium requirements increased (Park et al. 2005a). These experiments suggest that the thoughtful application of particular calcium/proton antiporters may have widespread utility in enhancing calcium concentration and bioavailability in transgenic plants.

18.4 Iron

18.4.1 Iron Biology and Human Nutrition

Iron and zinc deficiency in humans occur at approximately equal frequency, affecting on the order of 2 billion people (Welch and Graham 2004). Iron deficiency is far easier to diagnose than zinc deficiency, as inexpensive and robust clinical assays exist to estimate iron status. The impact of iron deficiency on human health is also relatively easy to quantify, both in terms of effects on the individual and on society. Iron deficiency diminishes cognitive development in children, reduces productivity in adults, and generally increases disease symptoms and severity (Caballero 2002; FAO/WHO 2004c). The cumulative effects of iron deficiency and anemia reduce national productivity. Economists have modeled the drag on Gross Domestic Product for countries with high rates of iron deficiency in children and adults (Horton and Ross 2003). For countries like Honduras, where iron deficiency rates were estimated to be between 12 and 17 % for women, men, and children, the reduction to GDP was estimated at 2.0 % per year. For countries like Bangladesh, where iron deficiency rates were estimated at 73 % for women and children and 60 % for men, the reduction to GDP was estimated at 7.9 % per year (Horton and Ross 2003). Clearly, alleviating iron deficiency could have a major impact on the health and productivity of the world as a whole and for many of the least developed countries in particular.

18.4.2 Conventional Genetic Approaches

18.4.2.1 Quantitative Trait Locus Mapping

One of the best-understood genetic systems for the inheritance of iron concentration is common bean, *Phaseolus vulgaris*. Legumes in general are well known to be excellent sources of both micronutrients and protein, making them an important component of a diverse and balanced diet. It is also recognized that the two ancestral races of beans, Andean and Mesoamerican, have important differences in micronutrient quality as Andean accessions had generally higher iron but lower zinc concentrations than those found in Mesoamerican germplasm (Islam et al. 2002). Before QTL mapping was applied to the Arabidopsis seed ionome, common bean was studied using a Mesoamerican × wild cross (Guzman Maldonado et al. 2003). While this study used an AFLP-based map, which makes comparison difficult with other studies, four later studies are quite easily comparable (Fig. 18.3) (Blair et al. 2009, 2010c, 2011; Cichy et al. 2009). These four studies from the International Center for Tropical Agriculture (CIAT) shared common research farms, which allowed for an excellent system to test the role of environmental effects, while examining genetic diversity within (Blair et al. 2010c, 2011) and between (Blair et al. 2009; Cichy et al. 2009) the Mesoamerican and Andean gene pools. There were at least

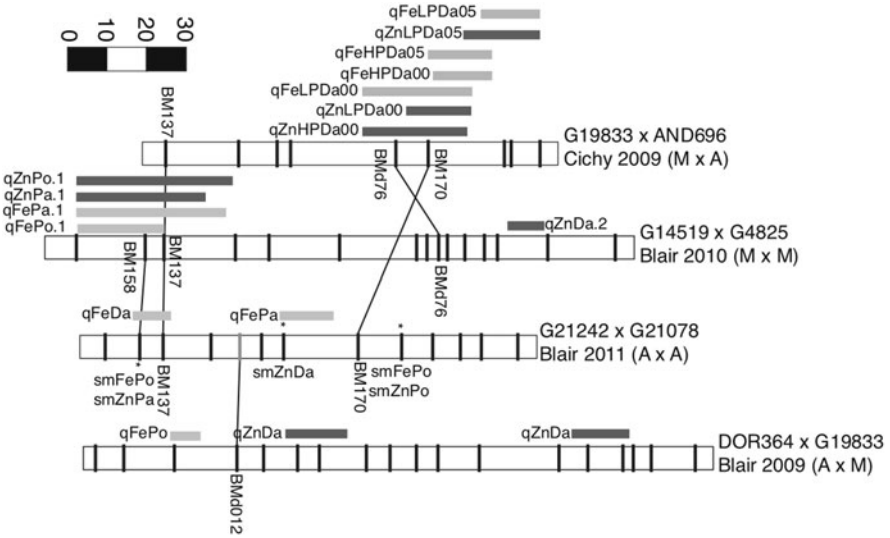

Fig. 18.3 Comparative mapping of iron and zinc QTL within *Phaseolus vulgaris*. A subset of QTL mapping results from four studies were compared and combined into the single analysis presented here (Bentsink et al. 2003; Blair et al. 2009, 2011; Welch 2002). Plants were grown at one of three sites: Da, Darien; Po, Popayán; Pa, Palmira. An additional genetic map was required to act as a bridge to join these maps together. Long open bars depict common bean chromosome 6, with the first author and year of publication appearing as a label. Shorter bars indicate confidence intervals for QTL identified by composite interval mapping, with light gray corresponding to iron and deep gray corresponding to zinc. QTL are named by the element, the site where detected, a two digit code for year or single digit code if more than one QTL was identified. QTL identified by single marker analysis are indicated with an asterisk and are referred to as "sm" but otherwise use the same naming convention. Lines connect shared markers between maps, where genetic distances are constant

three clusters of repeatedly detected QTLs for iron and/or zinc concentration on bean chromosome 6 (Fig. 18.3). In these populations, QTL for iron and zinc co-located more often than not, which makes introgression and improvement that much easier as one introgressed region could enhance multiple traits. The effects of these QTLs were observed on multiple research farm sites, suggesting high penetrance against environmental effects. A common underlying mechanism is plausible, at least for the clusters detected in the G19833 × AND696 and G14519 × G4825 populations, as the same parental allele confers the superior effect in both clusters (AND696 and G14519, respectively). Analysis of metal transporters in Arabidopsis has demonstrated that iron and zinc can share common components for seed loading (Walker and Waters 2011; Waters et al. 2006). Another cluster of QTL for iron and zinc concentration was detected in the two crosses between Mesoamerican and Andean varieties on chromosome 11 (Blair et al. 2009; Cichy et al. 2009). Coincidence of the *ferric reductase* gene on chromosome 7 with a cluster of iron concentration QTL in the same vicinity was explored (Blair et al. 2010b). While ferric reductase

activity did not co-segregate with QTL for iron concentration, this result is still interesting from a comparative genomics perspective. A pair of QTLs for seed iron and zinc was detected in *Lotus japonicus* on chromosome 1, in close proximity to the *LjFRO1* locus (Klein and Grusak 2009). As the data from bean argue against ferric reductase activity as a driver for iron concentration differences in seeds, the syntenic relationship between bean and Lotus QTL suggest comparative genomic analysis may assist both research efforts. By the same token, repeated detection of iron QTL on Arabidopsis chromosome 1 between two wholly unrelated mapping populations suggests that genomics-based solutions may help answer physiological questions (Fig. 18.2).

Breeding for iron concentration alone may not be sufficient to enhance iron nutritional quality. Plant foods have especially poor iron bioavailability, such that improving bioavailability may have a greater impact than increasing iron concentration (Frossard et al. 2000). Given the role of phytate in reducing iron bioavailability, including phytate in this discussion might be worthwhile. Phenotyping for phytate requires liquid chromatography or other specialized analytical chemistry techniques (Bentsink et al. 2003). However, the majority of phosphorous in the seed is in the form of phytate, such that total phosphorous can serve as a proxy for phytate (Raboy et al. 2000). Once one has produced an ionomic profile, calculating the ratio between iron and phosphorous would be a zero-cost method to estimate bioavailable iron (Simic et al. 2009). Iron, phosphorous, and zinc concentrations were measured in F_4 families derived from B84 and Os6-2, temperate North American inbreds of contrasting heteroic groups (Simic et al. 2009). Both phosphorous/iron and phosphorous/zinc ratios were calculated to estimate bioavailable iron and bioavailable zinc in the population. Heritability for these traits and others were estimated using replicated trials over a two-year period and found to be moderate (0.46 for iron, 0.59 for zinc and 0.63 for phosphorous). While all traits were estimated to be responsive to selection, the relatively high molar ratios of phosphorous to iron and zinc suggested that the B84 × Os6-2 population would not be a good one to derive new materials with enhanced nutritional quality (Simic et al. 2009). Simic et al. came to this conclusion due to experiments performed with a human cell culture bioassay capable of estimating iron bioavailability in foodstuffs (Glahn et al. 1998). The human cell culture (Caco-2) provides a highly tractable system to explore nutritional quality differences using a human epithelial cell line, where results are well correlated with human feeding studies (Yun et al. 2004). Caco-2 studies that varied phytate to iron ratios suggested that maximal inhibition occurred at 10:1 (Glahn et al. 2002). The phosphorous to iron ratios from the B84 × Os6-2 population suggested that this 10:1 value was already exceeded, and thus they concluded their population was not suitable for selection (Simic et al. 2009). However, bioavailable iron levels were not directly tested and thus the value for these ratio estimates is somewhat questionable.

A straighter path of inquiry into the genetic bases for iron bioavailability would be to use the Caco-2 bioassay as a phenotyping tool and estimate the nutritional quality of a mapping population directly. Lung'aho et al. did exactly this, evaluating the Intermated B73 × Mo17 (IBM) RI population using both ionomic profiling and the Caco-2 bioassay (Lung'aho et al. 2011). The IBM population offers at least two

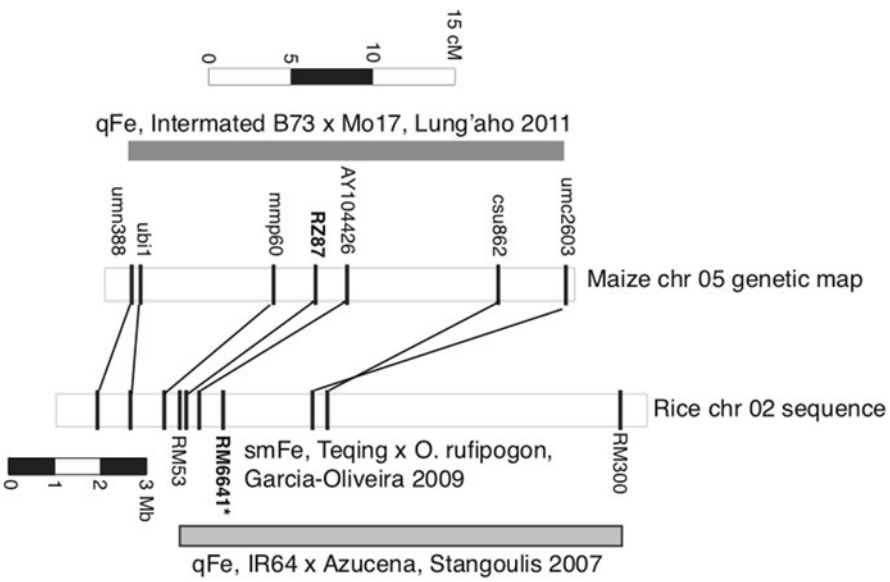

Fig. 18.4 Comparative mapping of iron QTL within the Poaceae. A subset of QTL mapping results from two rice mapping studies and one from maize were combined into the single analysis here (Garcia-Oliveira et al. 2009; Lung'aho et al. 2011; Stangoulis et al. 2007). Genetic markers from maize chromosome 5 were located within the rice chromosome 2 pseudomolecule using BLAST (Lawrence et al. 2008). The scale bar at left indicates genetic distances in maize, while the scale bar at right indicates physical distances in rice. Confidence intervals for the two QTL detected using composite interval mapping are shown as gray bars; the QTL detected single marker analysis is indicated with an asterisk

advantages over other RI sets in maize; four rounds of intermating before genotypes were fixed means that an increased number of meioses were captured and thus genetic resolving capability is markedly increased (Lee et al. 2002). The B73 inbred was also the target of the US maize genome sequencing program, such that a high quality, annotated genome sequence exists (Schnable et al. 2009). Ionomic profiling was conducted on IBM samples from two sites with dramatically different soils (Cornell University Farm: Lima Silt Loam (alfisol); North Carolina State University Farm: Norfolk Loamy Sand (ultisol), where only the iron results were reported in detail (Lung'aho et al. 2011). One replicate season was also analyzed using the Caco-2 bioassay to estimate bioavailable iron. Only a small positive correlation existed between total and bioavailable iron, suggesting that differences in the grain metabolome were larger determinants for iron nutritional quality than iron concentration. QTL were detected for both iron concentration and bioavailability but were not coincident, reaffirming the distinction between the two components to iron nutritional quality. One of the maize seed QTL for iron found on chromosome 5 was in a syntenic region to two seed iron QTL found in rice, which may suggest conserved processes are at work in the Poaceae as they apparently are in the Brassicaceae and Fabaceae (Fig. 18.4) (Garcia-Oliveira et al. 2009; Stangoulis et al.

2007). Comparison of QTL locations between more recent maize studies from each of these groups did not identify common QTL within maize (ie. B84 × Os6-2 vs. B73 × Mo17), while the potential for conserved QTL between maize and rice is still plausible (Simic et al. 2012; Baxter et al. 2013).

There is a healthy debate in the human nutrition community on the utility and accuracy of the Caco-2 bioassay for iron nutritional quality evaluation (Fairweather-Tait et al. 2005). In the ideal world, animal or human feeding studies should be used to verify predictions made in the laboratory, no matter the basis of selection. Animal and human feeding studies are quite expensive undertakings, such that they can only be practically applied as validation and not phenotyping tools. With this in mind, Lung'aho et al. created derivatives from the IBM RI set with altered iron nutritional quality, using molecular markers to select and the Caco-2 bioassay to verify (Lung'aho et al. 2011). Creating a small panel of highly related lines enabled both multi-environment testing and opened the possibility of animal testing. Much like Blair et al. demonstrated efficacy of their QTL across multiple sites, nearly isogenic derivatives from the IBM collection were reported to produce significant differences in iron bioavailability with maize grown on five sites (Lung'aho et al. 2011). As the derivatives were constructed into both parental backgrounds, nearly isogenic hybrids could be made to facilitate large-scale grain production and animal feeding studies. A small scale poultry feeding study demonstrated that the predictions made with Caco-2 were accurate; further, the hybrid predicted to have high iron bioavailability was capable of maintaining broiler chickens in an iron replete state (Hoekenga et al. 2011). This result was in spite of maize providing the sole source of iron in the ration and that both the presumptive high and low iron bioavailability hybrids had the same concentration of iron in the grain (Hoekenga et al. 2011). This demonstrates that the disconnection between iron concentration and iron bioavailability observed with the Caco-2 bioassay is quite firm; manipulating the chemistry of the grain, while holding iron concentration constant, can dramatically enhance iron bioavailability. A larger feeding study, using maize produced in a subsequent season on different plots, confirmed the initial poultry feeding study (Tako et al. 2013). These studies also highlight the utility of broiler (meat) chickens as a low-cost system for animal feeding studies (Tako et al. 2010). While young pigs are the typical human analog for nutritional studies, chickens eat less, have greater dietary flexibility, and are far easier to rear than piglets (Tako et al. 2009). The broiler chicken model has been used to validate iron biofortified beans (based on increased iron concentration) and maize (based on increased bioavailable iron) (Tako et al. 2011; Tako et al. 2013).

18.4.2.2 Germplasm Evaluation and Development

The observations made with the maize IBM derivatives validate the utility of the Caco-2 bioassay as a phenotyping tool. While these results were confirmed using animal feeding studies, these new materials are hardly ready to be released for widespread cultivation in the Global South. Instead, this result should be seen for the

case study that is it and as a guide for future efforts. Much as HarvestPlus supported bean breeding at CIAT, HarvestPlus has also supported maize breeding at CIMMyT, the International Maize and Wheat Improvement Center. Elite accessions from the CIMMyT Highland Yellow and Highland Zinc germplasm programs were evaluated along with experimental materials from the pro-vitamin A biofortification breeding program, to create a series of three test panels to evaluate iron nutritional quality at multiple sites (Pixley et al. 2011). Materials were evaluated using both ionomic and Caco-2 profiling, to examine the heritability and combining ability of iron, zinc, and pro-vitamin A concentration with iron bioavailability. As all materials were agronomically superior, grain yield was also evaluated. Trends were not consistent across all germplasm pools and trial sites. However, zinc concentration and iron bioavailability were negatively correlated in at least some trials, consistent with the previous observations of zinc interference with iron uptake (Glahn et al. 2002). Neither iron concentration nor grain yield was correlated with iron bioavailability; pro-vitamin A carotenoids did have a positive correlation, consistent with previous reports of the antioxidant activity increasing iron bioavailability (Garcia-Casal et al. 1998). Heritability for each of the traits examined was moderate and quite sensitive to environmental effects, which is also consistent with the observations made in the IBM population (Lung'aho et al. 2011; Pixley et al. 2011). These results suggest that breeding for enhanced iron bioavailability is an achievable goal, but one that will require the development of lower cost selection tools like molecular markers until the program is small enough to be reasonable to evaluate with the Caco-2 bioassay or animal feeding trials.

As iron deficiency is a problem throughout the world, it may be more practical to identify useful alleles within existing breeding programs than to import them from outside. Disease and pest pressures, together with product quality issues associated with traditional cooking practices, may make introgression of new alleles especially time consuming. With these issues in mind, Blair et al. have begun to examine the ionomic and genetic characteristics in African bean germplasm, as the Great Lakes region is a secondary site of genetic diversity (Blair et al. 2010a). Average iron concentrations were higher in the Central African genotypes compared with earlier surveys of Latin American varieties, while average zinc levels were very similar (Blair et al. 2010a; Islam et al. 2002).

18.4.2.3 Mutant Analysis

For comprehensive reviews of the use of mutants in understanding iron acquisition, transport, storage, and utilization, the reader is directed to one of several recent reviews (Conte and Walker 2011; Jeong and Guerinot 2009; Morrissey and Guerinot 2009; Walker and Waters 2011; White and Broadley 2009). A recent report identifies a new component for root/shoot signaling and determination of the leaf ionome (Chao et al. 2011). Several years ago, a large-scale ionomic survey of mutagenized Arabidopsis identified a number of putative mutants important for the accumulation

of particular elements in leaves (Lahner et al. 2003). One such mutation, *tsc10*, increased foliar potassium and sodium while decreasing calcium, iron, magnesium, and molybdenum (Chao et al. 2011). Microarray-enabled bulk segregant analysis rapidly mapped the mutation to a gene essential for sphingolipid biosynthesis, 3-ketodihydrosphinganine reductase (3-KDS). Sphingolipids are essential components to membranes and are hypothesized to help define lipid rafts, or physio-chemically distinct portions of particular membranes (Mongrand et al. 2010). These lipid rafts apparently restrict certain proteins, such as proton ATPases, to particular regions within a cell, creating microdomains. Lipid rafts are also associated with signaling processes, as leucine-rich repeat protein receptor kinases are enriched in these fractions relative to the plasma membrane at large (Mongrand et al. 2010). A mutation within the 3-KDS gene was responsible for the *tsc10* mutation, with the *tsc10* mutant having markedly less 3-KDS activity and significant changes to sphingolipid composition within plasma membranes (Chao et al. 2011). Grafting experiments demonstrated that while the mutation was originally detected using an ionomic screen of leaves, the site of function is in the root. Mutant roots grafted to wild type shoots have the same outcome as intact mutant plants, while wild type roots grafted to mutant shoots do not have a disturbed leaf ionome (Chao et al. 2011). The *tsc10* mutant also had alterations in suberin composition of the root, the water impermeable barrier that is crucial to xylem-mediated transport of water and minerals from the root to shoot. Interestingly, a suberin-related mutation, *esb1*, was also identified from the same mutagenesis and ionomic analysis (Baxter et al. 2009). While it is not yet clear how these alterations to suberin and lipid raft composition affect mineral transport and homeostasis, these experiments affirm the utility of ionomic profiling of mutant collections to discover novel participants in these physiological processes.

18.4.3 Transgenic Approaches

There have been several strategies to increase bioavailable iron in seeds or leaves, including the addition of promoters of iron bioavailability such as nicotianamine (Lee et al. 2009), the addition of iron-rich storage proteins to increase the pool of iron (Murray-Kolb et al. 2002), and addition of phytases, enzymes that degrade phytate (Drakakaki et al. 2005). Stacking transgenes has also been shown to be effective, where bioavailable iron in maize grain was enhanced using a combination of a fungal phytase with a soybean ferritin (Drakakaki et al. 2005). While these approaches have been discussed elsewhere, it is worth mentioning that the degree of improvement observed by Drakakaki et al., according to the Caco-2 bioassay, was similar to the gain observed by Lung'aho et al. (Lung'aho et al. 2011). As the conventional breeding approach relied on neither iron concentration nor decreasing phytate, it is an open question as to how much gain would be observed in bioavailable iron by transforming the ferritin/phytase transgenes into the high iron bioavailable derivatives from the IBM. To reach the target levels for iron, calcium, or zinc bioavailability may require the combination of conventional and transgenic plant improvement technologies.

18.5 Zinc

18.5.1 Zinc Biology and Human Nutrition

Zinc is often considered together with iron during discussions of micronutrient malnutrition. Often, populations are simultaneously affected by both iron and zinc deficiency, but this is not always true. This is in part due to the fact that iron is the fourth most abundant element in the earth's crust, while zinc is deficient in many agricultural soils (Cakmak 2009). A study of pregnant women in Ethiopian farming communities identified that while one-third were iron deficient, all were zinc deficient due to very low zinc concentrations in the locally produced foods (Abebe et al. 2008). Zinc deficiency is more difficult to assess clinically than iron deficiency, as serum zinc concentrations are not always an accurate descriptor for zinc status. Often, subject interviews and cognitive testing are required to fully diagnose zinc deficiency, making data collection far more involved and expensive (Stoecker et al. 2009). Zinc is required as a co-factor for more than 300 proteins in the human body, and is essential for DNA synthesis, degradation of carbohydrates, lipids, and proteins, and for immune system function (FAO/WHO 2004d). While iron and zinc deficiency both cause chronic health problems, zinc deficiency is more likely to be associated with acute health problems and death (Cakmak 2009).

18.5.2 Conventional Genetic Approaches

18.5.2.1 Quantitative Trait Locus Mapping

One of the best examples for the map-based cloning of a grain quality QTL is the case of *GPC-1B* found in wheat (Uauy et al. 2006). In addition to regulating grain protein concentration, this NAC1 family transcription factor also regulates the translocation of iron and zinc into the developing grain (Waters et al. 2009). The superior allele was found in wild emmer wheat and conveys a $\sim 15\,\%$ increase in both zinc and iron (Uauy et al. 2006). Due to zinc deficient soils in wheat growing regions of countries such as Turkey and India, where zinc deficiency is quite prevalent in both the crops and people, the presence of the *GPC-1B* allele alone may not be sufficient to biofortify the wheat. Wheat breeders may also have to consider a second trait, zinc efficiency, to guarantee success of a biofortified wheat cultivar. Plants with high zinc efficiency are more capable of growing robustly at low available zinc, protecting growth and yield potential. Physiological studies have made some progress at describing the underlying mechanisms for zinc efficiency (Hacisalihoglu et al. 2001, 2003; Khoshgoftarmanesh et al. 2009). More recently, a pair of transcription factors (*bZIP19* and *bZIP23*) were identified in Arabidopsis that activate genes inducible by zinc deficiency; both the genes and their palindromic recognition sequence seem to be highly conserved in plants, with clear homologs in rice and poplar (Assunção et al. 2010).

Zinc efficiency has also been studied in common bean (Hacisalihoglu et al. 2004). This is fortunate, as the inheritance of seed zinc concentration is well described in this system (Fig. 18.3). Much of Latin America, southwestern and southern Asia, China and Australia have zinc deficient soils (Hacisalihoglu and Kochian 2003). For zinc biofortified varieties to be successful in these regions, varieties will need to combine both zinc efficiency and bioavailability. Selections made in the Great Lakes region of Africa, where soil zinc levels are sufficient, may not be appropriate in Mexico or Colombia, where soil zinc levels are low. The combination of both seed quality and zinc efficiency may require more sophisticated breeding strategies to achieve the desired outcome (Hamblin et al. 2011). Using either association mapping or genomic selection would require an investment in genome-wide markers. However the return on this investment would be at least two-fold: to develop well-adapted biofortified beans (taking full advantage of Latin American and African genetic diversity); to observe the differential genomic responses to selection in these two environments, which may answer many questions about what determines zinc efficiency in common bean.

18.5.2.2 Germplasm Evaluation and Development

One of the benefits of the post genomic era is the reduced cost of entry to new species and extant but orphan crops (Doust et al. 2009; Varshney et al. 2009). Where genetically and phenotypically diverse germplasm collections exist, characterization and utilization can quickly follow. For example, pearl millet (*Pennisetum glaucum*) is a cereal crop well adapted to arid conditions and largely planted in the Global South; only two of the top twenty millet producers are G8 countries (http://faostat.fao.org, accessed 08/08/11). Pearl millet has high seed zinc and iron concentrations, superior to sorghum, which would be the principal niche competitor as cereals with high levels of abiotic stress tolerance (White and Broadley 2009). While improvement of pearl millet has been hampered by lack of investment, new releases are being made with enhanced nutritional quality (Rai et al. 2008). As the cost of genotyping continues to fall and as statistical and bioinformatics methods improve, one can hope that traditional crops with low yields but superior qualities will be the beneficiaries of the systems biology revolution.

18.6 Future Directions

Effective biofortification improvement programs will require a combination of disciplines to achieve full success. Plant breeders, agronomists, biotechnologists, food scientists, animal scientists, and human nutritionists will all be involved at one point or another during the development, evaluation, and release of biofortified crops. With input from so many different kinds of science, a systems biology approach will be possible. The mechanisms that underlie calcium, iron, and zinc acquisition,

transport, loading, and bioavailability in seeds and leaves will need to be better understood so that realistic improvement targets can be met. The suites of genes and alleles necessary to achieve biofortification goals in some regions may not work well in others, due to differences in soil fertility or other environmental variables. One imagines that progress will be more rapid, at least initially, in maize, rice, and wheat due to the size of investment and research communities. However, as we learn more of the molecular mechanisms for calcium, iron, and zinc bioavailability, the pace of biofortification research and improvement should accelerate in many crop systems.

Acknowledgments This work was supported by USDA ARS. The author would like to thank Mrs. Meghan den Bakker and Ms. Ellie Taylor for their excellent work on the research farm during the writing of this review.

References

Abebe Y, Bogale A, Hambidge KM et al (2008) Inadequate intakes of dietary zinc among pregnant women from subsistence households in Sidama, Southern Ethiopia. Public Health Nutr 11: 379–386

Alonso-Blanco C, Peeters AJ, Koornneef M et al (1998) Development of an AFLP based linkage map of Ler, Col and Cvi *Arabidopsis thaliana* ecotypes and construction of a Ler/Cvi recombinant inbred line population. Plant J 14:259–271

Arsenault JE, Yakes EA, Hossain MB et al (2010) The current high prevalence of dietary zinc inadequacy among children and women in rural Bangladesh could be substantially ameliorated by zinc biofortification of rice. J Nutr 140:1683–1690

Assunção AG, Herrero E, Lin YF et al (2010) *Arabidopsis thaliana* transcription factors bZIP19 and bZIP23 regulate the adaptation to zinc deficiency. Proc Acad Natl Sci U S A 107:10296–10301

Atwell S, Huang YS, Vilhjalmsson BJ et al (2010) Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. Nature 465:627–631

Baxter I (2009) Ionomics: studying the social network of mineral nutrients. Curr Opin Plant Biol 12:381–386

Baxter I, Ouzzani M, Orcun S et al (2007) Purdue ionomics information management system. An integrated functional genomics platform. Plant Physiol 143:600–611

Baxter I, Hosmani PS, Rus A et al (2009) Root suberin forms an extracellular barrier that affects water relations and mineral nutrition in Arabidopsis. PLoS Genet 5:e1000492

Baxter IR, Gustin JL, Settles AM, Hoekenga OA (2013) Ionomic characterization of maize kernels in the Intermated B73x Mo17 (IBM) population. 53:209–220

Bentsink L, Yuan K, Koornneef M, Vreugdenhil D (2003) The genetics of phytate and phosphate accumulation in seeds and leaves of *Arabidopsis thaliana*, using natural variation. Theor Appl Genet 106:1234–1243

Blair MW, Astudillo C, Grusak MA et al (2009) Inheritance of seed iron and zinc concentrations in common bean (*Phaseolus vulgaris* L.). Mol Breeding 23:197–207
Blair MW, Chaves A, Tofino A et al (2010a) Extensive diversity and inter-genepool introgression in a world-wide collection of indeterminate snap bean accessions. Theor Appl Genet 120:1381–1391
Blair MW, Knewtson SJ, Astudillo C et al (2010b) Variation and inheritance of iron reductase activity in the roots of common bean (*Phaseolus vulgaris* L.) and association with seed iron accumulation QTL. BMC Plant Biol 10:215
Blair MW, Medina JI, Astudillo C et al (2010c) QTL for seed iron and zinc concentration and content in a Mesoamerican common bean (*Phaseolus vulgaris* L.) population. Theor Appl Genet 121:1059–1070
Blair MW, Astudillo C, Rengifo J et al (2011) QTL analyses for seed iron and zinc concentrations in an intra-genepool population of Andean common beans (*Phaseolus vulgaris* L.). Theor Appl Genet 122:511–521
Boccio JR, Iyengar V (2003) Iron deficiency: causes, consequences, and strategies to overcome this nutritional problem. Biol Trace Elem Res 94:1–32
Bouis HE (2003) Micronutrient fortification of plants through plant breeding: can it improve nutrition in man at low cost? Proc Nutr Soc 62:403–411
Bouis HE, Welch R (2010) Biofortification—a sustainable agricultural strategy for reducing micronutrient malnutrition in the global south. Crop Sci 50:20–32
Broadley MR, Hammond JP, King GJ et al (2008) Shoot calcium and magnesium concentrations differ between subtaxa, are highly heritable, and associate with potentially pleiotropic loci in *Brassica oleracea*. Plant Physiol 146:1707–1720
Broadley MR, White PJ (2010) Eats roots and leaves. Can edible horticultural crops address dietary calcium, magnesium and potassium deficiencies? Proc Nutr Soc 69:601–612
Buescher E, Achberger T, Amusan I et al (2010) Natural genetic variation in selected populations of *Arabidopsis thaliana* is associated with ionomic differences. PLoS One 5:e11081
Caballero B (2002) Global patterns of child health: the role of nutrition. Ann Nutr Metab 46(Suppl 1):3–7
Cakmak I (2009) Enrichment of fertilizers with zinc: an excellent investment for humanity and crop production in India. J Trace Elem Med Biol 23:281–289
Chao DY, Gable K, Chen M et al (2011) Sphingolipids in the root play an important role in regulating the leaf ionome in *Arabidopsis thaliana*. Plant Cell 23:1061–1081
Cichy KA, Caldas GV, Snapp SS, Blair MW (2009) QTL analysis of seed iron, zinc, and phosphorous levels in an andean bean population. Crop Sci 49:1742–1750
Conte SS, Walker EL (2011) Transporters contributing to iron trafficking in plants. Mol Plant 4:464–476
Doust AN, Kellogg EA, Devos KM, Bennetzen JL (2009) Foxtail millet: a sequence-driven grass model system. Plant Physiol 149:137–141
Drakakaki G, Marcel S, Glahn RP et al (2005) Endosperm-specific co-expression of recombinant soybean ferritin and Aspergillus phytase in maize results in significant increases in the levels of bioavailable iron. Plant Mol Biol 59:869–880
Elshire RJ, Glaubitz JC, Sun Q et al (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6:e19379
Fairweather-Tait S, Lynch S, Hotz C et al (2005) The usefulness of in vitro models to predict the bioavailability of iron and zinc: a consensus statement from the HarvestPlus expert consultation. Int J Vitam Nutr Res 75:371–374
FAO (2000) Food insecurity: when people live with hunger and fear starvation. The state of food insecurity in the world. FAO, Rome
FAO/WHO (2004a) Calcium. Joint FAO/WHO expert consultation on human vitamin and mineral requirements, 2nd edn. Bangkok, pp 59–93. Accessed from http://whqlibdoc.who.int/publications/2004/9241546123_chap4.pdf. Verified 11/14/13
FAO/WHO (2004b) Food as a source of nutrients. Joint FAO/WHO expert consultation on human vitamin and mineral requirements, 2nd edn. Bangkok, pp 318–337. Accessed from http://whqlibdoc.who.int/publications/2004/9241546123_chap17.pdf. Verified 11/14/13.

FAO/WHO (2004c) Iron. Joint FAO/WHO expert consultation on human vitamin and mineral requirements. FAO/WHO, Bangkok, pp 246–278. Accessed from http://whqlibdoc.who.int/publications/2004/9241546123_chap13.pdf. Verified 11/14/13.

FAO/WHO (2004d) Zinc. Joint FAO/WHO expert consultation on human vitamin and mineral requirements, 2nd edn. Bangkok, pp 230–245. Accessed from http://whqlibdoc.who.int/publications/2004/9241546123_chap12.pdf. Verified 11/14/13.

Frossard E, Bucher M, Mächler F et al (2000) Potential for increasing the content and bioavailability of Fe, Zn, and Ca in plants for human nutrition. J Sci Food Agric 80:861–879

Garcia-Casal MN, Layrisse M, Solano L et al (1998) Vitamin A and beta-carotene can improve nonheme iron absorption from rice, wheat and corn by humans. J Nutr 128:646–650

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

Georges F, Das S, Ray H, Bock C et al (2009) Over-expression of *Brassica napus* phosphatidylinositol-phospholipase C2 in canola induces significant changes in gene expression and phytohormone distribution patterns, enhances drought tolerance and promotes early flowering and maturation. Plant Cell Environ 32:1664–1681

Gibson RS, Bailey KB, Gibbs M, Ferguson EL (2010) A review of phytate, iron, zinc, and calcium concentrations in plant-based complementary foods used in low-income countries and implications for bioavailability. Food Nutr Bull 31:S134–S146

Glahn RP, Lee OA, Yeung A, Goldman MI, Miller DD (1998) Caco-2 cell ferritin formation predicts nonradiolabeled food iron availability in an in vitro digestion/Caco-2 cell culture model. J Nutr 128:1555–1561

Glahn RP, Wortley GM, South PK, Miller DD (2002) Inhibition of iron uptake by phytic acid, tannic acid, and ZnCl2: studies using an in vitro digestion/Caco-2 cell model. J Agric Food Chem 50:390–395

Gore MA, Chia JM, Elshire RJ et al (2009) A first-generation haplotype map of maize. Science 326:1115–1117

Guzman Maldonado H, Martinez O, Acosta GJAetal (2003) Putative quantitiative trait loci for physical and chemical components of common bean. Crop Sci 43:1029–1035

Hacisalihoglu G, Kochian LV (2003) How do some plants tolerate low levels of soil zinc? Mechanisms of zinc efficiency in crop plants. New Phytol 159:341–350

Hacisalihoglu G, Hart JJ, Kochian LV (2001) High- and low-affinity zinc transport systems and their possible role in zinc efficiency in bread wheat. Plant Physiol 125:456–463

Hacisalihoglu G, Hart JJ, Wang YH et al (2003) Zinc efficiency is correlated with enhanced expression and activity of zinc-requiring enzymes in wheat. Plant Physiol 131:595–602

Hacisalihoglu G, Hart JJ, Vallejos CE, Kochian LV (2004) The role of shoot-localized processes in the mechanism of Zn efficiency in common bean. Planta 218:704–711

Hambidge KM, Krebs NF, Westcott J et al (2005) Absorption of calcium from tortilla meals prepared from low-phytate maize. Am J Clin Nutr 82:84–87

Hamblin MT, Buckler ES, Jannink JL (2011) Population genetics of genomics-based crop improvement methods. Trends Genet 27:98–106

Harjes CE, Rocheford TR, Bai L et al (2008) Natural genetic variation in lycopene epsilon cyclase tapped for maize biofortification. Science 319:330–333

Hirschi KD (2009) Nutrient biofortification of food crops. Annu Rev Nutr 29:401–421

Hoekenga O, Gustin J, Flint-Garcia S et al (2010) Ionomics of the maize nested association mapping panel. Vitro Cell Dev Biol-Anim 46:S9–S9

Hoekenga OA, Lung'aho MG, Tako E et al (2011) Iron biofortification of maize grain. Plant Genet Res 9:327–329

Horton S, Ross J (2003) The economics of iron deficiency. Food Pol 28:51–75

Huang X, Wei X, Sang T et al (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. Nat Genet 42:961–967

Hunt JR (2003) Bioavailability of iron, zinc, and other trace minerals from vegetarian diets. Am J Clin Nutr 78:633S–639S

Hurrell R, Egli I (2010) Iron bioavailability and dietary reference values. Am J Clin Nutr 91:1461S–1467S

Islam FMA, Basford KE, Jara C et al (2002) Seed compositional and disease resistance differences among gene pools in cultivated common bean. Genet Resour Crop Evol 49:285–293

Jeong J, Guerinot ML (2009) Homing in on iron homeostasis in plants. Trends Plant Sci 14:280–285

Khoshgoftarmanesh AH, Sadrarhami A, Sharifi HR et al (2009) Selecting zinc-efficient wheat genotypes with high grain yield using a stress tolerance index. Agron J 101:1409–1416

Kim KM, Park YH, Kim CK et al (2005) Development of transgenic rice plants overexpressing the Arabidopsis H+/Ca2+ antiporter CAX1 gene. Plant Cell Rep 23:678–682

Kim SI, Tai TH (2011) Identification of genes necessary for wild-type levels of seed phytic acid in *Arabidopsis thaliana* using a reverse genetics approach. Mol Genet Genomics 286:119–133

Kim SI, Andaya CB, Goyal SS, Tai TH (2008) The rice OsLpa1 gene encodes a novel protein involved in phytic acid metabolism. Theor Appl Genet 117:769–779

Klein MA, Grusak MA (2009) Identification of nutrient and physical seed trait QTL in the model legume *Lotus japonicus*. Genome 52:677–691

Korth KL, Doege SJ, Park SH et al (2006) Medicago truncatula mutants demonstrate the role of plant calcium oxalate crystals as an effective defense against chewing insects. Plant Physiol 141:188–195

Krill AM, Kirst M, Kochian LV et al (2010) Association and linkage analysis of aluminum tolerance genes in maize. PLoS One 5:e9958

Lahner B, Gong J, Mahmoudian M et al (2003) Genomic scale profiling of nutrient and trace elements in *Arabidopsis thaliana*. Nat Biotechnol 21:1215–1221

Larson SR, Raboy V (1999) Linkage mapping of maize and barley myo-inositol 1-phosphate synthase DNA sequences: correspondence with a low phytic acid mutation. Theor Appl Genet 99:27–36

Lawrence CJ, Harper LC, Schaeffer ML et al (2008) MaizeGDB: the maize model organism database for basic, translational, and applied research. Int J Plant Genomics 2008:496957. doi: 10.1155/2008/496957.

Lee M, Sharopova N, Beavis WD et al (2002) Expanding the genetic map of maize with the intermated B73x Mo17 (IBM) population. Plant Mol Biol 48:453–461

Lee S, Jeon US, Lee SJ et al (2009) Iron fortification of rice seeds through activation of the nicotianamine synthase gene. Proc Acad Natl Sci U S A 106:22014–22019

Li YC, Ledoux DR, Veum TL et al (2000) Effects of low phytic acid corn on phosphorus utilization, performance, and bone mineralization in broiler chicks. Poult Sci 79:1444–1450

Lister C, Dean C (1993) Recombinant inbred lines for mapping RFLP and phenotypic markers in *Arabidopsis thaliana*. Plant J 4:745–750

Liu QL, Xu XH, Ren XL et al (2007) Generation and characterization of low phytic acid germplasm in rice (*Oryza sativa* L.). Theor Appl Genet 114:803–814

Lung'aho MG, Mwaniki AM, Szalma SJ et al (2011) Genetic and physiological anlaysis of iron biofortification in maize kernels. PLoS One 6:e20429

Martino HS, Martin BR, Weaver CM et al (2008) A soybean cultivar lacking lipoxygenase 2 and 3 has similar calcium bioavailability to a commercial variety despite higher calcium absorption inhibitors. J Food Sci 73:H33–H35

Mongrand S, Stanislas T, Bayer EM et al (2010) Membrane rafts in plant cells. Trends Plant Sci 15:656–663

Morris J, Hawthorne KM, Hotze T et al (2008) Nutritional impact of elevated calcium transport activity in carrots. Proc Natl Acad Sci U S A 105:1431–1435

Morris J, Nakata PA, McConn M et al (2007) Increased calcium bioavailability in mice fed genetically engineered plants lacking calcium oxalate. Plant Mol Biol 64:613–618

Morrissey J, Guerinot ML (2009) Iron uptake and transport in plants: the good, the bad, and the ionome. Chem Rev 109:4553–4567

Murphy AM, Otto B, Brearley CA et al (2008) A role for inositol hexakisphosphate in the maintenance of basal resistance to plant pathogens. Plant J 56:638–652

Murray-Kolb LE, Takaiwa F, Goto F et al (2002) Transgenic rice is a source of iron for iron-depleted rats. J Nutr 132:957–960

Nagy R, Grob H, Weder B, Green P et al (2009) The Arabidopsis ATP-binding cassette protein AtMRP5/AtABCC5 is a high affinity inositol hexakisphosphate transporter involved in guard cell signaling and phytate storage. J Biol Chem 284:33614–33622

Panzeri D, Cassani E, Doria E et al (2011) A defective ABC transporter of the MRP family, responsible for the bean lpa1 mutation, affects the regulation of the phytic acid pathway, reduces seed myo-inositol and alters ABA sensitivity. New Phytol 191:70–83

Park S, Kim CK, Pike LM et al (2004) Increased calcium in carrots by expression an Arabidopsis H+/Ca2+ transporter. Mol Breeding 14:275–282

Park S, Cheng NH, Pittman JK et al (2005a) Increased calcium levels and prolonged shelf life in tomatoes expressing Arabidopsis H+/Ca2+ transporters. Plant Physiol 139:1194–1206

Park S, Kang TS, Kim CK et al (2005b) Genetic manipulation for enhancing calcium content in potato tuber. J Agric Food Chem 53:5598–5603

Park S, Elless MP, Park J et al (2009) Sensory analysis of calcium-biofortified lettuce. Plant Biotechnol J 7:106–117

Pinstrup-Andersen P (2002) Food and agricultural policy for a globalizing world: preparing for the future. Amer J Agr Econ 84:1201–1214

Pixley KV, Palacio-Rojas N, Glahn R (2011) The usefulness of iron bioavailability as a target trait for breeding maize (*Zea mays* L.) with enhanced nutritional value. Field Crops Research 123:153–160

Pomper KW, Grusak MA (2004) Calcium Uptake and Whole-plant Water Use Influence Pod Calcium Concentration in Snap Bean Plants. J Amer Soc Hort Sci 129:890–895

Raboy V, Gerbasi PF, Young KA et al (2000) Origin and seed phenotype of maize low phytic acid 1-1 and low phytic acid 2-1. Plant Physiol 124:355–368

Rai KN, Hash CT, Singh AK, Velu G (2008) Adaptation and quality traits of a germplasm-derived commercial seed parent of pearl millet. Plant Genet Res News 154:20–24

Schnable PS, Ware D, Fulton RS et al (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112–1115

Simic D, Sudar R, Ledencan T et al (2009) Genetic variation of bioavailable iron and zinc in grain of a maize population. J Cereal Sci 50:392–397

Simic D, Drinic SM, Zdunic Z et al (2012) Quantitative trait loci for biofortification traits in maize. J Hered 103:47–54

Stangoulis J, Huynh BL, Welch R et al (2007) Quantitative trait loci for phytate in rice grain and their relationship with grain micronutrient content. Euphytica 154:289–294

Stoecker BJ, Abebe Y, Hubbs-Tait L et al (2009) Zinc status and cognitive function of pregnant women in Southern Ethiopia. Eur J Clin Nutr 63:916–918

Swarbreck D, Wilks C, Lamesch P et al (2008) The Arabidopsis Information Resource (TAIR): gene structure and function annotation. Nucleic Acids Res 36:D1009–D1014

Tako E, Laparra JM, Glahn RP et al (2009) Biofortified black beans in a maize and bean diet provide more bioavailable iron to piglets than standard black beans. J Nutr 139:305–309

Tako E, Rutzke MA, Glahn RP (2010) Using the domestic chicken (*Gallus gallus*) as an in vivo model for iron bioavailability. Poult Sci 89:514–521

Tako E, Blair MW, Glahn RP (2011) Biofortified red mottled beans (*Phaseolus vulgaris* L.) in a maize and bean diet provide more bioavailable iron than standard red mottled beans: Studies in poultry (*Gallus gallus*) and in vitro digestion/Caco-2 model. Nutrition J 10:113

Tako E, Hoekenga OA, Kochian LV, Glahn RP (2013) High bioavailability iron maize (Zea mays L.) developed through molecular breeding provides more bioavailable iron in vitro (Caco-2 model) and in vivo (Gallus gallus). Nutr J 12:3 doi: 10.1186/1475-2891-12-3

Thavarajah P, Thavarajah D, Vandenberg A (2009) Low phytic acid lentils (*Lens culinaris* L.): a potential solution for increased micronutrient bioavailability. J Agric Food Chem 57:9044–9049

Tontisirin K, Nantel G, Bhattacharjee L (2002) Food-based strategies to meet the challenges of micronutrient malnutrition in the developing world. Proc Nutr Soc 61:243–250

Uauy C, Distelfeld A, Fahima T, Blechl A, Dubcovsky J (2006) A NAC Gene regulating senescence improves grain protein, zinc, and iron content in wheat. Science 314:1298–1301
Varshney RK, Close TJ, Singh NK et al (2009) Orphan legume crops enter the genomics era! Curr Opin Plant Biol 12:202–210
Veum TL, Ledoux DR, Raboy V, Ertl DS (2001) Low-phytic acid corn improves nutrient utilization for growing pigs. J Anim Sci 79:2873–2880
Vreugdenhil D, Aarts MG, Koornneef M et al (2004) Natural variation and QTL analysis for cationic mineral content in seeds of *Arabidopsis thaliana*. Plant Cell Environ 27:828–839
Vucenik I, Shamsuddin AM (2003) Cancer inhibition by inositol hexaphosphate (IP6) and inositol: from laboratory to clinic. J Nutr 133:3778S–3784S
Vucenik I, Shamsuddin AM (2006) Protection against cancer by dietary IP6 and inositol. Nutr Cancer 55:109–125
Walker EL, Waters BM (2011) The role of transition metal homeostasis in plant seed development. Curr Opin Plant Biol 14:318–324
Waters BM, Chu HH, Didonato RJ et al (2006) Mutations in Arabidopsis yellow stripe-like1 and yellow stripe-like3 reveal their roles in metal ion homeostasis and loading of metal ions in seeds. Plant Physiol 141:1446–1458
Waters BM, Grusak MA (2008) Quantitative trait locus mapping for seed mineral concentrations in two *Arabidopsis thaliana* recombinant inbred populations. New Phytol 179:1033–1047
Waters BM, Uauy C, Dubcovsky J, Grusak MA (2009) Wheat (*Triticum aestivum*) NAM proteins regulate the translocation of iron, zinc, and nitrogen compounds from vegetative tissues to grain. J Exp Bot 60:4263–4274
Welch RM (2002) Breeding strategies for biofortified staple plant foods to reduce micronutrient malnutrition globally. J Nutr 132:495S–499S
Welch RM, Graham RD (2004) Breeding for micronutrients in staple food crops from a human nutrition perspective. J Exp Bot 55:353–364
White PJ, Broadley MR (2005) Biofortifying crops with essential mineral elements. Trends Plant Sci 10:586–593
White PJ, Broadley MR (2009) Biofortification of crops with seven mineral elements often lacking in human diets–iron, zinc, copper, calcium, magnesium, selenium and iodine. New Phytol 182:49–84
Yun S, Habicht J, Miller D, Glahn R (2004) An in vitro digestion/Caco-2 cell culture system accurately predicts the effects of ascorbic acid and polyphenolic compounds on iron bioavailability in humans. J Nutr 134:2712–2721
Zhao J, Paulo MJ, Jamar D et al (2007) Association mapping of leaf traits, flowering time, and phytate content in *Brassica rapa*. Genome 50:963–973
Zhao J, Jamar DC, Lou P et al (2008) Quantitative trait loci analysis of phytate and phosphate concentrations in seeds and leaves of *Brassica rapa*. Plant Cell Environ 31:887–900

Chapter 19
Optimising the Content and Composition of Dietary Fibre in Wheat Grain for End-use Quality

Peter R. Shewry, Luc Saulnier, Kurt Gebruers, Rowan A.C. Mitchell, Jackie Freeman, Csilla Nemeth and Jane L. Ward

Abstract The cell wall polysaccharides of wheat affect the quality for food processing, livestock feed and distilling, and also form a major source of dietary fibre (DF) for human health. Multisite field trials of wheat genotypes show extensive variation in the content of the major types of cell wall polysaccharide in wheat, arabinoxylan and β-glucan, much of which is heritable and hence available for exploitation by plant breeders. Furthermore, contents of DF components have not declined as a result of intensive wheat breeding. The identification of candidate genes for DF synthesis using bioinformatics and RNAi suppression in transgenic wheat will allow the content and composition of DF components to be fine-tuned for specific end uses.

19.1 Introduction

Dietary fibre (DF) accounts for about 10–14 % of the whole wheat grain and about 2.5–4.5 % of white flour. The major components are cell wall polysaccharides and in particular arabinoxylan (AX) and (1–3,1–4)-β-D-glucan (β-glucan) which account for about 70 and 20 % of the total, respectively, in the starchy endosperm (i.e. white flour) and 65 and 30 %, respectively, in the aleurone. The compositions of the cell wall of the different grain tissues are summarised in Table 19.1.

P. R. Shewry (✉) · R. A.C. Mitchell · J. Freeman · C. Nemeth · J. L. Ward
Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ , UK
e-mail: peter.shewry@rothamsted.ac.uk

L. Saulnier
INRA UR1268 Biopolymers, Interactions Assemblies, 44316 Nantes, France

K. Gebruers
Laboratory of Food Chemistry and Biochemistry, and Leuven Food Science and Nutrition Research Centre (LFoRCe), Catholic University Leuven, Kasteelpark Arenberg 20, Box 2463, 3001 Leuven, Belgium

C. Nemeth
School of Biosciences, University of Nottingham, Sutton Bonnington Campus, Sutton Bonnington, Leicestershire LE12 5RD, UK

R. Tuberosa et al. (eds.), *Genomics of Plant Genetic Resources,*
DOI 10.1007/978-94-007-7575-6_19,

Table 19.1 The cell wall composition (%) of wheat grain tissues

Tissue	Cellulose	β-Glucan[a]	AX[b]	Glucomannan	Lignin
Starchy endosperm (flour)	2–4	20	70	2–7	0
Aleurone (bran)	2	30	65	2	0
Pericarp (bran)	30	0	60	0	12

[a](1–3,1–4)-β-D-glucan. Traces of (1–3)-β-D glucan (callose) may also be present in grain fractions
[b]Arabinoxylan. This varies in structure within and between tissues, with forms which are highly cross-linked and substituted with glucuronic acid (GAX) being present in the pericarp

The content and composition of the cell walls of wheat affects the quality for all major end uses.

19.1.1 For Human Health

DF, including cereal grain fibre, has significant benefits in reducing the incidence of a number of diseases, including those associated with the "metabolic syndrome" (Nugent 2004). DF is not digested in the upper gastrointestinal (GI) tract and many of the beneficial effects result from fermentation in the colon. However, it also has effects throughout the GI tract, with insoluble fibre providing faecal bulk to speed up transit while soluble fibre may delay the uptake of nutrients in the small intestine, and hence lower the glycaemic index, by increasing the viscosity of the digesta. Wheat products contribute about 20 % of the total intake of DF in UK adults, with about 11 % coming from white bread and 5 % from wholemeal bread (Steer et al. 2008). However, the average daily intake of DF in UK adults is only about 13.5 g/day, which falls far short of the dietary recommendation of 18 g/day (www.heartstat.org). Even higher daily intakes are recommended by other health authorities, for example, at least 25 g/day by the European Food Safety Authority (EFSA 2010) and 30 and 25 g/day for male and female adults, respectively, in Australia and New Zealand by NHMRC (NHMRC 2006). In France, the average daily intake is 21.6 g for men and 17.5 g for women and a 50 % increase in dietary fibre consumption is recommended (Bertrais et al. 2005). Increasing the consumption of whole grain products, or the DF content of white products, are therefore viable strategies for increasing the DF intakes of populations at low cost.

19.1.2 For Alcohol Production

Substantial volumes of wheat are used for alcohol production, either for distillation to produce alcoholic drinks or for bioethanol production. A low content of DF is desirable for these processes, not only to increase alcohol yield (due to a higher proportion of starch) but also to reduce the viscosity which results in technical problems during processing.

19.1.3 For Livestock Feed

High contents of DF components, and particularly of soluble fibre, are also detrimental for livestock feed, by increasing the viscosity (Pettersson and Åman 1989). This is a particular problem for feeding chickens and other poultry where it leads to sticky faeces. Distillers dried grains and soluble (DDGS) from distilling and biofuel production are also enriched in fibre (over 30 % dry weight) which may result in excessive viscosity when used for livestock feed.

19.1.4 For Food Processing

Wheat AX also affect the functional properties of wheat flour and dough and xylanases are often used in starch/gluten separation processes and in commercial breadmaking and to increase the ratio of water-extractable:water-unextractable AX (Courtin and Delcour 2002; Goesaert et al. 2005).

Little was known about the extent of variation in the composition of wheat flour cells walls, and the relative effects of genotype and environment on this, until a detailed study was carried out under the EU HEALTHGRAIN programme. HEALTHGRAIN was supported as part of the Sixth Framework Food Research Programme (2005–2010) with the aim of improving the well-being of EU consumers and decreasing the incidence of the metabolic syndrome by increasing the uptake of beneficial components in wholegrain cereals and fractions derived from them (Poutanen et al. 2008, 2010). The programme included a range of bioactive components, including DF, and the studies ranged from crop genetics to processing, dietary intervention and consumer expectations.

19.2 Variation in the Content of Wheat DF

In order to determine the extent of variation in the total contents of DF components a selection of 150 wheat genotypes were grown in field plots on a single site in Martonvasar (Hungary) in 2005 (Ward et al. 2008). Most were commercial cultivars, comprising 107 from Europe, 28 from the USA and 15 from Asia, the Near East and Australia. 103 of the lines were modern cultivars released within the last 50 years but they also included 10 land races and old varieties and 23 germplasm and breeding lines. The grain was milled and white flour and bran fractions analysed for total AX (TOT-AX) and water-extractable AX (WE-AX). Wholemeal fractions were also analysed for β-glucan. The results of these analyses (which are reported in full by Gebruers et al. 2008) are summarised in Table 19.2. Wide variation in the contents of all fractions was observed, from 1.36–2.74 % dry weight for flour TOT-AX, 0.29–1.38 % for flour WE-AX, 12–22.6 % for bran TOT-AX, 0.3–0.9 % for bran WE-AX, 0.5–1.0 % for wholemeal β-glucan. It is notable that although the

Table 19.2 Means, ranges and heritabilities of dietary fibre components in wheat

Component	Mean content in 150 lines[a] (mg/g dm)	Range of content in 150 lines (fold)[a]	Mean A:X ratio in 150 lines[a]	Range of A:X ratio in 150 lines[a]	Heritability in 26 lines (approx)[b] (%)
β-Glucan	7.16	1.93			51
WE-AX flour	5.15	4.76	0.48	0.39–0.57	60
WE-AX bran	4.20	2.36	1.01	0.71–1.63	48
TOT-AX flour	19.31	2.01	0.58	0.49–0.71	72
TOT-AX bran	177.909	1.27	0.62	0.53–0.71	39

[a]Based on analyses of 150 lines grown on a single site (Gebruers et al. 2008) [b]Based on 26 lines grown in multiple environments

bran fraction is rich in AX only a small proportion of this (a mean of about 2.4 %) is soluble, whereas soluble AX accounts for about a third of the total in flour.

However, only single samples of grain were analysed and, although these were grown in adjacent plots on the same site, at least some of the variation observed could result from the effects of environment or interactions between the genotype and the environment (G × E interactions). In order to determine these effects, 23 lines were selected from the 150 and were grown together with three additional lines on the same site for 2 further years (2006, 2007) and on sites in the UK and France in 2007 only (Shewry et al. 2010a). The 26 lines comprised 24 winter type and two spring type and the former only were also grown in Poland in 2007. This gave data sets for six environments (sites or years) for 21 lines and five environments for five lines. Analysis of the flour, bran and wholemeal fractions for DF components allowed the variation in content to be partitioned into the effects of genotype, environment and G × E. These calculations indicated that about 72 % of the variance in flour TOT-AX, 60 % of the variance in flour WE-AX, 39 % of the variance in bran TOT-AX, 48 % of the variance in bran WE-AX and 51 % of the variance in wholemeal β-glucan could be ascribed to the genotype (Table 19.2). This high level of heritability indicates that the content of AX is amenable to selection by plant breeders, either to increase the content (particularly of WE-AX in flour) for human health, or to decrease the content for livestock feed, distilling and biofuel production.

However, a study of two sets of spring and winter wheats (25 lines of each) grown at three locations each in the USA showed that environment had a much greater effect on WE-AX and TOT-AX than genotype in winter wheats, and on WE-AX but not TOT-AX in spring wheats (Li et al. 2009). The authors concluded that the relative effects of genotype and environment on AX amount and composition depended on the precise genotypes and environments which were studied.

19.3 Variation in the Structure of Wheat DF

AX has a linear backbone of (1→4) linked β-D-xylopyranosyl units which are either unsubstituted, monosubstituted with arabinose on position O-3 or disubstituted with arabinose on the O-2 and O-3 positions, with the degree of substitution varying between genotypes (Fig. 19.1). Arabinose residues monosubstituted at the O-3 position

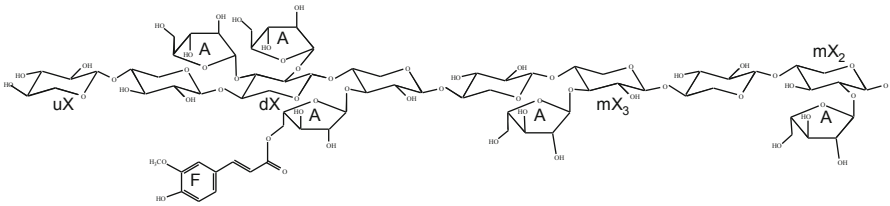

Fig. 19.1 Main structural features of cereal arabinoxylan. *A* arabinose, *X* xylose, *F* ferulic acid, *uX* unsubstituted xylose, *dX* di-substituted xylose, mX_3 O-3 mono-substituted xylose, mX_2 O-2 mono-substituted xylose

may also be substituted with ferulic acid, which can form diferulate cross-links by oxidation.

Variation in the structure of AX in wholemeal samples of the 150 wheat genotypes in the HEALTHGRAIN study was determined by digesting the TOT-AX with endoxylanase and separating the oligosaccharides (AXOS) by HP-AEC (Shewry et al. 2010b; Toole et al. 2011). Previous analyses had shown that the AXOS contained monosubstituted, disubstituted or both mono and disubstituted xylose residues, and principal component analysis of their proportions showed that almost 80 % of the variation in AX structure between the 150 genotypes could be ascribed to differences in the arabinose substitution pattern with cultivars on the left hand side of Fig. 19.2 containing a higher proportion of monosubstitution and cultivars on the right side a higher proportion of disubstitution (Fig. 19.2). This structural variation has been observed in other wheat populations and is related to genotype. Although the Healthgrain diversity screen did not provide a balanced view of the global diversity in wheat germplasm (cultivars from Western Europe being over-represented) we did not observe any differences in the structural diversity of AX related to the origin of the wheat. Saulnier et al. (2007) also analysed 90 lines derived from a cross between wheat varieties differing in total AX content and showed that 60.6–70.3 % of the xylose units in AX were unsubstituted, 16.1–27.3 % were monosubstituted and 7.7–19.3 % were disubstituted.

Differences in the proportion of mono, di- and unsubstituted xylose residues result in differences in the ratio of arabinose to xylose residues on hydrolysis (A:X ratio). Analysis of the HEALTHGRAIN lines showed that the A:X ratio ranged from 0.39 to 0.57 for flour WE-AX, from 0.49–0.71 for flour TOT-AX, from 0.71–1.63 for bran WE-AX and from 0.53–0.71 for bran TOT-AX in the HEALTHGRAIN lines (Table 19.2). A similar range in the A:X ratio of WE-AX (0.39–0.57) was reported for a population of 90 lines from the cross Synthetic × Oparta (Saulnier et al. (2007).

β-glucan comprises glucose residues joined by (1→3) and (1→4) linkages (Fig. 19.3). Single (1→3) linkages are usually separated by two or three (1→4) linkages, resulting in the release of trisaccharide (G3) and tetrasaccharide (G4) units on digestion with a specific lichenase (*endo*-(1–3)(1–4)-β-D-glucan-4-glucanohydrolase) enzyme. However, longer stretches of (1→4) linked glucan of up to 14 units have been reported for wheat bran β-glucan (Liu et al. 2006). Such regions are sometimes referred to as "cellulose-like", as cellulose is (1→4)-β-D-glucan without any (1→3) linkages.

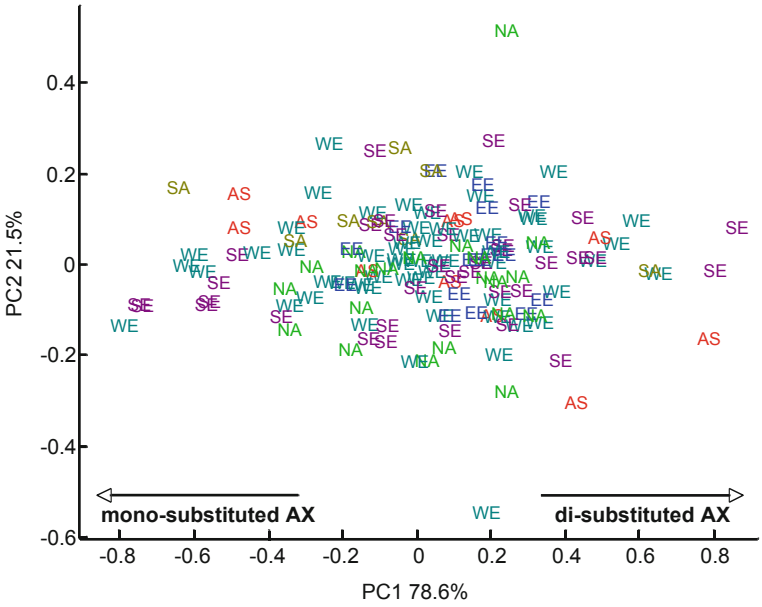

Fig. 19.2 Similarity map from Principal Component Analysis of enzyme mapping of AX from lines in the HEALTHGRAIN diversity screen (150 lines). PC1 and PC2 are the first two principal components in PCA. WE: cultivars originating from Western Europe (58), Austria (4), France (23), Germany (12), United Kingdom (15), Switzerland (3), Netherlands (1); SE: Southern Europe and Mediterranean countries (23): Italy (15), Bulgaria (3), Romania (5), Turkey (4), Israel (1); EE: Eastern and Central Europe (25): Hungary (8), Yugoslavia (6), Croatia (1), Czech Republic (1), Kazakhstan (1), Russia (5), Poland (3); AS: Asia and Oceania (14): Australia (8), China (3), Korea (2), New Zealand (1); NA: North America (21): Canada (6), USA (15); SA: South America (8): Argentina (2), Mexico (6)

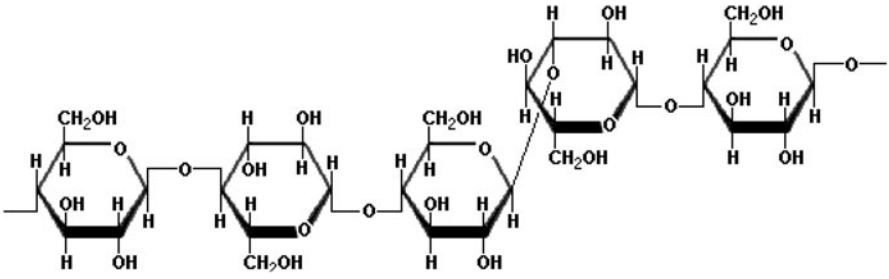

Fig. 19.3 Structure of (1–3,1–4)-β-D-glucan

The proportions of glycan fragments of DP3, DP4 and DP5 and above have been reported to be 70.4:21.3:8.5 (Liu et al. 2006) and 67.1:24.2:8.7 (Lazaridou et al. 2004), with DP3:DP4 molar ratios of 4.38 and 3.66, respectively. However, these studies analysed only single preparations of β-glucan and the extent of genetic variation is not known.

19.4 Genetic Analysis of DF Content

The genetic control of grain AX has been studied using classical genetic analysis and association genetics, focusing on the WE-AX fraction and the viscosity of aqueous extracts (which is mainly determined by WE-AX).

Analyses of several populations of doubled haploid lines, recombinant inbred lines and single seed descent lines allowed the identification of 12 QTL which were reduced to three "meta-QTL" for WE-AX viscosity on chromosomes 1B, 3B and 6B (Martinant et al. 1998; Charmet et al. 2009), with the QTL on 6B accounting for up to 59 % of the variation in AX viscosity in two of the crosses (Charmet et al. 2009). Quraishi et al. (2010) combined this meta-QTL analysis with association genetic analysis of the HEALTHGRAIN diversity collection. The association analysis identified seven loci for AX viscosity, including the three meta-QTL on chromosomes 1B, 3B and 6B and additional loci on chromosomes 3A, 5B, 7A and 7B.

Analyses of these loci should lead to the identification of genes controlling AX synthesis in wheat and Quraishi et al. (2010) have shown that the most significant QTL on chromosome 1B contains four genes which may contribute to AX viscosity. Their identification is also facilitating the development of molecular markers for marker-assisted selection of high (or low) AX viscosity in wheat breeding (Quraishi et al. 2010).

The genes controlling β-glucan synthesis have not been mapped in wheat, but the *CSLF6* gene encoding β-glucan synthase has been mapped to chromosome 7H in barley (Wenzl et al. 2006; Burton et al. 2008). Analysis of a 7H wheat/barley addition line and a 4BS.7HL wheat/barley translocation line have also shown increases in β-glucan content of about two-fold (Cseh et al. 2011). Manickavelu et al. (2011) have recently mapped four QTL for β-glucan content, on chromosomes 3A, 1B, 5B and 6D, in a cross between Chinese Spring and spelt wheat, although the contents of β-glucan in the parental lines did not differ significantly (1.82 and 1.55 %, respectively).

19.5 Effects of Intensive Plant Breeding on DF Components

It has also been suggested that intensive breeding has resulted in decreased health benefits of wheat by focusing on yield and processing quality (Morris and Sands 2003). In particular, it has been suggested that modern wheats have an increased proportion of amylopectin in starch and an increased content of gluten proteins, due to the emphasis on selecting for baking and processing performance (Sands et al. 2009). The HEALTHGRAIN diversity collection provided an ideal opportunity to test this hypothesis, as it was possible to determine registration dates for 146 lines, ranging from 1842 (Red Fife) to 2004. Comparison of the contents of individual DF components (TOT-AX and WE-AX in bran and flour, β-glucan in wholemeal) showed no relationships between registration date and composition (Shewry et al. 2011). This is summarised in Fig. 19.4, in which the values for the five DF components are combined to give a Total Fibre Score. This shows no decline in content

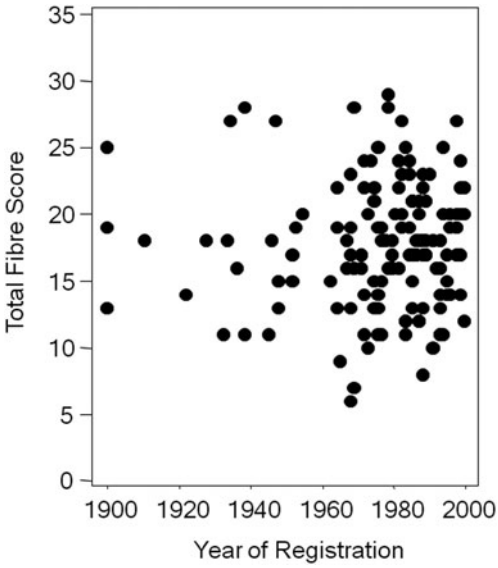

Fig. 19.4 The relationship between the year of registration and the total dietary fibre scores of 146 wheat lines grown in Hungary in 2005. To calculate the total dietary fibre score the lines were grouped into six classes for each dietary fibre component (β-glucan %, flour WE-AX %, flour total AX%, bran WE-AX %, and bran total AX %), with group 6 comprising the 25 lines with the highest contents and group 1 the 25 lines with the lowest contents. Addition of the scores for the individual fibre components means that line with the highest levels of all components would have a maximum score of 30. Two landraces (Chinese Spring and Nap Hal) and Red Fife (released in 1842) are shown with release dates of 1900. (Taken from Shewry et al. (2011))

with registration date, with the range of scores being greater in the more recent lines. However, this greater range may relate to the fact that larger numbers of modern lines were analysed than older lines.

19.6 Genetic Engineering of DF Content and Composition

The existence of extensive heritable genetic variation in the content and composition of AX in wheat means that it should be possible for breeders to select for changes in amount and composition. However, genetic manipulation provides an opportunity to make more substantial changes which are outside the current range of variation. For example, the walls of the starchy endosperm cells of barley differ from those of wheat in being rich in β-glucan (about 75 %) and poor in AX (about 20 %) (MacGregor and Fincher 1993). Furthermore, a lower proportion of the β-glucan in wheat flour is soluble in water (about 10–20 %, Nemeth et al. 2010) than in barley grain (about 30–44 %, Izydorczyk et al. 2000). These differences in amount and solubility are of particular importance in relation to the health benefits of β-glucans, with the role of β-glucan in reducing coronary heart disease (CHD) having been accepted by the US FDA for health claims on food products (Anonymous 2008). The limited variation in β-glucan content in wheat therefore means that manipulation to increase

the amount and solubility may be easier to achieve using genetic engineering than using conventional breeding.

Burton et al. (2006) showed that the *CSLF6* gene of rice mediated the synthesis of β-glucan in the cell walls of leaves of transgenic Arabidopsis plants and subsequently showed that over-expression of *CSLF6* in transgenic barley resulted in increases of over 80 % in the content of β-glucan in seeds (Burton et al. 2011). We have similarly shown that down-regulation of *CSLF6* in developing wheat grain using RNAi suppression resulted in a mean decrease in total β-glucan of 42 % in five transgenic lines (Fig. 19.5a), and a mean decrease of 55 % in hot water-extractable β-glucan (Nemeth et al. 2010). The proportion of β-glucan that was extracted in hot water was slightly lower in the transgenic lines (9.71–13.67 %, mean 12.49 %) than in the control lines (13.57–18.24 %, mean 15.78 %) (Fig. 19.4b) and the ratio of DP3 to DP4 fragments released by digestion with lichenase slightly higher (mean 2.58 compared with 2.33). The Calcoflour average M_r of the hot water-extractable β-glucan was also lower in the transgenic lines (36–57 $\times$ 10^4 g/mol compared with 79–85 $\times$ 10^4 g/mol). Although these studies indicate that *CSLF6* encodes a β-glucan synthase enzyme it is not known whether this enzyme synthesises both (1→3) and (1→4) linkages in β-glucan, or if a second enzyme is required.

The synthesis of (1→3) and (1→4) linkages in β-glucan is of particular interest because it has been suggested that the distribution of these linkages and the occurrence of "cellulose-like" regions with (1→4) linkages determines the solubility of β-glucan (Lazaridou and Biliaderis 2007).

There is similar uncertainty about the precise number of enzymes responsible for arabinoxylan synthesis, but we are using a combination of bioinformatics and RNAi suppression in transgenic wheat to identify candidate genes. This has indicated the existence of at least two genes encoding enzymes with xylan synthase activity in the glycosyl transferase GT43 and GT47 families, and several genes encoding enzymes with arabinosyl transferase activity in the GT61 family (Mitchell et al. 2007; Anders et al. 2012; Pellny et al. 2012; Lovegrove et al. 2013).

The feruloylation of wheat AX is also of interest in relation to the effects of diferulate cross-linking on the solubility and viscosity of the AX, and in relation to the delivery of ferulic acid (a phenolic acid with antioxidant properties) into the colon (Vitaglione et al. 2008). Mitchell et al. (2007) predicted that genes in the acyl-CoA transferase superfamily (PF02458) encoded arabinoxylan feruloyl transferases, based on bioinformatic analyses of expression profiles. Piston et al. (2010) subsequently showed that the simultaneous down-regulation of four genes in this family resulted in a modest but significant decrease in the feruloylation of cell wall AX in leaves of rice plants. However, further work is required to confirm their function in cereal grain. Nothing is currently known about the mechanism of formation of diferulate and triferulate cross-links, but this is assumed to occur after deposition in the cell wall and could be catalysed by a peroxidase enzyme.

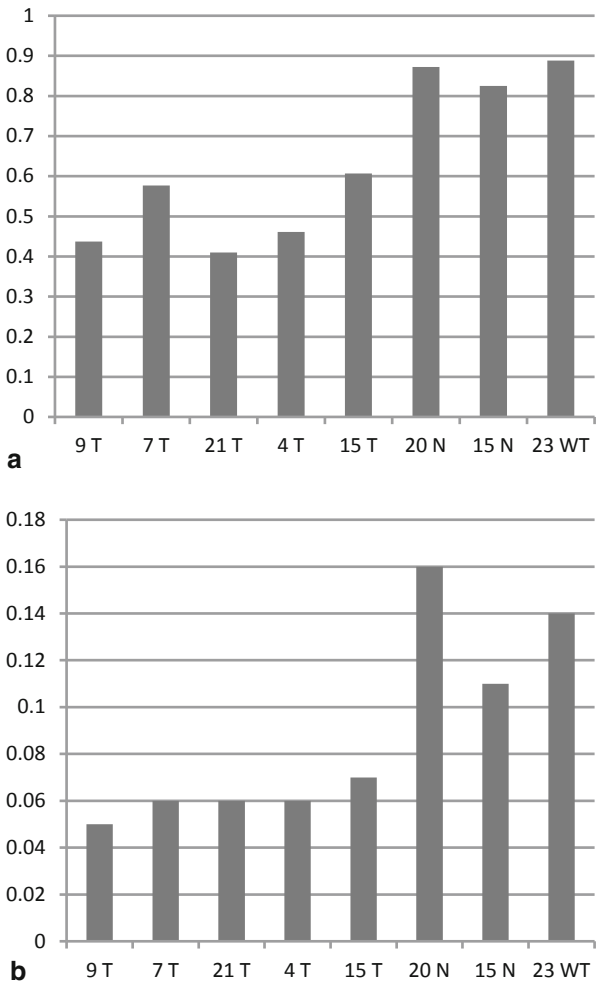

Fig. 19.5 Effect of RNAi suppression of the *CSLF6* gene on the content of: (**a**) total β-glucan (% dry weight) and (**b**) water-extractable β-glucan (% dry weight) in wheat flour. Lines labelled *T* are transgenic, *N* are null segregants and *WT* is untransformed control

19.7 Conclusions

Wheat cell wall polysaccharides play important roles in grain utilisation, whether for food, feed or distilling. The existence of heritable genetic variation in the content and composition of AX and β-glucan will allow the development of new types of wheat by plant breeding. The identification of candidate genes for key enzymes in AX and β-glucan synthesis will allow specific changes in cell wall composition and properties to be made. These lines will provide a basis for determining structure:function relationships and may also be exploited in plant breeding, subject to public acceptance. Alternatively, it should be possible to exploit mutations in key genes, identified using TILLING technology (Slade et al. 2005).

Acknowledgements This publication is financially supported by the European Commission in the Communities Sixth Framework Programme, Project HEALTHGRAIN (FOOD-CT-2005–514008). It reflects the authors' views and the Community is not liable for any use that may be made of the information contained in this publication. Rothamsted Research and the Institute of Food research receive grant-aided support from the Biotechnology and Biological Sciences Research Council of the UK.

References

Anders N, Wilkinson M, Lovegrove A et al (2012) Glycosyl transferases in family 61 mediate arabinofuranosyl transfer onto xylan in grasses. Proc Natl Acad Sci 109:989–993

Anonymous (2008) Final ruleon soluble fiber from certain foods and risk of coronary heart disease (73 FR 47828). US FDA

Bertrais S, Castetbon K, Deheeger M et al (2005) Situation et évolution des apports alimentaires de la population en France, 1997–2003. Saint-Maurice (Fra) Institut de veille sanitaire; (http://opac.invs.sante.fr/index.php?lvl=notice_display&id=5797)

Burton RA, Wilson SM, Hrmova M et al (2006) Cellulose synthase-like *CslF* genes mediate the synthesis of cell wall (1,3;1,4)-β-D glucans. Science 311:1940–1942

Burton RA, Jobling SA, Harvey AJ et al (2008) The genetics and transcriptional profiles of the cellulose synthase-like *HvCslF* gene family in barley. Plant Phys 146:1821–1833

Burton RA, Collins HM, Kibble NAJ et al (2011) Over-expression of specific HvCslF cellulose synthase-like genes in transgenic barley increases the levels of cell wall (1,3;1,4)-β-D-glucans and alters their fine structure. Plant Biotech J 9:117–135

Charmet G, Masood-Quraishi U, Ravel C et al (2009) Genetics of dietary fibre in bread wheat. Euphytica 170:155–168

Courtin CM, Delcour JA (2002) Arabinoxylans and endoxylanases in wheat flour breadmaking. J Cereal Sci 35:225–243

Cseh A, Kruppa K, Molnár I et al (2011) Characterization of a new 4BS.7HL wheat/barley translocation line using GISH, FISH and SSR markers and its effect on the β-glucan content of wheat. Genome 54:795–804

EFSA Panel on Dietetic Products, Nutrition and Allergies (2010) Scientific opinion on dietary reference values for carbohydrates and dietary fibre. EFSA J 8:1462

Gebruers K, Dornez E, Boros D et al (2008) Variation in the content of dietary fibre and components thereof in wheats in the HEALTHGRAIN diversity screen. J Agric Food Chem 56:9740–9749

Goesaert H, Brijs K, Veraverbeke WS et al (2005) Wheat flour constituents: how they impact bread quality, and how to impact their functionality. Trends Food Sci Technol 16:12–30

Izydorczyk MS, Storsley J, Labossiere D et al (2000) Variation in total and soluble β-glucan content in hulless barley: effects of thermal, physical and enzymic treatments. J Agric Food Chem 48:982–989

Lazaridou A, Biliaderis CG (2007) Molecular aspects of cereal β-glucan functionality: physical properties, technological applications and physiological effects. J Cereal Sci 46:101–118

Lazaridou A, Biliaderis CG, Micha-Screttas M, Steele BR (2004) A comparative study on structure-function relations of mixed-linkage (1 → 3),(1 → 4) linear β-D-glucans. Food Hydrocolloid 18:837–855

Li S, Morris CF, Bettge AD (2009) Genotype and environment variation for arabinoxylans in hard winter and spring wheats of the UK Pacific Northwest. Cereal Chem 86:88–95

Liu W, Cui SW, Kakuda Y (2006) Extraction, fractionation, structural and physical characterization of wheat β-D-glucans. Carbohyd Polym 63:408–416

Lovegrove A, Wilkinson MD, Freeman J, et al (2013) RNA interference suppression of genes in glycosyl transferase families 43 and 47 in wheat starchy endosperm causes large decreases in arabinoxylan content. Plant Physiol 163:95–107

MacGregor AW, Fincher GB (1993) Carbohydrates of the barley grain. In: MacGregor AW, Bhatty RS (eds) Barley: chemistry and technology. AACC, St Paul, MN

Manickavelu A, Kawaura K, Imamura H et al (2011) Molecular mapping of quantitative trait loci for domestication traits and β-glucan content in a wheat recombinant inbred line population. Euphytica 177:179–190

Martinant JP, Cadelen T, Billot A, Chartier S (1998) Genetic analysis of water-extractable arabinoxylans in bread wheat endosperm. Theor Appl Genet 97:1069–1075

Mitchell RAC, Dupree P, Shewry PR (2007) A novel bioinformatics approach identifies candidate genes for the synthesis and feruloylation of arabinoxylan. Bioinformatics 144:43–53

Morris CE, Sands DC (2003) The breeder's dilemma—yield or nutrition? Nature Biotechnol 24:1078–1080

Nemeth C, Freeman J, Jones HD et al (2010) Down-regulation of the *CSLF6* gene results in decreased (1,3;1,4)-β-D glucan in endosperm of wheat. Plant Physiol 152:1209–1218

NHMRC (2006) Nutrient reference values for Australia and New Zealand. The National Health and Medical Research Council, Canberra BC 2610, Commonwealth of Australia, pp 317

Nugent AP (2004) The metabolic syndrome. Nutr Bull 29:36–43

Pellny TK, Lovegrove A, Freeman J et al (2012) Cell walls of developing wheat (*Triticum aestivum* L.) endosperm: comparison of composition and RNA Seq transcriptome. Plant Phys 158:612–627

Pettersson D, Åman P (1989) Enzyme supplementation of a poultry diet containing rye and wheat. Br J Nutr 62:139–149

Piston F, Uauy C, Fu L et al (2010) Down-regulation of four putative arabinoxylan fruloyl transferase genes from family PF02458 reduces ester-linked ferulate content in rice cell walls. Planta 231:677–691

Poutanen K, Shepherd R, Shewry PR et al (2008) Beyond whole grain: the European HEALTHGRAIN project aims at healthier cereal foods. CFW 53:32–35

Poutanen K, Shepherd R, Shewry PR et al (2010) More of the grain - Progress in the HEALTHGRAIN project for healthy cereal foods. CFW 55:79–84

Quraishi U-M, Murat F, Abrouk M et al (2010) Combined meta-genomics analysis unravel candidate genes for the grain dietary fibre content in bread wheat (*Triticum aestivum* L.). Funct Integ Genom 11:71–83

Sands DC, Morris CE, Dratz EA, Pilgeram AL (2009) Elevating optimal human nutrition to a central goal of plant breeding and production of plant-based foods. Plant Sci 177:377–389

Saulnier L, Sado P-E, Branlard G et al (2007) Wheat arabinoxylans: exploiting variation in amount and composition to develop enhanced varieties. J Cereal Sci 46:261–281

Shewry PR, Piironen V, Lampi A-M et al (2010a) The HEALTHGRAIN wheat diversity screen: effects of genotype and environment on phytochemicals and dietary fiber components. J Agric Food Chem 58:9291–9298

Shewry PR, Saulnier L, Guillon F et al (2010b) Improving the benefits of wheat as a source of dietary fibre. In: van der Kamp JW, Jones J, McCleary B, Topping D (eds) Dietary fibre: new frontiers for food and health. Wageningen Academic Publishers, Wageningen

Shewry PR, Gebruers K, Andersson AAM et al (2011) Analysis of wheat lines from the HEALTHGRAIN diversity screen shows that intensive plant breeding has not resulted in decreased contents of bioactive components in grain. J Agric Food Chem 59:928–933

Slade AJ, Fuerstenberg SI, Loeffler D et al (2005) A reverse genetic, nontransgenic approach to wheat crop improvement by TILLING. Nat Biotechnol 23:75–81

Steer T, Thane C, Stephen A, Jebb S (2008) Bread in the diet: consumption and contribution to nutrient intakes of British adults. Proc Nutr Soc 67:E363

Toole GA, Le Gall G, Colquhoun IJ et al (2011) Spectroscopic analysis of diversity of arabinoxylan structures in endosperm cell walls of wheat cultivars (*Triticum aestivum*) in the HEALTHGRAIN diversity collection. J Agric Food Chem 59:7075–7082

Vitaglione P, Napolitano A, Fogliano V (2008) Cereal dietary fibre: a natural functional ingredient to deliver phenolic compounds into the gut. Trends Food Sci Tech 19:451–463

Ward JL, Poutanen K, Gebruers K et al (2008) The HEALTHGRAIN cereal diversity screen: concept, results and prospects. J Agric Food Chem 56:9699–9700

Wenzl P, Li H, Carling J et al (2006) A high-density consensus map of barley linking DArT markers to SSR, RFLP and STS loci and agricultural traits. BMC Genomics 7:206

Chapter 20
TILLING for Improved Starch Composition in Wheat

F. Sestili, E. Botticella and D. Lafiandra

Abstract Starch, the main component of wheat flour, is known to significantly influence the quality of wheat-based food products. In the last 20 years, research efforts have enabled the development of a number of wheat lines differing in starch composition and characterized by new chemical physical properties potentially able to confer new added value to food products. Scientists have focused on the opportunity to modify starch composition by targeting the main actors of its biosynthetic pathway: switching off the various starch synthetic enzymes has allowed for the production of a set of wheat starches with an amylose content ranging from 0 up to 75 %. Actually, amylose/amylopectin (AM/AP) ratio is considered the main factor affecting starch properties. Low amylose wheat is currently being investigated for its potential to improve the shelf life of baked products, frozen quality and the texture of noodles. High amylose wheat is of great interest for its healthy and nutritional properties comparable to those of a functional dietary fiber. In this context, reverse-genetics approaches based on the discovery and investigation of new allelic variants are becoming increasingly important. In particular, TILLING (Targeting Induced Local Lesions IN Genomes) has been widely adopted in wheat as well as in other important crops. Herewith, we review how TILLING is being exploited to improve starch composition in wheat.

D. Lafiandra (✉) · F. Sestili · E. Botticella
Department of Agriculture, Forests, Nature and Energy,
University of Tuscia, Viterbo, Italy
e-mail: lafiandr@unitus.it

F. Sestili
e-mail: francescosestili@unitus.it

E. Botticella
e-mail: e.botticella@unitus.it

R. Tuberosa et al. (eds.), *Genomics of Plant Genetic Resources,*
DOI 10.1007/978-94-007-7575-6_20,

20.1 Introduction

Wheat is one of most important cereal crops cultivated all over the world with a total harvest of about 685 million tons in 2009 (http://faostat.fao.org). The peculiarities that make wheat unique in comparison to the other cereals are both its agronomic and qualitative characteristics. The adaptability to critical climatic and environmental conditions and the high yields are also key factors of the success of wheat in the world. Wheat yields can exceed 10 t/ha in optimal conditions, comparing well with other cereals (Shewry et al. 2009). In addition, wheat has a good nutritional profile and is an optimal source of carbohydrates, amino acids, minerals, vitamins and phytochemicals. However, the most important peculiarity of wheat in comparison to other cereal crops is due to the unique viscoelastic properties of its dough that results in the possibility of producing a vast range of end products such as bread, pasta, noodles, snacks and different types of baked foods.

While the role of the proteins present in the wheat kernel in influencing the quality of certain end products is well established, less is known on starch, which makes up from 70 to 80 % of the endosperm. Therefore, a great attention is being addressed to the opportunity deriving from the manipulation of starch composition. In addition, starch is a renewable material and has important applications in several industrial sectors with a high impact on the world economy. Currently, a perspective of great interest is the use of starch as a raw material for the production of alternative sustainable energy and biodegradable polymers.

Wheat starch is the major component of several common foods, playing an important role for human diet as source of carbohydrates. Consequently, intense research activity has been developed with the aim of modulating the chemical composition of the main component of wheat endosperm that can deeply influence the quality of the derived end products and their utilization in food industries.

20.2 Chemical Structure and Composition of Starch

In cereals, reserve starch is produced during the development and maturation of the endosperm and is used during plant germination as source of carbon and energy. The highest level of starch structural organization is represented by the granule, a semicrystalline structure characterized by a large variation in the size and shape in different crop species. Chemically, starch is composed of two polymers of D-glucose known as amylose and amylopectin that differ in their degree of polymerization (DP) and frequency of ramification. In general, all starch polymers are made up by D-glucose molecules linked through α-1,4 bonds resulting in linear chains whereas the points of ramification are introduced on linear polymer by α-1,6 linkages. Amylose is a linear polymer with low DP ($< 3{,}000$) and a frequency of branching inferior to 1 %. Amylopectin has a higher DP ($> 5{,}000$) and frequency of ramification (3–4 %). Several studies have investigated the variability of the content of amylose and amylopectin in wheat, including cultivated, wild and progenitor species, and demonstrated that the amount of amylose ranges between 18 and 35 % of total starch (amylopectin 65–82 %) (Zeng et al. 1997; Mohammadkhani et al. 1998; Stoddard and Sarker 2000).

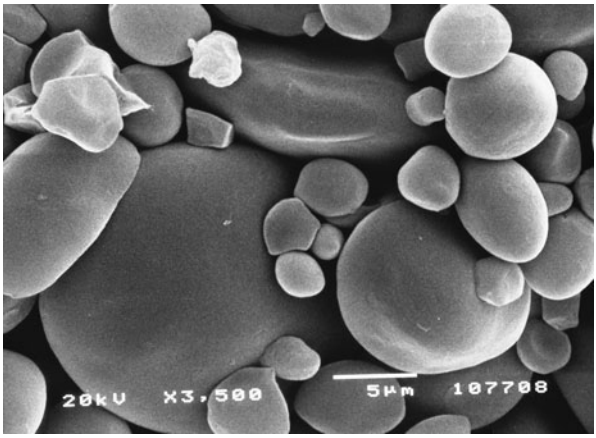

Fig. 20.1 Starch granule isolated from the Durum wheat cultivar Svevo

20.3 The Starch Granule

The members of the *Triticeae* tribe, wheat, barley and rye, have a bimodal size distribution of their granules with the larger ellipsoidal A-type granules ranging from 10–35 μm in diameter and the smaller spherical B-type granules between 1–10 μm in diameter (Fig. 20.1). The A-type granules are synthesized during the first part of the starch accumulation period whereas the B-type granules are initiated and synthesized from 15 days post-anthesis (dpa) and until maturity (Parker 1985). In wheat, A-type granules are only 2–4 % of total granule stored in mature endosperm, but correspond to about 75 % of the starch weight. Numerous studies have demonstrated the influence of granule size and distribution on the functionality and the pasting properties of starch (Chiotelli and Le Meste 2002; Sahlstrom et al. 2003; Stoddard 2003; Park et al. 2004; Ao and Jane 2007; Liu et al. 2007). Sahlstrom et al. (1998) demonstrated that B-type granules have a higher ability of water retention than A-type granules, causing a minor bread weight loss during baking. On the contrary, a high percentage of B-type granules determines the worsening of dough visco-elastic properties due to the subtraction of water from the gluten fraction. In durum wheat, Soh et al. (2006) showed that dough enriched with B granules produced an improvement of visco-elastic characteristics, increase in pasta firmness and a slight reduction in stickiness. Furthermore, the inclusion of B-type granules in the gluten matrix was correlated with an improvement of pasta firmness (Soh et al. 2006).

The increase of A-type granules represents a desirable trait in several applications such as the industrial production of starch and gluten. In fact B-type granules, being small in size, are lost during starch extraction and cause a marked decrease of yield; in addition in the process of gluten extraction, the entrapping of small starch granules in the matrix worsens the purity of the final product (Stoddard 2003).

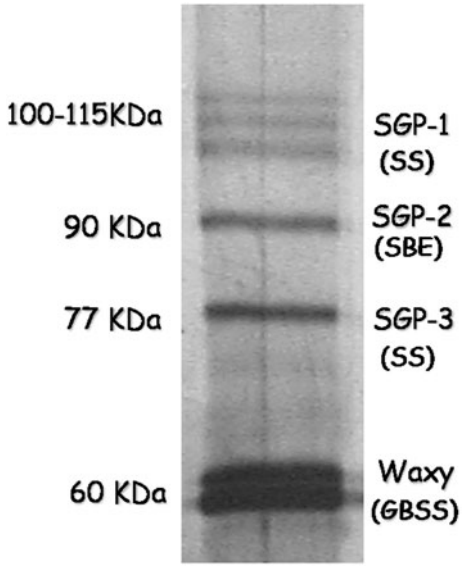

Fig. 20.2 SDS-PAGE separation of starch granule proteins extracted from the bread wheat cultivar Cadenza

20.4 Components Associated to Starch Granules

Lipids are the most important non-starch components located in the granule. They are mainly concentrated on the granule surface (Baldwin et al. 2001) and influence the enzymatic interactions and water absorption of starch granules. In particular, lipids interact with amylose chains promoting the formation of crystalline structures called V crystallites that influence starch functionality by decreasing its susceptibility to enzymatic attack and its ability to interact with water. Starch granules include intrinsic proteins that are embedded in the starch matrix and proteins associated with the granule surface (Kasarda et al. 2008). The intrinsic proteins are enzymes involved in starch synthesis and are known as starch granule proteins (SGP). The most abundant SGPs are the product of the waxy genes that have a molecular weight of 58–60 kDa (Ainsworth et al.1993; Nakamura et al. 1993; Rahman et al. 1995) and play a crucial role in amylose synthesis (Preiss 1991). In bread wheat, Wx-A1, Wx-D1 and Wx-B1 (Fig. 20.2) are three waxy proteins whose encoding genes are located on chromosome arms 7AS (Wx-A1), 7DS (Wx-D1) and 4AL (Wx-B1). The latter was originally located on chromosome 7BS before a translocation occurred between chromosomes 7BS and 4AL during wheat evolution (Miura et al. 1994; Yamamori et al. 1994). In Durum wheat, only the Wx-A1 and Wx-B1 proteins are present.

The SGPs with a molecular weight of 77 and 90 kDa (Fig. 20.2) resulted to be starch synthases by amino terminal sequencing, enzymatic assay and reactions with monoclonal antibodies. Yamamori and Endo (1996) have demonstrated that the three proteins of 100, 108 and 115 kDa (called Sgp-1) correspond to the three isoforms of the class IIa starch synthase (SSIIa). Li et al. (1999) have observed the presence of these proteins in the soluble fraction during the early stages of endosperm

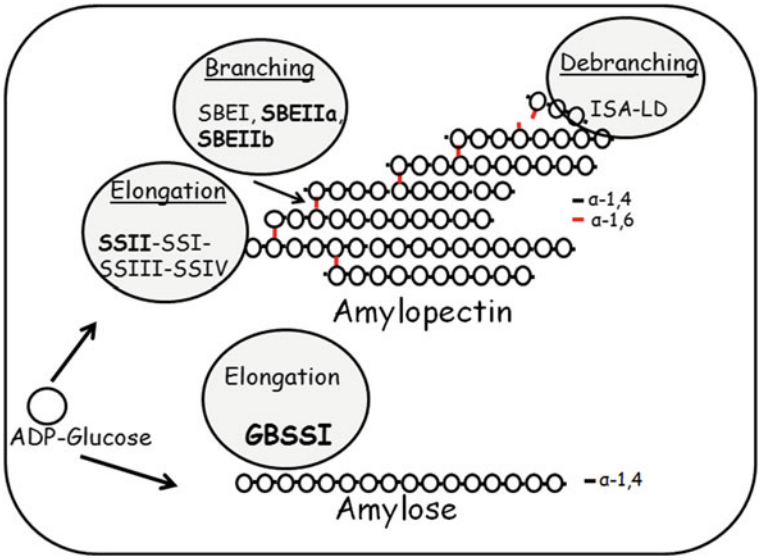

Fig. 20.3 Biosynthesis of amylose and amylopectin

development. The protein Sgp-2 (85 kDa) shows a high homology with the class IIb starch branching enzyme (SBEIIb) of maize (Rahman et al. 1995; Bàga et al. 1999). The proteins Sgp-3 (75 kDa) co-migrate on SDS-PAGE gel and correspond to the three isoforms of the class I starch synthase (SSI) (Yamamori and Endo 1996). Peng et al. (2000) have reported the presence of two high molecular weight SGPs called SGP-140 and SGP-145 preferentially associated with A-type starch granules and determined to be the class I starch branching isoform (SBEI).

20.5 Starch Biosynthesis

Starch biosynthesis has been well established and reviewed in several papers (Preiss 1988; Ball and Morell 2003; James et al. 2003; Geigenberger et al. 2004; Stitt et al. 2010; Vriet et al. 2010; Zeeman et al. 2010). The primary starch is one of the main products of photosynthesis in the leaves. During the night, it is converted into sucrose and used as source for starch biosynthesis in sink organs. In cereals, the reserve starch is produced and accumulated into the amyloplasts of the endosperm. Amylose and amylopectin are synthesized by two different pathways having as common substrate the ADP-glucose. The following classes of enzymes are involved: ADP-glucose pyrophosphorilases (AGPase), starch synthases (SS), branching and debranching enzymes (BE and DBE) (Fig. 20.3).

It is generally accepted that the first step in starch biosynthesis is catalyzed by AGPase, a tetrameric enzyme that consists of two large and two small subunits

(Preiss and Sivak 1996). This reaction takes part predominantly in the cytosol of cereal endosperm and converts D-Glucose in ADP-Glucose that is transported into the amyloplast by specific carriers. Mutations in AGPase subunits and overexpression experiments in barley, maize and wheat have demonstrated an important role of this enzyme for crop yields (Hannah and Greene 1998; Smidansky et al. 2002; Patron et al. 2004; Li et al. 2011). The absence of AGPase drastically reduces the amount of starch; conversely, the enhancement of ADP-glucose pyrophosphorylase activity significantly increases grain yield.

20.6 Amylose Synthesis

The Granule Bound Starch Synthase (GBSS), better known as waxy protein, is the enzyme responsible for amylose synthesis. In cereals, the adjective "waxy" is used to indicate the aspect of the caryopsis lacking amylose.

By crossing the waxy *null* wheat lines identified in natural populations, a set of partial and complete waxy genotypes has been produced in bread and durum wheat (Nakamura et al. 1995; Lafiandra et al. 2010). The complete waxy lines are characterized by starch with a very low amylose content (0–2 %) and peculiar functionality.

20.7 Amylopectin Synthesis

Amylopectin biosynthesis involves several classes of starch synthases, branching and debranching enzymes. At least four classes of SS (SSI, SSII, SSIII and SSIV) take part in amylopectin synthesis in the endosperm. The identification and the study of SSI defective mutants in *Arabidopsis* and rice have pointed out a role of this enzyme in the synthesis of short chains with DP 8–12 (Delvalle et al. 2005; Fujita et al. 2006).

The SSIIa (or Sgp-1) enzymes of wheat play a crucial role in starch biosynthesis and are located both in the soluble fraction and embedded in the granule (Li et al. 1999). In bread wheat the three Sgp-1 isoforms (Sgp-A1, Sgp-B1 and Sgp-D1) are encoded by three genes located on the short arm of each homoeologous chromosome of group 7 (Li et al. 1999). Genotypes lacking one of three isoforms have been identified in wheat (Yamamori 1998). Complete Sgp-1 null mutants have been produced by crossing single *null* genotypes in bread and durum wheat (Yamamori et al. 2000; Lafiandra et al. 2010). These materials showed an amylose content remarkably increased (35 and 42 %, in bread and durum wheat, respectively) coupled with unique viscoelastic properties and also a drastic reduction in total starch and grain yield.

The genes encoding SSIII enzymes have been isolated and localized on chromosome 1 in wheat (Li et al. 2000), but mutants in these genes have not been identified. Functional studies of SSIII enzymes in other species have highlighted their possible role in the synthesis of amylopectin chains of 8–50 DP (Edward et al. 1999).

SSIV enzymes are also encoded by genes located on chromosomes of group 1 in wheat (Leterrier et al. 2008). Their function in cereals is not completely understood, because of failure to identify mutants. Study of two mutants in SSIV gene of Arabidopsis led to hypothesize their role in the control of the correct number of granules in the chloroplast or their involvement in the priming of amylopectin semicrystalline structure (Roldàn et al. 2006).

The starch branching enzymes (SBEs) are transglycosylases that split an α-1,4 glucosidic linkage in α-1,6 with the formation of a ramification point on starch chains. These isoforms have been classified into two different classes, SBEI and SBEII, based on biochemical and genetic evidence (Burton et al. 1995). SBEI-defective mutants have been identified and characterized in several species (maize, rice, potato and wheat) and no significant effects on starch structure and functionality have been detected suggesting a minor role in amylopectin biosynthesis for these enzymes (Jobling et al. 1999; Blauth et al. 2002; Satoh et al. 2003; Regina et al. 2004). In wheat starch, branching enzymes of class II, SBEIIa and SBEIIb, are encoded by different genes present on the long arm of the homoeologous group 2 chromosomes (Regina et al. 2005). RNA interference experiments have permitted to investigate their function in starch synthesis in bread and durum wheat (Regina et al. 2006; Sestili et al. 2010). The silencing of SBEIIa genes produces a phenotype with a strong increase of amylose content ($> 70\,\%$ of total starch) and drastically modifies the granule morphology. Still in wheat, the RNAi suppression of SBEIIb did not cause visible alterations in starch structure (Regina et al. 2006), differently from what has been observed in other species such as maize (Garwood et al. 1976) and rice (Mizuno et al. 1993) where the reduction of SBEIIb activity resulted in a strong increase of amylose content.

Two groups of debranching enzymes (DBEs) have been characterized in higher plants known as isomylases (ISA1, ISA2 and ISA3) and pullulanases (PUL or LD, limit dextrinase). Both types hydrolyze α-1,6 linkages of amylopectin but differ in the specificity for substrate. Though the knowledge of DBE function in wheat has been not clearly determined, the important role of the isoamylase enzymes in starch biosynthesis has been confirmed in several mutants of other species, such as maize, Chlamidomonas, rice, Arabidopsis and barley (James et al. 1995; Mouille et al. 1996; Zeeman et al. 1998; Kubo et al. 1999; Burton et al. 2002; Fujita et al. 2003). The absence of isoamylase activity is associated with lower starch content, formation of a high branched polymer (phytoglycogen) and changes in starch granule distribution and number. In particular, a strong increase of the total number of granules has been observed, with irregular shape and intermediate size in comparison to A/B granules normally present in the *Triticeae*.

In barley, Stahl et al. (2004) increased LD activity by silencing an inhibitor of LD enzymes (LDI), showing that these isoforms play a possible role in the formation of A-type granules.

Amylose	Nutrition and Health	Industrial Material	Food Quality
High	Low Glycemic Index Obesity, CDV, Diabetes	Colon Delivery of drugs	Additive for crispness
	Prebiotic Function Colon diseases	Biodegradable plastics	Major firmness of pasta
Low	Animal Feeding	Paper Textile Adhesives	Improved shelf life Increased frozen stability

Fig. 20.4 Role of high- and low-amylose starch on nutrition, quality and non-food uses

20.8 Novel Starches: Applications in Food and Non-Food Industries

Traditionally, chemical treatments are necessary to obtain desired starch functionality but, in recent years, genetic crop improvement has being applied in the production of novel starch variants as it presents several advantages both in terms of cost savings and pollution management. The relationship between starch chemical–physical structure and its functional properties has been established in cereals identifying as main parameters amylose/amylopectin ratio, type of granule population (size and number), and, as minor ones, glucan chain lengths in amylopectin and starch associated components, notably lipids. Manipulation of amylose/amylopectin ratio, obtained targeting starch biosynthetic enzymes, has produced durum and bread wheat lines with an amylose content ranging from 0 to 70 % of the total starch. The possible uses of high and low amylose wheat are discussed below and schematized in Fig. 20.4.

20.9 Waxy Wheats

Wheat starches with amylose content between 0 and 20 % are defined as low amylose and present peculiar characteristics that make them suitable for several uses in the food industry. Low amylose wheat, obtained by reducing the activity of waxy genes, have a high capacity of water absorption due to the branched structure of amylopectin and different water retention properties due to the low grade of amylopectin retrogradation. In fact, retrogradation is a physical process essentially involving the amylose linear chains that during cooling reassociate, thus releasing water previously absorbed. Loss of water following retrogradation in wheat-based food products is responsible for important food quality parameters; the use of waxy wheat flour permits

to modulate the rate and the extent of starch retrogradation thus increasing shelf life of baked products, retarding bread staling and improving the stability of frozen foods (Yamamori et al. 1992; Miura et al.1994; Ellis et al. 1998; Bhattacharya et al. 2002).

Morita et al. (2002) reported that bread obtained from flours containing 40 % waxy wheat presents an increased retention of water in the crumb resulting in an increase of the product shelf life. Moreover, loaf size made from 40–100 % waxy flour increased in both bread and durum wheat (Morita et al. 2002; Jonnala et al. 2010) that is a desired quality trait in bread making.

Waxy wheat flour has also been found to improve the texture of noodles, common Chinese spaghetti (Yamamori et al. 1992; Miura et al. 1994; Zhao et al. 1998; Yasui et al. 1999; Abdel-Aal et al. 2002; Park and Baik 2004).

On the contrary, the use of waxy durum wheat was deleterious for the quality of pasta, resulting in a worsening of its texture; in addition, pasta cooked faster, absorbed more water and become sticky compared with the one made from non waxy durum wheat cultivars. Moreover, the lack of interactions between amylose and proteins caused an elevated cooking loss (Gianibelli et al. 2005; Vignaux et al. 2005).

20.10 High Amylose Wheat

High amylose wheat has become particularly interesting for its novel nutritional and healthy properties. The current diffusion of important diet-related diseases, such as type II diabetes, cardiovascular diseases (CVD) and colon cancer (Mascie Taylor and Karim 2003; Jemal et al. 2005) indicates the value to manipulate the rate and the extent of starch digestibility. Cereal starch is normally degraded to glucose in the large bowel by α-amylases and amyloglucosidase which target the α-1,4 and α-1,6 glucosidic bonds, respectively, nonetheless, a variable proportion escapes the enzymatic degradation to glucose in the small intestine and acts as a substrate of the intestinal bacteria. The interest of food scientists has focused on the possibility of increasing the quantity of starch resistant to the digestion termed “resistant starch” that is considered functionally similar to the “dietary fibers”. High amylose starches produced, in both bread and durum wheat, have been associated to a substantial increase in the content of “resistant starch”: the resistance of the carbohydrate to the enzymatic attack has been attributed to the ability of long amylose glucan chains to re-associate creating strong double helix structures. This behavior that makes this fraction functionally similar to dietary fibers, as mentioned above, provides several overall benefits for colon health. The fermentation of RS in the large bowel produces small molecules known as short chain fatty acids (SCFAs) representing important metabolites for viscera that contribute to their energetic needs (Topping 2007). SCFAs are responsible for the lowering of pH that contributes to manage the proliferation of pathogen bacteria and induces the ionization of cytotoxic compounds such as ammonium inhibiting their adsorption from the organism. While most of these acids make this general function, one of these, butyrate, has been associated to unique benefits such as the ability to promote a normal phenotype in colonocytes repairing damaged DNA (Le Leu et al. 2005; Toden et al. 2007) and inducing death in

transformed cells (Candido et al. 1978). These mechanisms are believed to be effective in the prevention of serious diseases such as colon-rectal cancer. RS metabolism in the large bowel is beneficial because it is associated with the production of higher levels of butyrate compared to dietary fibers (Topping and Clifton 2001). A further advantage of RS is represented by the ability to lower the glycaemic index of the foods in which is present: foods rich in resistant starch releases less glucose in the blood and more slowly (Nugent 2005). This makes RS able to prevent important disturbs such as obesity and diabetes. Furthermore, resistant starch is considered a prebiotic substrate for intestinal bacteria and it has been shown to promote the growth of lactic bacteria coming from food assumption (probiotics) (Brown et al. 1998).

Although the most important feature of high amylose wheats is represented by its healthy function it is also known to confer improved quality features to several food products. High amylose starch has desirable properties of low water binding capacity, small particle size and bland flavor, providing good impact on dough rheological characteristics and product palatability (Sharma et al. 2008). Furthermore the use of whole grains, as source of dietary fibre, can negatively affect dough texture, that results stickier due to the high water absorption power associate to bran; on the contrary the incorporation of commercial high amylose starch (40 %) does not alter the water absorption power of the flour and, consequently, bread quality results unaffected (Waring 2005). Incorporation studies of commercial high amylose starch in baked products such as cakes, muffins and biscuits have shown a significant increase in total dietary fiber content and, at the same time, an improvement in the texture that resulted softer (Waring 2005). Chanvrier et al. (2007) have analyzed the functional properties of SGP-1 *null* wheat flour finding an increase in resistant starch content and acceptable sensorial quality of the bread made up by 50 % Sgp-1 flour. In a recent work, Hallström et al. (2011) have reported an increased RS formation in bread based on whole wheat grain with elevated amylose content (38 %) compared to whole wheat with normal amylose content; the increase of resistant starch was found to correlate with the lowering of postprandial glycaemic response.

High amylose starch has also been recognized as texture improver in baked goods, firmness enhancer in pasta and as an additive to increase crispness of foods (Waring 2005; Sajilata et al. 2006; Soh et al. 2006; Chanvrier et al. 2007).

In the pharmaceutical sector, high amylose starch has resulted suitable for the production of films coating colon specific drugs (Freire et al. 2009). In fact, the properties of resistance of high amylose starch can guarantee the survival of the drug transiting the intestine and improve the action of the drug. A new interesting perspective is represented by the use of high amylose starches for the production of bioplastics with improved performances (Avella et al. 2002).

20.11 TILLING as Tool for Genetic Improvement

Classical mutagenesis has been widely used in crop breeding over the past 60 years and is lately emerging as an efficient alternative to exploit and modify functionality of genes controlling important traits in crops. Chemical mutagenic treatment provides

a valid tool to generate high density mutations in the genome of a target organism, though in polyploids the presence of multiple copies of a gene has represented a major limit for the detection of interesting phenotype for valuable traits by a forward genetics approach.

In the last 10 years the advances made in genome sequencing, the availability of gene sequences and the development of new techniques have remarkably increased the use of reverse genetics approaches to investigate gene function, to associate a phenotype to a known sequence and to improve quality traits in the species of interest. TILLING (Targeting Induced Local Lesions IN Genomes) is a high-throughput reverse genetics technique that combines the traditional chemical mutagenesis methods and PCR-based screening to indentify desired mutations in the genes of interest (McCallum et al. 2000).

Currently several TILLING platforms have been produced in bread and durum wheat and more recently also in the diploid species *T. monococcum* (Table 20.1), thus providing available resources for gene functional studies and the possibility to detect novel genetic variation to be used in crop breeding programs.

A further advantage of TILLING in crop improvement is that the genotypes produced are not considered genetically modified (GM). In fact, GM application is still not completely accepted, encountering resistance from the general public and governments and still presents limits in terms of stability and efficiency of the transgene. On the contrary TILLING relies on point mutations and results precise and efficient in the targeting of a specific gene of interest.

20.12 The TILLING Methodology

The schematic representation of the TILLING strategy is reported in Fig. 20.5. The first step of TILLING is the mutagen treatment of thousands of seeds. Compounds such as ethane methane sulfonate (EMS) and sodium azide (NaN3) have been chosen for many TILLING projects because induce single nucleotide changes by nucleotide alkylation, resulting in a high density of mutations randomly distributed through the genome (Greene et al. 2003). In order to ensure that any gene of interest carries out significant mutations, it is necessary to produce a mutagenized population of appropriate size depending on the dosage of mutagen used and the species targeted. The degree of embryo lethality is strongly related to the ploidy of the organism targeted: diploid organisms have less tolerance to the chemical treatment than polyploid species and it is necessary to increase the size of mutagenized population for an efficient screening. By contrast polyploid organisms have a higher tolerance of mutations due to the presence of multiple homeoalleles for each gene that complement the absence of mutated one. The use of a higher dosage of mutagen in polyploids lets down the number of mutants required for the achievement of a significant coverage and the costs of screening for mutant identification.

Table 20.1 TILLING platforms in wheat

Species	Population size	Mutation frequency (kb)	Targeted genes	SNP detection method	Reference
T. aestivum cv Express	10,000	1/24	GBSSI	Licor	Slade et al. (2005)
T. durum cv Kronos	8,000	1/40	GBSSI	Licor	Slade et al. (2005)
T. aestivum cv QAL2000	869	1/29	GBSSI, Pina, Pinb	Cel1 digestion and agarose gel	Dong et al. (2009b)
T. aestivum cv Ventura	1,168	1/33	GBSSI	Cel1 digestion and agarose gel	Dong et al. (2009b)
T. durum cv Cham1	4,500	1/52	RCA-A1	Licor	Parry et al. (2009)
T. aestivum cv Cadenza	4,500	1/40	SSIIa, SBEIIa	Licor HRM	Sestili et al. (2010); Botticella et al. (2011)
T. aestivum UC1041 + *Gpc-B1/Yr36*	1,536	1/38	SBEIIa, WKS	Cel1 digestion and agarose gel	Uauy et al. (2009)
T. durum cv Kronos	1,368	1/51	SBEIIa, WKS	Cel1 digestion and agarose gel	Uauy et al. (2009)
T. monococcum	1,532	1/92	PAL, HCT, COMT1, GBSSI	Cel1 digestion and agarose gel	Rawat et al. (2012)
T. durum cv Svevo	4,000	–	SBEIIa	HRM	Bovina et al., 2013

TILLING analysis is carried out on genomic DNA pools of M2 plants. PCR amplified fragments are denatured and re-annealed, in order to favor the production of mismatched heteroduplexes between wild type and mutant. Different high-throughput methodologies are currently available for mutant detection based on the use of endonuclease CEL I or high resolution melting technology (Colbert et al. 2001; Gady et al. 2009). The traditional method for the detection of DNA polymorphisms used in TILLING is a heteroduplex mismatch cleavage assay based on endonuclease CEL I (McCallum et al. 2000). The use of forward and reverse primers, labeled with different fluorophores, allows the products of heteroduplex cleavage to be detected in separate channels in the Li-Cor DNA analyser (LI-COR-4300; LI-COR Biosciences; Lincoln, NE, USA).

Low-cost alternative methods based on CEL I have been developed using classical electrophoresis, making TILLING accessible to any small laboratory (Dong et al. 2009a; Uauy et al. 2009).

High resolution melting ™(HRM) is a scanning mutation technology of next generation derived from the extension of previous DNA melting analysis combined with new generation fluorescent dsDNA dyes (Wittwer et al. 2003).

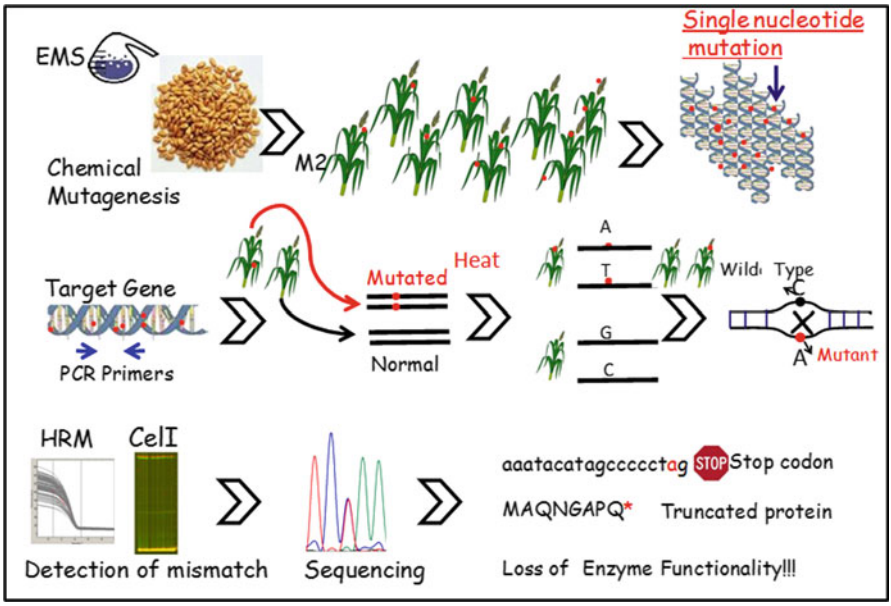

Fig. 20.5 TILLING approach

CEL I-based method permits to analyze amplicons of large dimension (up to 1500 bp) with the disadvantage to detect also mutations in the intronic regions that do not alter protein functionality. HRM-based method is sensitive and specific for the detection of mutations in small DNA fragments (up to 500 bp) (Dong et al. 2009b; Gady et al. 2009) and is advantageous for the screening of genes containing small exons both in terms of costs and time saving compared to CEL I-based method (Hofinger et al. 2009).

For polyploid crop species, such as wheat, the screening is more difficult because polymorphism of homoeologous genes can produce false positive mutants, so homoeologous specific primer pairs must be utilized for mutant detection.

Because it is not easy to produce homeoallele specific amplicons with optimal size for HRM analysis, a strategy of nested PCR could be necessary. In this strategy homoeologous PCR fragments, obtained in the first PCR, are used as template in a 2nd round PCR using primer pairs targeting the exons inside the 1st PCR fragments.

The combination of next-generation sequence (NGS) technology and bioinformatics constitutes a robust and efficient alternative for TILLING application compared to the above cited methods based on the mismatch identification (Varshney et al. 2009; Tsai et al. 2011; Wang et al. 2012). Tsai et al. (2011) developed a method based on Illumina sequencing of target genes amplified from 2D or 3D pooled templates representing 768 individuals per experiment. These authors demonstrated that the advantages of this new approach are both in terms of rapidity and cost. However, it is to be expected that the costs of TILLING will drive down with the development of new technologies, making its uses more and more attractive in the next years.

20.13 Bioinformatics Tools for the Analysis of Nucleotide Polymorphisms

Nucleotide changes produced by EMS treatment can be classified as non-sense (both as changes to a stop codon in coding sequence and as changes in a splice junction at the beginning or end of an intron), missense and silent (in both coding and non-coding sequences) mutations. Non-sense mutations are normally responsible for the loss of gene function because produce truncated proteins or the shifting of reading frame. In some cases, the missense changes can be particularly interesting, as they can have an effect on the activity of the target gene. Several bioinformatics programs offer support for the TILLING analyses. The program CODDLE (www.proweb.org/coddle/) is useful to determinate the region of the target gene where there is a higher probability of finding deleterious mutations (Till et al. 2003). This program, using a software of alignment of nucleotide sequences and the Primer3 algorithm, designs specific primer pairs that can amplify gene regions up to 1.5 kb. The severity of the effect of a missense mutation can be investigated by the web program SIFT, that predicts with approximately 75 % accuracy whether an amino acid change affects the function of a protein of interest (Ng and Henikoff 2001).

PARSESNP (Project Aligned Related Sequences and Evaluate SNPs) (http://www.proweb.org/parsesnp/) is a bioinformatics web platform, that gives information on the mapping of the mutations on genomic sequence, the changes in restriction enzyme recognition sites and the effects of SNPs on the enzyme activity (Taylor and Greene 2003).

20.14 TILLING Application to Modify Starch Composition in Wheat

The potential of the TILLING technology for genetic improvement of wheat starch quality has been shown, for the first time, by Slade et al. (2005), which screened 1,152 individuals of a bread wheat EMS-mutagenized population and 768 lines of a durum wheat TILLING population, identifying 256 new alleles in the waxy genes. The new allelic series detected included non-sense, splice junction as well as missense mutations, demonstrating that the use of chemical mutagens can generate large genetic variation. Slade et al (2005) produced a complete waxy null genotype combining the two null waxy mutations in *Wx-A1* and *Wx-D1* genes identified by TILLING and a preexistent mutation in the third *waxy* homoeolog. The seeds of the novel complete waxy line showed a phenotype characterized by low amylose content. A similar strategy to produce a complete waxy genotype has been undertaken by Dong et al. (2009a), who used a modified TILLING method with unlabeled primers and agarose gels for mutation detection. These authors identified 121 waxy mutants by analyzing two bread wheat populations, the soft wheat line QAL2000 and the hard

wheat cultivar Ventura, both lacking the Wx-B1 protein. Among the 121 novel alleles, knockout (truncation) mutations have been identified in *Wx-A1* and *Wx-D1* genes. The intercrossing of two *null* mutants permitted to obtain complete waxy genotypes. Phenotypic analysis of these mutants showed a deep change of starch structure, characterized by a drastic reduction of amylose content.

With the aim of producing low and high amylose lines, Sestili et al. (2010) screened an EMS-mutagenized population of the bread wheat cultivar Cadenza combining SDS–PAGE analysis of starch granule proteins with a TILLING approach. In particular they focused on *Sgp-1* and *Wx* genes involved, respectively, in amylopectin and amylose biosynthesis and found single null genotypes associated with each of the three homoeologous loci of the target genes.

Another application of TILLING to modify starch composition in wheat has been carried out by Uauy et al. (2009). These authors targeted genes coding starch branching enzymes IIa and IIb through a novel protocol based on CEL-I digestion and non-denaturing polyacrylamide gel detection. The screening of a bread wheat TILLING population permitted to identify 65 mutations on *SBEIIa* alleles. In the tetraploid library, 58 and 35 mutants were detected for *SBEIIa* and *SBEIIb* genes, respectively.

The same genes have been targeted by a TILLING strategy by Botticella et al. (2011) that reported the identification of 123 novel allelic variants for the three homoeologous genes coding SBEIIa in the bread wheat cultivar Cadenza. The crossing of knockout mutants permitted to realize a set of single and double null *SBEIIa* lines. Pyramiding of two null homeologs resulted in an increase of amylose content up to 21 % compared to the control line. Although total starch content decreased slightly, the comparison of 100 seed weight did not highlight significant differences among the single and double *null* genotypes in respect to the control. Isolation of triple *SBEIIa* null lines has also been reported (Lafiandra et al. 2012).

More recently, Hazard et al. (2012) using single gene mutants of *SBEIIa* showed no significant increase in amylose and resistant starch content, but when a double mutant was obtained, combining a *SBEIIa*-A knock-out mutation with a *SBEIIa-B* splice-site mutation, an increase of 22 % in amylose content and 115 % in resistant starch content was observed.

Similarly, Slade et al. (2012) using TILLING to identify novel genetic variation in each of the A and B genomes in tetraploid durum wheat and the A, B and D genomes in hexaploid bread wheat, have identified deleterious mutations in the form of single nucleotide polymorphisms (SNPs) in starch branching enzyme IIa genes (*SBEIIa*). Combination of these new alleles of *SBEIIa* through breeding resulted in the development of high amylose durum and bread wheat lines containing 47–55 % amylose and having elevated resistant starch levels compared to wild-type wheat. High amylose lines also had reduced expression of *SBEIIa*, changes in starch granule morphology and altered starch granule protein profiles as evaluated by mass spectrometry.

Recently, starch metabolism has been modified with success by using the TILLING strategy in barley. In particular, Bovina et al. (2011) have reported the identification of 29 mutations in five starch-related barley genes (*Bmy1*, *GBSSI*, *LDA1*, *SSI* and *SSII*) through the screening of a sodium azide-mutagenized barley

population. Non-sense mutations (two truncation mutations and two splice junction mutations) were identified for *SSI* and *SSII* genes that play an important role in amylopectin biosynthesis. The absence of SSI and SSII proteins is often associated to a reduced content of total starch and amylopectin.

20.15 Future Perspectives for TILLING in Wheat

TILLING is powerful tool to modify amylose/amylopectin ratio in wheat. In the future, it will be of interest to exploit other factors as granule size and distribution, glucan chains length distribution and starch yield.

Currently, the role of a few classes of starch biosynthetic enzymes has been determined. The existence of multiple isoforms for each class of the enzymes involved in starch biosynthesis offers the opportunity to "design" a large variety of novel starches that could satisfy different requirements necessary for different applications. A further opportunity of TILLING could derive from the identification and determination of the role played by key amino acids in the biosynthetic enzymes. The targeting of catalytic or binding sites could permit the modulation of enzyme activity and the realization of novel wheat starches.

A future interesting perspective is the use of TILLING in some crops, such as rice, potato and maize, representing the major sources of starch both for food and non-food applications. For example, the production of high amylose rice could be a means of delivering new healthy benefits to the population of developing countries whose diet is characterized by a high consumption of rice.

Moreover, because of the growing food demand and decreasing available lands, the increase of starch yield is more and more becoming a target trait in crop genetic improvement and TILLING represents a valid tool to realize this aim.

In conclusion, the application of TILLING technology to wheat plants has steadily increased in the last few years and it is predictable that this trend will continue with the completion of wheat genome sequencing, that will provide many potential genes of interest that may be targeted for plant genetic improvement.

References

Abdel-Aal ESM, Hucl P, Chibbar RN et al (2002) Physicochemical and structural characteristics of flours and starches from waxy and nonwaxy wheats. Cereal Chem 79:458–464

Ainsworth C, Clark J, Balsdon J (1993) Expression, organization and structure of the genes encoding the waxy protein (granule-bound starch synthase) in wheat. Plant Mol Biol 22:67–82

Ao Z, Jane JL (2007) Characterization and modeling of the A-and B-granule starches of wheat, triticale, and barley. Carbohyd Polym 67:46–55

Avella M, Errico ME, Rimedio R, Sadocco P (2002) Preparation of biodegradable polyesters/high-amylose-starch composites by reactive blending and their characterization. J Appl Polym Sci 83:1432–1442

Båga M, Glaze S, Mallard CS, Chibbar RN (1999) A starch-branching enzyme gene in wheat produces alternatively spliced transcripts. Plant Mol Biol 40:1019–1030

Baldwin PM (2001) Starch granule-associated proteins and polypeptides: a review. Starch 53: 475–503
Ball SG, Morell MK (2003) From bacterial glycogen to starch: understanding the biogenesis of the plant starch granule. Ann Rev Plant Biol 54:207–233
Bhattacharya M, Erazo-Castrejón SV, Doehlert DC, McMullen MS (2002) Staling of bread as affected by waxy wheat flour blends. Cereal Chem 79:178–182
Blauth SL, Kim KN, Klucinec J et al (2002) Identification of mutator insertional mutants of starch-branching enzyme 1 (sbe1) in *Zea mays* L. Plant Mol Biol 48:287–297
Botticella E, Sestili F, Hernandez-Lopez A et al (2011) High Resolution Melting analysis for the detection of EMS induced mutations in wheat *Sbella* genes. BMC Plant Biol 11:156
Bovina R, Brunazzi A, Gasparini G et al (2013) Development of a TILLING resource in durum wheat for forward- and reverse-genetics analyses. Crop & Pasture Sci. In press
Bovina R, Talamè V, Silvio S et al (2011) Starch metabolism mutants in barley: A TILLING approach. Plant Genet Resour 9:170–173
Brown IL, Wang X, Topping DL et al (1998) High amylose maize starch as a versatile prebiotic for use with probiotic bacteria. Food Aust 50:603–610
Burton RA, Bewley JD, Smith AM et al (1995) Starch branching enzymes belonging to distinct enzyme families are differentially expressed during pea embryo development. Plant J 7:3–15
Burton RA, Jenner H, Carrangis L et al (2002) Starch granule initiation and growth are altered in barley mutants that lack isoamylase activity. Plant J 31:97–112
Candido EPM, Reeves R, Davies JR (1978) Sodium butyrate inhibits histone deacetylation in cultured cells. Cell 14:105–113
Chanvrier H, Appelqvist IAM, Bird AR et al (2007) Processing of novel elevated amylose wheats: functional properties and starch digestibility of extruded products. J Agric Food Chem 55: 10248–10257
Chiotelli E, Le Meste M (2002) Effect of small and large wheat starch granules on thermomechanical behavior of starch. Cereal Chem 79:286–293
Colbert T, Till BJ, Tompa R et al (2001) High-throughput screening for induced point mutations. Plant Physiol 126:480–484
Delvalle D, Dumez S, Wattebled F et al (2005) Soluble starch synthase I: a major determinant for the synthesis of amylopectin in *Arabidopsis thaliana* leaves. Plant J 43:398–412
Dong C, Dalton-Morgan J, Vincent K, Sharp P (2009a) A modified TILLING method for wheat breeding. Plant Gen 2:10–12
Dong C, Vincent K, Sharp S (2009b) Simultaneous mutation detection of three homoeologous genes in wheat by High Resolution Melting analysis and Mutation Surveyor. BMC Plant Biol 9:143
Edwards A, Fulton DC, Hylton CM et al (1999) A combined reduction in activity of starch synthases II and III of potato has novel effects on the starch of tubers. Plant J 17:251–261
Ellis RP, Cochrane MP, Dale MFB et al (1998) Starch production and industrial use. J Sci Food Agr 77:289–311
Freire A, Podczeck F, Veiga F, Sousa J (2009) Starch-based coatings for colon-specific delivery Part II: Physicochemical properties and in vitro drug release from high amylose maize starch films C. Eur J Pharmac Biopharmac 72:587–594
Fujita N, Kubo A, Suh DS et al (2003) Antisense inhibition of isoamylase alters the structure of amylopectin and the physicochemical properties of starch in rice endosperm. Plant Cell Physiol 44:607–618
Fujita N, Yoshida M, Asakura N et al (2006) Function and characterization of starch synthase I using mutants in rice. Plant Physiol 140:1070–1084
Gady ALF, Hermans FWK, Van de Wal MHBJ et al (2009) Implementation of two high throughput techniques in a novel application: detecting point mutations in large EMS mutated plant populations. Plant Methods 5:13
Garwood DL, Shannon J, Creech R (1976) Starches of endosperms possessing different alleles at the amylose-extender locus in *Zea mays* L. Cereal Chem 53:355–364

Geigenberger P, Stitt M, Fernie AR (2004) Metabolic control analysis and regulation of the conversion of sucrose to starch in growing potato tubers. Plant Cell Environ 27:655–673
Gianibelli MC, Sissons MJ, Batey IL (2005) Effect of source and proportion of waxy starches on pasta cooking quality. Cereal Chem 82:321–327
Greene EA, Codomo CA, Taylor NE et al (2003) Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in Arabidopsis. Genetics 164:731–740
Hallström E, Sestili F, Lafiandra D et al (2011) Novel wheat variety with elevated content of amylose beneficially influence glycaemia in healthy subjects. Food Nutr Res 55:7074
Hannah LC, Greene TW (1998) Maize seed weight is dependent on the amount of endosperm ADP-glucose pyrophosphorylase. J Plant Physiol 152:649–652
Hazard B, Zhang X, Colasuonno P et al (2012) Induced Mutations in the starch branching enzyme II (SBEII) genes increase amylose and resistant starch content in durum wheat. Crop Sci 52: 1754–1766
Hofinger BJ, Jing HC, Hammond KE, Kanyuka K (2009) High-resolution melting analysis of cDNA-derived PCR amplicons for rapid and cost-effective identification of novel alleles in barley. Theor Appl Genet 119:851–865
James MG, Robertson DS, Myers AM (1995) Characterization of the maize gene sugary1, a determinant of starch composition in kernels. Plant Cell 7:417–429
James MG, Denyer K, Myers AM (2003) Starch synthesis in the cereal endosperm. Curr Opin Plant Biol 6:215–222
Jemal A, Ward E, Hao Y, Thun M (2005) Trends in the leading causes of death in the United States. JAMA 294:1255–1259
Jobling SA, Schwall GP, Westcott RJ et al (1999) A minor form of starch branching enzyme in potato (*Solanum tuberosum* L) tubers has a major effect on starch structure: cloning and characterisation of multiple forms of SBEA. Plant J 18:163–171
Jonnala RS, MacRitchie F, Smail VW et al (2010) Protein and quality characterization of complete and partial near-isogenic lines of waxy wheat. Cereal Chem 87:538–545
Kasarda DD, Dupont FM, Vensel WH, Altenbach SB et al (2008) Surface-associated proteins of wheat starch granules: suitability of wheat starch for celiac patients. J Agric Food Chem 56:10292–10302
Kubo A, Fujita N, Harada K et al (1999) The starch debranching enzymes isoamylase and pullulanase are both involved in amylopectin biosynthesis in rice endosperm. Plant Physiol 121:399–409
Lafiandra D, Sestili F, D'Ovidio R et al (2010) Approaches for modification of starch composition in Durum wheat. Cereal Chem 87:28–34
Lafiandra D, Sestili F, Botticella E, Phillips A (2012) Improving wheat health benefits through manipulation of starch composition. In: Bedo Z and Làng L (eds) Proceedings of the 19th EUCARPIA General Congress. Agricultural Institute, Centre for Agricultural Research, Hungarian Academy of Sciences, Màrtonvàsar, Hungary Canada pp 180–183
Le Leu RK, Brown IL, Hu Y et al (2005) A symbiotic combination of resistant starch and Bifidobacterium lactis facilitates apoptotic deletion of carcinogen damaged cells in rat colon. J Nutr 135:996–1001
Leterrier M, Holappa LD, Broglie KE, Beckles DM (2008) Cloning, characterisation and comparative analysis of a starch synthase IV gene in wheat: Functional and evolutionary implications. BMC Plant Biol 8:98
Li N, Zhang S, Zhao Y, Li B, Zhang J (2011) Over-expression of AGPase genes enhances seed weight and starch content in transgenic maize. Planta 233:241–250
Li ZY, Chu XS, Mouille G et al (1999) The localization and expression of the class II starch synthases of wheat. Plant Physiol 120:1147–1155
Li ZY, Mouille G, Kosar-Hashemi B et al (2000) The structure and expression of the wheat starch synthase III gene. Motifs in the expressed gene define the lineage of the starch synthase III gene family. Plant Physiol 123:613–624
Liu Q, Gu Z, Donner E et al (2007) Investigation of digestibility in vitro and physicochemical properties of A-and B-type starch from soft and hard wheat flour. Cereal Chem 84:15–21

Mascie-Taylor NCG, Karim E (2003) The burden of chronic disease. Sci 302:1921–1922
McCallum CM, Comai L, Greene EA, Henikoff S (2000) Targeting induced local lesions in genomes (TILLING) for plant functional genomics. Plant Physiol 123:439–442
Miura H, Tanii S, Nakamura T, Watanabe N (1994) Genetic control of amylose content in wheat endosperm starch and differential effects of three Wx genes. Theor Appl Genet 89:276–280
Mizuno K, Kawasaki T, Shimada H et al (1993) Alteration of the structural properties of starch components by the lack of an isoform of starch branching enzyme in rice. J Biol Chem 268:19084–19091
Mohammadkhani A, Stoddard FL, Marshall DR (1998) Survey of amylose content in *Secale cereale*, *Triticum monococcum, T. turgidum* and *T. tauschii*. J Cereal Sci 28:273–280
Morita N, Maeda T, Miyazaki M et al (2002) Dough and baking properties of high-amylose and waxy wheat flours. Cereal Chem 79:491–495
Mouille G, Maddelein ML, Ball S (1996) Preamylopectin processing: a mandatory step for starch biosynthesis in plants. Plant Cell 8:1353–1366
Nakamura T, Yamamori M, Hirano H, Hidaka S (1993) Decrease of waxy (Wx) protein in two common wheat cultivars with low amylose content. Plant Breeding 111:99–105
Nakamura T, Yamamori M, Hirano H et al (1995) Production of waxy (amylose-free) wheats. Mol Gen Genet 248:253–259
Ng PC, Henikoff S (2001) Predicting deleterious amino acid substitutions. Genome Res 11:863–874
Nugent AP (2005) Health properties of resistant starch. Nutr Bulletin 30:27–54
Park CS, Baik BK (2004) Significance of amylose content of wheat starch on processing and textural properties of instant noodles. Cereal Chem 81:521–526
Park SH, Wilson JD, Chung OK, Seib PA (2004) Size distribution and properties of wheat starch granules in relation to crumb grain score of pup-loaf bread. Cereal Chem 81:699–704
Parker ML (1985) The relationship between A-type and B-type starch granules in the developing endosperm of wheat. J Cereal Sci 3:271–278
Parry MAJ, Madgwick PJ, Bayon C et al (2009) Mutation discovery for crop improvement. J Exp Bot 60:2817–2825
Patron NJ, Greber B, Fahy BE et al (2004) The lys5 mutations of barley reveal the nature and importance of plastidial ADP-Glc transporters for starch synthesis in cereal endosperm. Plant Physiol 135:2088–2097
Peng M, Gao M, Baga M et al (2000) Starch-branching enzymes preferentially associated with A-type starch granules in wheat endosperm. Plant Physiol 124:265–272
Preiss J (1991) Biology and molecular biology of starch synthesis and its regulation. In: Miflin BJ (ed) Oxford Survey of Plant Molecular and Cellular Biology, vol 7. Oxford University Press, Oxford, pp. 59–114
Preiss J (1988) Biosynthesis of starch and its regulation. In: Preiss J (ed) The biochemistry of plants, carbohydrates, vol 14. Academic Press, San Diego, pp. 181–254
Preiss J, Sivak M (1996) Starch synthesis in sinks and sources. In: Zamski E, Schaffter AA (eds) Photoassimilate distribution in plants and crops: sink-source relationships. M Dekker, New York, pp. 63–96
Rahman S, Kosar-Hashemi B, Samuel MS et al (1995) The major proteins of wheat endosperm starch granules. Aust J Plant Physiol 22:793–803
Rawat N, Sehqal SK, Rothe N et al (2012) A diploid wheat TILLING resource for wheat functional genomics. BMC Plant Biol 12:205
Regina A, Kosar-Hashemi B, Li Z et al (2004) Multiple isoforms of starch branching enzyme-I in wheat: lack of the major Sbe-I isoform does not alter starch phenotype. Funct Plant Biol 31:591–601
Regina A, Kosar-Hashemi B, Li Z et al (2005) Starch branching enzyme IIb in wheat is expressed at low levels in the endosperm compared to other cereals and encoded at a non-syntenic locus. Planta 222:899–909
Regina A, Bird A, Topping D et al (2006) High-amylose wheats generated by RNA interference improves indices of large-bowel health in rats. Proc Natl Acad Sci USA 103:3546–3551

Roldán I, Wattebled F, Lucas MM et al (2006) The phenotype of soluble starch synthase IV defective mutants of *Arabidopsis thaliana* suggests a novel function of elongation enzymes in the control of starch granule formation. Plant J 49:492–504
Sahlstrom S, Brathen E, Lea P, Autio K (1998) Influence of starch granule size distribution on bread characteristics. J Cereal Sci 28:157–164
Sahlstrom S, Baevre AB, Brathen E (2003) Impact of starch properties on hearth bread characteristics. II. Purified A-and B-granule fractions. J Cereal Sci 37:285–293
Sajilata MG, Singhal RS, Kulkarni PR (2006) Resistant starch: a review. Compr Rev Food Sci Food Safety 5:5–17
Satoh H, Nishi A, Yamashita K et al (2003) Starch-branching enzyme I-deficient mutation specifically affects the structure and properties of starch in rice endosperm. Plant Physiol 133:1111–1121
Sestili F, Botticella E, Bedo Z et al (2010) Production of novel allelic variation for genes involved in starch biosynthesis through mutagenesis. Mol Breed 25:145–154
Sharma A, Yadav BS, Ritika (2008) Resistant starch: physiological roles and food applications. Food Rev Int 24:193–234
Shewry PR (2009) Wheat. J Exp Bot 60:1537–1553
Slade AJ, Fuerstenberg SI, Loeffler D et al (2005) A reverse genetic, nontransgenic approach to wheat crop improvement by TILLING. Nat Biotechnol 23:75–81
Slade AJ, McGuire C, Loeffler D et al (2012) Development of high amylose wheat through TILLING. BMC Plant Biol 12:69
Smidansky ED, Clancy M, Meyer FD et al (2002) Enhanced ADP-glucose pyrophosphorylase activity in wheat endosperm increases seed yield. Proc Natl Acad Sci USA 99:1724–1729
Soh NH, Sissons MJ, Turner MA (2006) Effect of starch granule size distribution and elevated amylose content on durum dough rheology and spaghetti cooking quality. Cereal Chem 83: 513–519
Stahl Y, Coates S, Bryce JH, Morris PC (2004) Antisense downregulation of the barley limit dextrinase inhibitor modulates starch granule size distribution, starch composition and amylopectin structure. Plant J 39:599–611
Stitt M, Lunn J, Usadel B (2010) Arabidopsis and primary photosynthetic metabolism: more than the icing on the cake. Plant J 61:1067–1091
Stoddard FL (2003) Genetics of starch granule size distribution in tetraploid and hexaploid wheats. Aust J Agric Res 54:637–648
Stoddard FL, Sarker R (2000) Characterization of starch in *Aegilops* species. Cereal Chem 77: 445–447
Taylor NE, Greene EA (2003) PARSESNP: a tool for the analysis of nucleotide polymorphisms. Nucleic Acids Res 31:3808–3811
Till BJ, Reynolds SH, Greene EA et al (2003) Large-scale discovery of induced point mutations with high-throughput TILLING. Genome Res 13:524–530
Toden S, Bird AR, Topping DL, Conlon MA (2007) Dose-dependent reduction of dietary protein induced colonocyte DNA damage by resistant starch in rats correlates more highly with caecal butyrate than with other short chain fatty acids. Cancer Biol Ther 6:253–258
Topping D (2007) Cereal complex carbohydrates and their contribution to human health. Trends Food Sci Technol 46:220–229
Topping DL, Clifton PM (2001) Short-chain fatty acids and human colonic function: roles of resistant starch and nonstarch polysaccharides. Physiol Rev 81:1031–1064
Tsai H, Howell T, Nitcher R et al (2011) Discovery of rare mutations in populations: TILLING by sequencing. Plant Physiol 27:1254–1268
Uauy C, Paraiso F, Colasuonno P et al (2009) A modified TILLING approach to detect induced mutations in tetraploid and hexaploid wheats. BMC Plant Biol 9:115
Varshney RK, Nayak SN, May GD, Jackson SA (2009) Next-generation sequencing technologies and their implications for crop genetics and breeding. Trends Biotechnol 27:522–530
Vignaux N, Doehlert DC, Elias EM et al (2005) Quality of spaghetti made from full and partial waxy durum wheat. Cereal Chem 82:93–100

Vriet C, Welham T, Brachmann A et al (2010) A suite of *Lotus japonicus* starch mutants reveals both conserved and novel features of starch metabolism. Plant Physiol 154:643–655

Wang TL, Uauy C, Robson F, Till B (2012) TILLING in extremis. Plant Biotech J doi: 10.1111/j.1467-7652.2012.00708

Waring S (2005) Functionality of resistant starch in food applications. http://www.foodinnovations.com/pdf/functresist.pdf

Wittwer CT, Reed GH, Gundry CN et al (2003) High-resolution genotyping by amplicon melting analysis using LCGreen. Clin Chem 49:853–860

Yamamori M (1998) Selection of a wheat lacking a putative enzyme for starch synthesis, SGP-1. In: Slinkard AE (ed) Proceedings of the Ninth International Wheat Genetics Symposium. University Extension Press, University of Saskatchewan, Saskatoon, Canada, pp 300–302

Yamamori M, Endo TR (1996) Variation of starch granule proteins and chromosome mapping of their coding genes in common wheat. Theor Appl Genet 93:275–281

Yamamori M, Nakamura T, Kuroda A (1992) Variations in the content of starch-granule bound protein among several Japanese cultivars of common wheat (*Triticum aestivum* L). Euphytica 64:215–219

Yamamori M, Nakamura T, Endo TR, Nagamine T (1994) Waxy protein deficiency and chromosomal location of coding genes in common wheat. Theor Appl Genet 89:179–184

Yamamori M, Fujita S, Hayakawa K et al (2000) Genetic elimination of a starch granule protein, SGP-1, of wheat generates an altered starch with apparent high amylose. Theor Appl Genet 101:21–29

Yasui T, Matsuki J, Sasaki T, Matsuki J (1999) Milling and flour pasting properties of waxy endosperm mutant lines of bread wheat (*Triticum aestivum* L.). J Sci Food Agric 79:687–692

Zeeman SC, Umemoto T, Lue WL, Auyeung P et al (1998) A mutant of Arabidopsis lacking a chloroplastic isoamylase accumulates both starch and phytoglycogen. Plant Cell 10:1699–1711

Zeeman SC, Kossmann J, Smith AM (2010) Starch: its metabolism, evolution, and biotechnological modification in plants. Annu Rev Plant Biol 61:209–234

Zeng M, Morris CF, Batey IL, Wrigley CW (1997) Sources of variation for starch gelatinization, pasting, and gelation properties in wheat. Cereal Chem 74:63–71

Zhao XC, Batey IL, Sharp PJ et al (1998) A single genetic locus associated with starch granule properties and noodle quality in wheat. J Cereal Sci 27:7–13

Chapter 21
Molecular Breeding for Quality Protein Maize (QPM)

Raman Babu and B. M. Prasanna

Abstract Maize endosperm protein is deficient in two essential amino acids, lysine and tryptophan. Several spontaneous and induced mutations that affect amino acid composition in maize have been discovered amongst which the *o2* gene (*opaque2*) has been used in association with endosperm and amino acid modifier genes for developing Quality Protein Maize (QPM), which contains almost double the amount of endosperm lysine and tryptophan as compared to the normal/non-QPM maize. These increases have been shown to have dramatic impacts on human and animal nutrition, growth and performance. A range of hard endosperm QPM germplasm has been developed at the International Maize and Wheat Improvement Center (CIMMYT) mostly through conventional breeding approaches to meet the requirements of various maize growing regions across the world. Microsatellite markers located within the *o2* gene provided opportunities for accelerating the pace of QPM conversion programs through marker-assisted selection (MAS). More recently, CIMMYT scientists are striving to develop reliable, easy-to-use markers for endosperm hardness and free amino acid content in the maize endosperm. Recent technological developments including high throughput, single seed-based DNA extraction, coupled with low-cost, high density SNP genotyping strategies, and breeder-ready markers for some key adaptive traits in maize, promise enhanced efficiency and cost effectiveness of MAS in QPM breeding programs. Here, we present a summary of QPM research and breeding with particular emphasis on genetic and molecular basis of *o2*, epistasis between *o2* and other high-lysine mutant genes, and the recent advances in genomics technologies that could potentially enhance the efficiency of molecular breeding for QPM in the near future.

R. Babu (✉)
International Maize and Wheat Improvement Center (CIMMYT), Km 45 Carretera, Mexico-Veracruz, Texcoco, 56130 Mexico D.F., Mexico
e-mail: r.babu@cgiar.org

B. M. Prasanna
CIMMYT, ICRAF House, United Nations Avenue, Gigiri, Nairobi 00621, Kenya

R. Tuberosa et al. (eds.), *Genomics of Plant Genetic Resources,*
DOI 10.1007/978-94-007-7575-6_21,

21.1 Introduction

Maize is one of the most important food crops in the world and together with rice and wheat, providing at least 30 % of the food calories to more than 4.5 billion people in 94 developing countries, which includes 900 million poor consumers. In parts of Africa and Mesoamerica, maize alone contributes over 20 % of food calories. The role of maize for human consumption, expressed in terms of the share of calories from all staple cereals, varies significantly across regions (Shiferaw et al. 2011). This ranges from 4 % in South Asia to 29 % in the Andean region to 61 % in Mesoamerica. In sub-Saharan Africa, maize is mainly a food crop accounting for 73 and 64 % of the total demand in eastern and southern Africa (ESA) and West and Central Africa (WCA). This makes maize particularly important to the poor in many developing regions of Africa, Latin America and Asia as a means of overcoming hunger and improving food security.

A typical kernel of a modern maize hybrid contains 73 % starch, 9 % protein, 4 % oil and 14 % other constituents (mainly fibre). The two major structures of the kernel, the endosperm and the germ (embryo), constitute approximately 80 and 10 % of the mature kernel dry weight, respectively. The endosperm is largely starch (approaching 90 %) and the germ contains high levels of oil (30 %) and protein (18 %). Bulk of the proteins in a mature maize kernel is in the endosperm (80 %) and the rest in the germ. While the germ protein is superior in both quantity and quality, the endosperm protein is deficient in lysine and tryptophan, and therefore, maize needs to be eaten with complementary protein sources such as legumes or animal products (Prasanna et al. 2001).

In the 1920s in a Connecticut (USA) maize field, a natural spontaneous mutation of maize with soft, opaque grains was discovered, which was eventually named as *o2 (opaque2)* (Singleton 1939). In 1964, Dr. Oliver Nelson's team at Purdue University, USA, discovered that the homozygous recessive *o2* allele had substantially higher lysine (+ 69 %) in grain endosperm compared to normal maize (Mertz et al. 1964). It was further determined that this mutation results in two-to-threefold increase in the level of two amino acids, lysine and tryptophan in comparison with normal genotype. The increased concentration of these two essential amino-acids (normally deficient in the maize grain endosperm) effectively doubles the biological value of maize protein (Bressani 1991) with the considerably profitable result that only half the amount of *o2* maize (relative to normal maize), needs to be consumed to obtain the same biologically usable protein (FAO 1992). In addition, other amino acids such as histidine, arginine, aspartic acid and glycine showed increase, while the decrease was observed for some amino acids such as glutamic acid, alanine and leucine. Decrease in leucine is considered desirable as it makes leucine–isoleucine ratio more balanced, which in turn helps to provide more tryptophan for niacin biosynthesis, and thus helps in combating pellagra.

Decades of efforts by researchers at CIMMYT led to the development of "Quality Protein Maize" (QPM) with enhanced nutritional value, especially through higher levels of lysine and tryptophan in the endosperm and better amino acid balance. In

Ethiopia, Tanzania and Uganda, randomized trials showed significantly improved height and weight of children consuming such varieties, particularly in Southern Ethiopia where the population relies heavily on maize (Gunaratna et al. 2008).

In this chapter, we describe the early efforts towards development of high-lysine maize, the biochemical and molecular bases of QPM, some successful examples of molecular marker-assisted breeding for QPM development, and suggest an integrated approach for enhancing the nutritional quality in maize, including low-cost and reliable markers for the QPM component traits, seed DNA-based genotyping, and genomic technologies.

21.1.1 Maize Protein

Maize endosperm protein is comprised of different fractions. Based on their solubility, these can be classified into albumins (water soluble), globulins (soluble in saline solution), zein or prolamine (soluble in alcohol) and glutelins (soluble in alkali). In normal maize endosperm, the average proportions of various fractions of protein are albumins 3 %, globulin 3 %, zein (prolamine) 60 % and glutelin 34 %, while the embryo protein is dominated by albumins (+ 60 %), which is superior in terms of nutritional quality. The zein in maize endosperm is low in lysine content (0.1 g/100 g of protein), which negatively affects growth of animals (Osborne and Mendel 1914). In *o2* maize, the zein fraction is markedly reduced by roughly 50 % with a concomitant increase in the relative amounts of nutritionally superior fractions such as albumins, globulins and glutelins. The endosperm of *o2* maize contains twice as much lysine and tryptophan and 30 % less leucine than normal maize. The decreased level of zein (5–27 %) in *o2* maize along with reduced leucine, leads to more tryptophan for niacin synthesis and thus helps to combat pellagra and significantly improves its nutritional quality (Prasanna et al. 2001).

21.1.2 High-Lysine Mutants in Maize

Several mutants have been detected that favorably influence maize endosperm protein quality by elevating levels of two essential amino acids, lysine and tryptophan. The discovery of *o2* (Mertz et al. 1964) was followed by recognition of the biochemical effects of *floury2* (*fl2*) (Nelson et al. 1965). Searches for new mutants continued and resulted in the discovery of several others such as *opaque7*(*o7*) (Misra et al. 1972), *opaque6*(o6) and *floury3* (*fl3*) (Ma and Nelson 1975), *mucronate* (*Mc*) (Salamini et al. 1983) and *defective endosperm* (*De-B30*) (Salamini et al. 1997). Attempts were also made to find genotypes with high-lysine genes that retained a high level of zein fraction. Two such mutants, *opaque7749* and *opaque7455* (*o11*) (Nelson 1981) are particularly interesting as they have markedly higher levels of lysine as well as a high prolamine fraction. The specific chromosomal location is known for some of

the mutants. For example, the *o2* mutant is located on chromosome7, *fl2*on chr.4, *o7* on chr.10, *fl3* on chr.8 and *de-B30* on chr.7. The genetic action of some of the mutants is also known, for example, *o2*, *o6*, *o7* and *o11* are completely recessive. The two floury mutants are semi-dominant and exhibit variable expression for kernel opacity and protein quality depending on the presence of one or more recessives in the triploid endosperm. The mutant *De-B30* is dominant and shows dosage effects on kernel opacity and zein content (Soave et al. 1982).

21.1.3 Pleiotropic and Secondary Effects of o2 and Other High-Lysine Mutants

Genes and gene combinations that bring about drastic alterations in either plant or kernel characteristics also produce several secondary or undesirable effects. The low prolamine or high-lysine mutants are no exception. In addition to influencing several biochemical traits, they adversely affect a whole array of agronomic and kernel characteristics. The *o2*and other mutants adversely affect dry matter accumulation resulting in lower grain yield due to increased endosperm size. The kernels dry slowly following physiological maturity of the grain and have a higher incidence of ear rots. Other changes generally associated with high-lysine mutants include thicker pericarp, larger germ size, reduced cob weight, increased color intensity in yellow maize grains, and reduction in kernel weight and density. Thus, despite the nutritional superiority of *o2* maize, it did not become popular with farmers as well as consumers mainly because of reduced grain yield, chalky and dull kernel appearance and susceptibility to ear rots and stored grain pests. Hence, CIMMYT undertook to improve the phenotype of *o2* kernels to facilitate greater acceptability by developing hard endosperm grain types with protein quality of chalky *o2* strains. CIMMYT received funding support beginning in 1965 from the United Nations Development Program and introduced gene modifiers that changed the soft, starchy endosperm to a vitreous type preferred by farmers and consumers whilst retaining the elevated levels of lysine and tryptophan. CIMMYT has subsequently developed a range of hard endosperm *o2* genotypes with better protein quality through genetic selection, which are popularly known as quality protein maize (QPM). Today's QPM is essentially interchangeable with normal maize in both cultivation and agronomic characteristics as well as competitive in terms of yield, lodging, disease and pest resistance, moisture level while retaining the superior lysine and tryptophan content (Vasal 2001). In 2005, QPM was planted on 695,200 ha across 24 developing countries.

21.2 QPM Development Through Conventional Breeding

There are various breeding options for developing hard endosperm, high-lysine maize that is competitive in agronomic performance and market acceptance which could be based on specific endosperm high-lysine mutants or other donor stocks. The past

approaches involving normal maize breeding populations have centered on altering germ-endosperm ratio, selection for multiple aleurone layers, and recurrent selection to exploit natural variation for high-lysine content. Altering the germ-endosperm ratio to favor selection of larger germ size will have the dual advantage of increasing both protein quantity and quality (Bjarnason and Pollmer 1972) but it is not practical to attain lysine levels approaching those of *o2* maize. Besides, increased germ size has the disadvantage of contributing to poor shelf life of maize. Recurrent selection for high lysine in normal endosperm breeding populations has been largely unsuccessful due to the narrow genetic variation and heavy dependence on laboratory facilities of this approach. Alternatively, high-lysine endosperm mutants provided two attractive options: (1) exploiting double mutants involving *o2* and (2) simultaneous use of the *o2* gene with endosperm and amino acid modifier genes. In most instances, double mutant combinations involving *o2* and other mutants associated with endosperm quality were not vitreous (Vasal 2001). The most successful and rewarding option exploited combined use of *o2* with the associated endosperm and amino acid modifier genes.

Segregation and analysis of kernels with a range of endosperm modification began at CIMMYT as early as in 1969 by John Lonnquist and V. L. Asnani. Initial efforts towards development of QPM donor stocks with good kernel phenotypes as well as good protein quality proved to be highly challenging. Two effective approaches, i.e., intra-population selection for genetic modifiers in *o2* backgrounds exhibiting a higher frequency of modified *o2*kernels and recombination of superior hard endosperm *o2* families, resulted in development of good quality QPM donor stocks with a high degree of endosperm modification. This was followed by the large-scale development of QPM germplasm with a wide range of genetic backgrounds, representing tropical, subtropical and highland maize germplasm and involving different maturities, grain color and texture. A summary of characteristics of promising QPM gene pools and populations developed at CIMMYT is provided in Table 21.1. An innovative breeding procedure designated as 'modified backcross cum recurrent selection' was designed to enable rapid and efficient conversion programs (Vasal et al. 2001). More recently pedigree back crossing schemes have been used to convert elite QPM lines to maize streak virus (MSV) resistance for deployment in Africa as well as conversion of elite African lines to QPM (Krivanek et al. 2007).

A QPM hybrid breeding program was initiated at CIMMYT in 1985 as the QPM hybrid product has several advantages over open pollinated QPM cultivars: (1) higher yield potential comparable to the best normal hybrids, (2) assured seed purity, (3) more uniform and stable endosperm modification and (4) less monitoring of protein quality required during seed production. Several QPM hybrid combinations were derived and tested through international trial series at multiple CIMMYT and NARS locations in Asia, Africa and Latin America. Current QPM breeding strategies at CIMMYT focus on pedigree breeding wherein the best performing inbred lines with complementary traits are crossed to establish new segregating families. Both QPM × QPM and QPM × non-QPM crosses are made depending upon the specific requirements of the breeding project. In addition, backcross conversion is also followed to develop QPM versions of parental lines of popular hybrid cultivars that

Table 21.1 Characteristics of QPM gene pools and populations developed at CIMMYT (using *o2* and associated modifiers) including protein, tryptophan and lysine contents in the whole grain. (Based on Vasal 2001)

QPM pop/pool	Adaptation	Maturity	Color	Texture	Protein (%)	Tryptophan in protein (%)	Lysine in protein (%)	Quality index
Population 61	Tropical	Early	Y	Flint	9.2	0.98	4.2	3.8
Population 62	Tropical	Late	W	Semi-flint	9.9	0.92	3.9	4.4
Population 63	Tropical	Late	W	Dent	9.1	0.97	4.3	4.3
Population 64	Tropical	Late	W	Dent	9.6	1.00	3.8	4.3
Population 65	Tropical	Late	Y	Flint	9.2	0.96	4.2	4.4
Population 66	Tropical	Late	Y	Dent	9.3	1.01	4.3	4.3
Population 67	Subtropical	Medium	W	Flint	9.9	1.04	3.9	4.8
Population 68	Subtropical	Medium	W	Dent	9.5	1.01	4.0	4.3
Population 69	Subtropical	Medium	Y	Flint	10.0	0.98	4.2	4.4
Population 70	Subtropical	Medium	Y	Dent	9.3	1.10	4.3	4.7
Pool 15 QPM	Tropical	Early	W	Flint-Dent	9.1	0.94	4.2	4.6
Pool 17 QPM	Tropical	Early	Y	Flint	8.9	1.04	4.5	4.5
Pool 18 QPM	Tropical	Early	Y	Dent	9.9	0.93	4.0	4.6
Pool 23 QPM	Tropical	Late	W	Flint	9.1	1.03	3.8	4.2
Pool 24 QPM	Tropical	Late	W	Dent	9.4	0.92	3.8	4.0
Pool 25 QPM	Tropical	Late	Y	Flint	9.8	0.94	4.0	4.0
Pool 26 QPM	Tropical	Late	Y	Dent	9.5	0.90	4.1	4.3
Pool 27 QPM	Subtropical	Early	W	Flint-Dent	9.5	1.05	4.2	4.8
Pool 29 QPM	Subtropical	Early	Y	Flint-Dent	9.2	1.06	4.3	4.8
Pool 31 QPM	Subtropical	Medium	W	Flint	10.2	0.96	4.1	4.5
Pool 32 QPM	Subtropical	Medium	W	Dent	8.9	1.04	4.2	4.5
Pool 33 QPM	Subtropical	Medium	Y	Flint	9.3	1.05	–	4.2
Pool 34 QPM	Subtropical	Medium	Y	Dent	9.1	1.10	4.1	4.5

are widely grown in CIMMYT's target regions. Inbred lines developed through this process are then used in formation of QPM hybrids and QPM synthetics (Krivanek et al. 2007).

The 2000 World Food Prize jointly honored two CIMMYT Scientists, Dr. S. K. Vasal (Plant Breeder) and Dr. Evangelina Villegas (Cereal Biochemist) for their combined efforts and stellar achievements in developing and promoting QPM varieties to improve productivity and nutrition in malnourished and poverty-stricken areas of the world.

The genetic make-up of the QPM necessitates their cultivation in isolation from normal maize, as any contamination with *O2* allele will be apparent in the form of normal transparent kernels in contrast to the marble-like appearance of *o2*kernels. As isolation distance of 300–400 m is adequate; with the increase in the number of border rows, this distance can be suitably reduced. Even in the absence of isolation, the farmers planting 2–4 ha can save the seed from the middle of the field, whereas the rest of the crop can be used as nutritionally superior grain.

21.3 Molecular Basis of QPM: O2 and Modifier Gene Action

The breeding of QPM involves manipulation of three distinct genetic systems (Bjarnason and Vasal 1992; Krivanek et al. 2007): (1) the recessive mutant allele of the *O2* gene, (2) the endosperm hardness modifier genes and (3) the amino acid modifiers/genes influencing free amino acid content in the endosperm. The *O2*gene was cloned using a transposon tagging strategy with the maize mobile genetic elements, *Spm*(Schmidt et al. 1990) and *Ac*(Motto et al. 1988). The *O2* gene encodes a leucine-zipper class transcription factor required mainly for the expression of 22 kDa α-zein-coding genes and a gene encoding a ribosomal inactivating protein (Lohmer et al. 1991; Bass et al. 1992). Lower α-zein content in *o2* endosperm results in protein bodies that are about one-fifth to one tenth the normal size, which is presumed to alter packing of starch grains during seed desiccation, thereby conferring a characteristic soft texture to the kernel. With the reduction of α-zeins in the endosperm due to *o2* mutation, there is a usually concomitant increase in the level of γ-zeins (Habben et al. 1993). The homozygous recessive allele causes a decrease of the production of these zeins resulting in a corresponding increase in non-zein proteins, rich in lysine and tryptophan (Gibbon and Larkins 2005). Additionally, the recessive allele of the *o2* transcription factor also reduces the production of the enzyme, lysine keto-glutarate reductase, involved in free lysine degradation resulting in enhanced free lysine in the endosperm of *o2* maize. In the segregating generations, this recessive allele is selected either visually (identifying mosaic ears on F_2 harvests) or using molecular markers.

The mechanism(s) by which the endosperm modifier genes convert the starchy endosperm of *o2* to a normal phenotype is still not completely understood, but some important clues have been obtained through analysis of biochemical changes in modified *o2* endosperm. Studies suggest that the products of the modifier genes interact

with γ-zein mRNA transcripts and enhance their transport from the nucleus or increase their stability and translation. The overproduction of γ-zein appears to enhance protein body number and result in the formation of more vitreous endosperm. The endosperm hardness modifier genes, which convert the soft/opaque endosperm to a hard/vitreous endosperm without much loss of protein quality, are selected through a low-cost but effective method of light-box screening, where light is projected through the vitreous grains or blocked by the opaque grains. Endosperm modification is polygenically controlled. However, genetic and molecular analyses revealed some major loci involved in *o2* modification; for example, one locus maps near the centromere of chromosome 7 and the second maps near the telomere on the long arm of chromosome 7 (Lopes et al. 1995).

Despite the presence of *o2* and associated endosperm hardness modifier genes, the lysine and tryptophan levels in segregating families vary widely indicating the existence of third set of genes that modify the amino acid content, which necessitates systematic biochemical evaluation of lysine and/or tryptophan levels in each breeding generation. The lysine content of normal maize is around 2 %, whereas it is approximately 4 % (of the total protein) in QPM, with a range 1.6–2.6 % in normal maize and 2.7–4.5 % in QPM. Three genes associated with lysine level have been mapped to locations on chromosome 2, 4 and 7, besides several major *o*2 modifier-QTLs on chromosomes 1, 7 and 9 (Gibbon and Larkins 2005). Therefore, it is possible to get favorable responses to selection for endosperm texture modification as well as relative content of the essential amino acids, if they are monitored efficiently, during the QPM breeding programs.

21.4 Molecular Marker-Assisted Selection (MAS) in QPM Breeding

The *o2* gene is recessive and the modifiers are polygenic. Their introgression into elite inbred lines is not straight forward because of three major factors, (1) each conventional backcross generation needs to be selfed to identify the *o2* recessive gene and a minimum of six backcross generations are required to recover satisfactory levels of the recurrent parent genome, (2) in addition to maintaining the homozygous *o2* gene, multiple endosperm modifiers must also be selected, and (3) rigorous biochemical tests to ensure enhanced lysine and tryptophan levels in the selected materials in each breeding generation require enormous labor, time and financial resources. Although conventional breeding procedures have been used to convert commercial lines to QPM forms, these procedures are tedious and time consuming. Rapid advances in genomics research and technologies has led to the use of MAS which holds promise in enhancing selection efficiency and expediting the development of new cultivars with higher yield potential (Ribaut and Hoisington 1998; Xu and Crouch 2011). While marker-assisted foreground selection (Tanksley 1983; Melchinger 1990) helps in identifying the gene of interest without extensive phenotypic assays, marker-assisted background selection (Young and Tanksley 2005; Hospital et al. 1992; Frisch et al. 1999a, b) significantly expedites the rate of genetic

gain/recovery of recurrent parent genome in a backcross breeding program. With the development and access to reliable PCR-based allele-specific markers such as simple sequence repeats (SSRs) and single nucleotide polymorphisms (SNPs), MAS is becoming an attractive option, particularly for oligogenic traits such as QPM (Babu et al. 2004).

A rapid line conversion strategy for QPM has been developed (Babu et al. 2005), consisting of a two-generation backcross program that employs foreground selection for the *o2* gene in both backcross (BC) generations, background selection at non-target loci in the BC_2 generation, and phenotypic selection for kernel modification and other desirable agronomic traits in two subsequent selfed generations. The rapid line conversion strategy outlined in this investigation brings together the salient features of both marker-aided and phenotypic-based selection approaches such as fixing the large segregating generation for the target locus (*o2*), reducing the linkage drag by selection of flanking markers for recipient allele type, recovering maximum amount of recurrent parent genome within two BC generations and providing scope for precise phenotypic selection for desirable agronomic and biochemical traits on a reduced number of progeny.

21.4.1 Low-Cost and Reliable Markers for o2 and Modifier Genes

The cloning and characterization of the *o2* gene, followed by detection of three SSR markers (*phi057*, *phi112* and *umc1066*) within the gene (Lin et al. 1997; Yang et al. 2008), led to effective differentiation of the *O2* and *o2* alleles (Bantte and Prasanna 2003; Babu et al. 2005). These *o2*-specific SSR markers provide an excellent foundation for MAS but this alone is not sufficient to bring to bear the full effectiveness of molecular breeding for QPM genotypes. Each of the microsatellite markers located within the *o2* gene are associated with factors that challenge their routine use in MAS programs. *umc1066* is easily visualized on agarose gels but is commonly not polymorphic in CIMMYT breeding populations; *phi057* is difficult to visualize on agarose gels, usually requiring the use of polyacrylamide gels; *phi112* is a dominant marker and hence cannot be used in the identification of heterozygotes in F2/BC populations. However, *phi112*, which is based on a deletion in the promoter region, has the advantage of being a widely conserved marker, consistent with the phenotype in QPM germplasm tested. In order to overcome these difficulties, CIMMYT is in the process of identifying functional and more discriminative SNP markers that could be used in high throughput genotyping platforms for selection of the *o2* allele.

Effective markers associated with modifying loci for both endosperm hardness and amino acid levels need to be identified. Unfortunately relatively little is known about the number, chromosomal location and mechanism of action of these modifier genes. A complex system of genetic control of these modifier loci with dosage effects, cytoplasmic effects, incomplete and unstable penetrance in different QPM germplasm

creates a major bottleneck to the accelerated development of QPM germplasm. Using a limited set of restriction fragment length polymorphism (RFLP) markers and bulked segregant analysis (Lopes et al. 1995), two chromosomal regions on the long arm of chromosome 7 that are associated with *o2* endosperm hardness modification were identified. The locus near the centromere is linked with the gene encoding the 27-kDa gamma zein. More recently, the analysis of two different QPM genotypes, Ko326Y and CM105Mo2 (derived from CIMMYT's Pool 33 QPM) corroborated the existence of a common quantitative trait locus (QTL) near the centromere of chromosome 7 that appears to have a major effect (30 % of the phenotypic variance) on *o2* endosperm modification, in addition to a QTL on 9.04/9.05 (Holding et al. 2008). In this specific F_2 population segregating for kernel vitreousness, these two loci accounted for 40 % of the phenotypic variation and thus may prove to be strong candidates for MAS for QPM breeding.

Precise information on genes controlling the level of amino acid modification especially with respect to lysine and tryptophan is relatively scarce and studies to date have found several quantitative trait loci (QTL) on many of the maize chromosomes (Wang and Larkins 2001; Wu et al. 2007). The free amino acid (FAA) content in Oh545o2 is 12 times greater than its wild-type counterpart, and three and 10 times greater than in Oh51Ao2 and W64Ao2, respectively. QTL mapping in these lines identified four significant loci that account for about 46 % of the phenotypic variance for FAA (Wang et al. 2001). One locus on the long arm of chromosome 2 is coincident with genes encoding a monofunctional *Aspartate kinase 2* (*Ask2*) whereas another locus on the short arm of chromosome 3 is linked with a cytosolic triose phosphate isomerase 4. Subsequent feedback inhibition analysis has suggested that *Ask2* is the candidate gene associated with the QTL on 2S and that a single amino acid substitution in the C-terminal region of the *Ask2* allele of Oh545o2 is responsible for altered basal activity of the enzyme (Wang et al. 2001). Using a RIL population from the cross between B73o2 (an *o2* conversion of B73) and a QPM line (CML161), it was possible to identify three QTL for lysine content and six QTL for tryptophan content, which explained 32.9 and 49.1 % of the observed variation, respectively (Gutierrez-rojas 2007). Thus a series of molecular markers for manipulation of different genetic components of QPM is available and hence their validation and fine mapping in appropriate breeding populations should now be carried out in order to establish a single cost-effective MAS assay for molecular breeding of QPM. Concerted research efforts to quantify the effect of these loci affecting endosperm hardness and amino acid levels coupled with marker development and validation will also accelerate the pace and precision of QPM development programs.

The previously cited studies involved very diverse germplasm, including tropical x temperate crosses, but current research at CIMMYT is seeking useful markers associated with kernel hardness and high amino acid levels in elite QPM $\times$ QPM crosses. A combination of bulked segregant analysis (BSA) and genome-wide SNP scan (using Illumina's GoldenGate assay) in phenotypically contrasting (opaque vs. modified and high vs. low tryptophan) progenies of seven QPM $\times$ QPM populations have identified several genomic regions putatively associated with kernel hardness and high tryptophan concentration. BSA of class5 (opaque) and class1 (completely

modified) bulks identified five regions [1.04/1.05 (1S), 7.02 (7S), 8.06 (8L), 9.04/9.05 (9L) and 10.04 (10S)] associated with kernel hardness modification, each indicated by more than three SNP markers in at least four populations. The region on 7.02, in which a major QTL has been previously reported to be linked to the 27 kDa gamma-zein gene responsible for kernel hardness modification in QPM, was consistently associated with kernel modification for all seven populations. Genome-wide SNP scan of the backcross-derived 'high-low' tryptophan lines revealed several polymorphic regions, notably 2.07, 5.03. and 10.03, which are currently being validated at CIMMYT using SSR and SNP markers and additional segregating populations. As a cross validation, whole genome SNP profiles of 12 isoline pairs of QPM and their normal versions were generated, which confirmed the importance of the above-mentioned eight genomic regions associated with kernel modification and tryptophan content (Babu et al. 2009). These genomic regions and their diagnostic markers may be useful for designing a comprehensive system for cost-effective marker-assisted QPM breeding.

21.4.2 Seed DNA-Based Genotyping in MAS for QPM Development

Leaf collection from the field, labeling and tracking back to the source plants after genotyping are rate limiting steps in leaf DNA-based genotyping. Recently, an optimized genotyping method using endosperm DNA sampled from single maize seeds was developed at CIMMYT, which has the potential to replace leaf DNA-based genotyping for marker-assisted QPM breeding (Gao et al. 2008). This method is suitable for various types of maize seeds, produces high quality and quantity of DNA and has minimal effects on subsequent germination and establishment. A substantial advantage of this approach is that it can be used to select desirable genotypes before planting, which can bring about dramatic enhancements in efficiency by planting only the plants containing *o2* gene in recessive form in BC_nF_2 generations of non-QPM × QPM crosses, and also by minimizing the labor costs and scoring error associated with light box screening of a large number of grains for endosperm hardness. Over several breeding cycles, this is likely to lead to cumulative and accelerated gains in selection pressure (such as light box screening for endosperm hardness modification and systematic biochemical evaluation of lysine and/or tryptophan) and improvements in overall QPM breeding efficiency.

21.4.3 MAS and Development of QPM Cultivars: Successful Examples

To hasten the pace of progress of QPM cultivar development, and most importantly, to diversify the genetic base of QPM cultivars for any targeted agro-ecology, it is

important to convert some of the highly diverse and agronomically elite non-QPM inbred lines into QPM versions and derive more heterotic QPM hybrid combinations. Conversion of a normal maize line to a QPM version through conventional backcross breeding requires at least 6–7 years, since the desired *o2* allele has to be in a homozygous recessive state, thereby warranting progeny testing after every backcross. In addition, as discussed previously, the introgression of the polygenic endosperm modifiers into elite inbreds is not a straight-forward procedure as this is presently based on phenotypic selection alone and rigorous biochemical tests have to be undertaken to ensure enhanced lysine and tryptophan levels in the selected genotypes. Consequently, such a breeding programme would involve significant land, labour and financial resources (Dreher et al. 2003).

The recent developments in plant biotechnology, including molecular mapping and marker-assisted selection (MAS) offer a choice of options for introgression of the target gene(s) in the genetic background of elite varieties of major crops. MAS refers to the manipulation of genomic regions that are involved in the expression of traits of interest using molecular markers (Babu et al. 2004). The PCR based molecular markers help in targeted 'foreground selection' of segregating/backcross progenies possessing the desired gene(s), besides shortening of the breeding cycle significantly through rapid recovery of recurrent parent genome using 'background selection'. Microsatellite or Simple Sequence Repeat (SSR) markers are particularly useful in undertaking MAS in crop plants like maize.

Identification and utilization of *o2*-specific SSR markers (Lin et al. 1997; Bantte and Prasanna 2003) offered tremendous advantage in molecular marker-assisted conversion of non-QPM lines into their QPM versions. 'Foreground selection' for the *o2* allele using SSR markers and 'background selection' (using markers polymorphic between the donor and recurrent parents) aid in recovering individuals with desired genotype at the target locus, besides high levels of recovery of recurrent parent genome, within two to three backcross generations. This program can be thus implemented in a cost- and time-effective manner as compared to that based on phenotypic selection alone (Dreher et al. 2003).

There are a few successful examples of MAS for maize improvement using *o2*-specific SSR markers (Babu et al. 2005; Gupta et al. 2009; Prasanna et al. 2010). The parental lines of 'Vivek Hybrid 9' (CM145 and CM212), developed at Vivekananda Parvatiya Krishi Anusandhan Sansthan (VPKAS), Almora, were converted into QPM versions through transfer of *o2* gene using MAS and phenotypic screening for endosperm modifiers. The MAS-derived QPM hybrid, 'Vivek QPM 9' has been released in the year 2008 for commercial cultivation in zones I and IV in India. Vivek QPM 9 shows 41 % increase in tryptophan, 30 % in lysine, 23 % in histidine, 3.4 % in methionine, coupled with 12 % reduction in leucine, as compared to Vivek Hybrid 9. Domestic consumption of such QPM grains will help in reducing protein malnutrition in the hills and mountains. In view of this, F_1 hybrid seeds of Vivek QPM 9 are being produced in large scale for distribution in Uttarakhand and other parts of the country. Few villages in Uttarakhand have also been identified for converting them into QPM villages. Vivek QPM 9 can potentially replace Vivek Hybrid 9 as well as composites without any yield penalty in these areas. The approach outlined above

was also used to develop QPM versions of several elite, early maturing inbred lines adapted to the hill regions of India (Gupta et al. 2009, 2013).

QPM versions of six elite inbred lines, which are the parents of three single-cross hybrids, PEHM2, Parkash and PEEHM5 have been recently developed by the Maize Genetics Unit, IARI, New Delhi, through the ICAR Network Project on Molecular Breeding (Prasanna et al. 2010). A Network Project on molecular breeding for QPM is also being implemented in India, funded by the Department of Biotechnology (DBT), Govt. of India, for conversion of several important Indian maize inbred lines into QPM versions.

21.5 Need for an Integrated Approach for Enhancing Maize Protein Quality

Research at CIMMYT is currently focused on developing a package of molecular markers for cost-effective large scale, marker-aided QPM breeding program. CIMMYT recently developed gene-based SNP markers and high throughput KASPar chemistry-based genotyping assay for the *o2* gene. This system is likely to be useful especially for NARS programs with limited or no lab facilities, wherein NARS researchers could effectively outsource the SNP genotyping process to cost-effective service providers in the region. In addition, intensive efforts are being made to develop and validate new as well as existing markers for the endosperm and amino acid modifier genes across a wide range of populations and improved pools. Recent investigations into the improved protein quality of the *o2* mutant and the genetic mechanisms that can suppress its starchy kernel phenotype provide new insights to support the continued improvement of QPM. Chief among these developments are the use of transgenic approaches to improve nutritional quality and the discovery that an important component of modified endosperm texture in QPM is related to altered starch granule structure (Gibbon and Larkins 2005).

Another possible opportunity to further improve the protein quality conferred by *o2* is to pyramid it with second high-lysine mutant, *o16*, which was selected from Robertson's *Mutator* (*Mu*) stock where the lysine content in the F_3 *o2o2o16o16* families derived from recombination of both *o2* and *o16*, was about 30 % higher than that of *o2o2* or *o16o16* F3 families (Yang et al. 2004). Using both genes together could lessen the requirement for phenotypic screening for amino acid content and genetic screening for amino acid modifiers.

Recent efforts in genetic transformation are focused on developing a dominant *o2* trait in maize. RNA interference (RNAi) technology has also been used to reduce 22-kDa (Segal et al. 2003) and 19-kDa alpha zeins (Huang et al. 2004, 2005) using antisense transformation constructs, which resulted in moderate increases (15–20 %) in lysine content. In a recent study, using an improved double strand RNA (dsRNA) suppression construct, Huang et al. (2006) reported lysine and tryptophan levels similar to conventionally bred QPM genotypes. While the dominant nature of the anti-sense transgene is a definite advantage as compared to recessive allele of *o2*, the opaque

endosperm still needs to be modified by endosperm modifier genes whose epistasis with the transgene has not yet been tested. Very recently, Wu and Messing (2011) reported a potential accelerated QPM selection scheme, which is based on RNA interference construct that is directed against 22 and 19 kDa zeins, fused with visible green fluorescent protein (GFP) marker gene. When such RNAi lines were crossed with QPM lines, carrying *o2* kernel hardness modifier genes, green and vitreous progenies could be selected in the segregating generations thereby demonstrating that high-lysine content and hard endosperm traits could be selected in dominant fashion. Although RNAi technology has emerged as a powerful tool to overcome the pleiotropic and secondary effects of the desirable mutant genes, social acceptance and biosafety concerns regarding genetically modified food crops still exist in some countries for large scale adoption.

21.6 Conclusion

The opportunities for implementing breeding for improved nutritional quality in crops like maize have increased tremendously in the recent years. Significant strides have also been made, particularly with regard to MAS for generating QPM versions for elite inbred lines (Babu et al. 2005; Gupta et al. 2013) and identification of genes/QTLs influencing diverse quality traits in maize (Babu et al. 2013). Yet, the application of molecular tools to accelerate breeding for improved nutritional quality in maize has barely begun, and there is vast potential and need to expand the scope and impact of such operations. Breeders will want to avail molecular tools to more efficiently add value to new maize cultivars, for example, by enhancing their nutritional or biochemical qualities for use as food, feed, and industrial material.

Biofortification of maize grains is an important area of research for which *o2* provides an ideal platform upon which a number of nutritionally important traits such as enhanced Fe and Zn, low phytate (for increased bioavailability of nutrients), high provitamin A and high methionine could be combined to derive multiple benefits. Considering the pace of technological developments in genome research, molecular breeding is likely to be the leading option in future for stacking a range of nutritionally important specialty traits.

References

Babu R, Nair SK, Prasanna BM, Gupta HS (2004) Integrating marker assisted selection in crop breeding-prospects and challenges. Curr Sci 87:607–619

Babu R, Nair SK, Kumar A et al (2005) Two-generation marker-aided backcrossing for rapid conversion of normal maize lines to quality protein maize (QPM). Theor Appl Genet 111: 888–897

Babu R, Atlin G, Vivek B et al (2009) Bulked segregant analysis using the high throughput maize GoldenGate SNP genotyping assay reveals multiple genomic regions associated with kernel hardness and tryptophan content in quality protein maize. Maize Genet Conf Abstracts 51:P178

Babu R, Rojas NP, Gao S, Pixley K (2013) Validation of the effects of molecular marker polymorphisms in LcyE and CrtRB1 on provitamin A concentrations for 26 tropical maize populations. Theor Appl Genet 126:389–399

Bantte K, Prasanna BM (2003) Simple sequence repeat polymorphism in Quality Protein Maize (QPM) lines. Euphytica 129:337–344

Bass HW, Webster C, Obrian GR et al (1992) A maize ribosome-inactivating protein is controlled by the transcriptional activator *Opaque-2*. Plant Cell 4:225–234

Bjarnason M, Pollmer WG (1972) The maize germ: Its role as contributing factor to protein quantity and quality. Zeitschrift Pflanzenzuchtung 68:83–89

Bjarnason M, Vasal SK (1992) Breeding of quality protein maize (QPM). Plant Breed Rev 9:181–216

Bressani R (1991) Protein quality of high lysine maize for humans. Cereal Food World 36:806–811

Dreher K, Khairallah M, Ribaut JM, Morris M (2003) Money matters (I): Costs of field and laboratory procedures associated with conventional and marker-assisted maize breeding at CIMMYT. Mol Breed 11:221–234

FAO (1992) Maize in human nutrition. Food and nutrition series, no. 25. Food and Agriculture Organization (FAO), Rome, Italy

Frisch M, Bohn M, Melchinger AE (1999a) Comparison of selection strategies for marker assisted back crossing of a gene. Crop Sci 39:1295–1301

Frisch M, Bohn M, Melchinger AE (1999b) Minimum sample size and optimum positioning of flanking markers in marker assisted back crossing for transfer of a target gene. Crop Sci 39: 967–975

Gao S, Martinez C, Skinner DJ et al (2008) Development of a seed DNA-based genotyping system for marker-assisted selection in maize. Mol Breed 22:477–494

Gibbon BC, Larkins BA (2005) Molecular genetic approaches to developing quality protein maize. Trends Genet 21:227–233

Graham GG, Lembcke J, Morales E (1990) Quality protein maize as the sole source of dietary protein and fat for rapidly growing young children. Pediatrics 85:85–91

Gunaratna NS, De Groote H, Nestel P, Pixley KV, McCabe, CP (2010). A meta-analysis of community-based studies on quality protein maize. Food Policy. 35:202–210.

Gupta HS, Agarwal PK, Mahajan V et al (2009) Quality protein maize for nutritional security: Rapid development of short duration hybrids through molecular marker assisted breeding. Curr Sci 96:230–237

Gupta HS, Babu R, Agarwal PK, Mahajan V et al (2013) Accelerated development of quality protein maize hybrid through marker-assisted introgression of *opaque-2* allele. Plant Breed 132:77–82

Gutierrez-rojas LA (2007) Quantitative trait loci analysis to identify modifier genes of the gene *o2* in maize endosperm. PhD Dissertation, Texas A & M University

Habben IE, Kirleis AW, Larkins BA (1993) The origin of lysine-containing proteins in *opaque-2* maize endosperm. Plant Mol Biol 23:825–838

Holding DR, Hunter BG, Chung T et al (2008) Genetic analysis of *o2* modifier loci in quality protein maize. Theor Appl Genet 117:157–170

Huang SS, Adams WR, Zhou Q et al (2004) Improving nutritional quality of maize proteins by expressing sense and antisense zein genes. J Agric Food Chem 52:1958–1964

Huang SS, Kruger DE, Frizzi A et al (2005) High-lysine corn produced by the combination of enhanced lysine biosynthesis and reduced zein accumulation. Plant Biotech J 3:555–569

Huang SS, Frizzi A, Florida CA et al (2006) High lysine and high tryptophan transgenic maize resulting from the reduction of both 19- and 22-kD alpha-zeins. Plant Mol Biol 61:525–535

Krivanek AF, De Groote H, Gunaratna NS et al (2007) Breeding and disseminating quality protein maize (QPM) for Africa. African J Biotech 6:312–324

Lin KR, Bockoit AJ, Smith JD (1997) Utilization of molecular probes to facilitate development of Quality Protein Maize. Maize Genet Coop News 71:22–23

Lohmer S, Maddaloni M, Motto M et al (1991) The maize regulatory locus *Opaque-2* encodes a DNA-binding protein which activates the transcription of the *b-32* gene. EMBO J 10:617–624

Lopes MA, Takasaki K, Botswick DE et al (1995) Identifications of two *o2* modifier loci in Quality Protein Maize. Mol Gen Genet 247:603–613
Ma Y, Nelson OE (1975) Amino acid composition and storage proteins in two new high lysine mutants in maize. Cereal Chem 52:412–419
Melchinger AE (1990) Use of molecular markers in plant breeding. Plant Breed 104:1–19
Mertz ET, Bates LS, Nelson OE (1964) Mutant genes that change protein composition and increase lysine content of maize endosperm. Science 145:279–280
Misra PS, Jambunathan R, Mertz ET et al (1972) Endosperm protein synthesis in maize mutants with increased lysine content. Science 176:1426
Motto M, Maddolini M, Panziani G et al (1988) Molecular cloning of the o2-*m5* allele of *Zea mays*, using transposon tagging. Mol Gen Genet 121:488–494
Nelson OE (1981) The mutants *opaque9* through *opaque13*. Maize Genetics Coop. Newsletter 55:68
Nelson OE, Mertz ET, Bates LS (1965) Second mutant gene affecting the amino acid pattern of maize endosperm proteins. Science 150:1469–1470
Osborne TB, Mendel LB (1914) Nutritive properties of protein of the maize kernel. J Biol Chem 18:1–16
Prasanna BM, Pixley K, Warburton ML, Xie C (2010) Molecular marker-assisted breeding for maize improvement in Asia. Molecular Breeding. 26:339–356
Prasanna BM, Vasal SK, Kassahun B, Singh NN (2001) Quality protein maize. Curr Sci 81: 1308–1319
Ribaut JM, Hoisington DA (1998) Marker assisted selection: new tools and strategies. Trends Plant Sci 3:236–239
Salamini F, Di Fonzo N, Fornasari E et al (1983) *Mucronate*, *mc*, a dominant gene of maize which interacts with *o2* to suppress zein synthesis. Theor Appl Genet 65:123–128
Salamini F, Di Fonzo N, Gentinetta E, Soave C (1997) A dominant mutation interfering with protein accumulation in maize seeds. In: Seed protein improvement in cereals and grain legumes, IAEA, Vienna, p. 97
Schmidt RJ, Burr FA, Aukerman MJ, Burr B (1990) Maize regulatory gene *opaque-2* encodes a protein with a "leucine-zipper" motif that binds to zein DNA. Proc Natl Acad Sci USA 87:46–50
Segal G, Song RT, Messing J (2003) A new opaque variant of maize by a single dominant RNA-interference-inducing transgene. Genetics 165:387–397
Shiferaw B, Prasanna B, Hellin J, Banziger M (2011) Crops that feed the world 6. Past successes and future challenges to the role played by maize in global food security. Food Security 3:307–327
Singleton WR (1939) Recent linkage studies in maize: V. *opaque endosperm-2* (*o2*). Genetics 24:61–63
Soave C, Reggiani R, Di Fonzo N, Salamini F (1982) Clustering of genes for 20 kd zein sub units in the short arm of maize chromosome 7. Genetics 97:363–377
Tanksley SD (1983) Molecular markers in plant breeding. Plant Mol Bio Rep 1:1–3
Vasal SK (2000) The quality protein maize story. Food Nutr Bull 21:445-450
Vasal SK (2000) The quality protein maize story. Food Nutr Bull 21:445–450
Vasal SK (2001) High quality protein corn. In: Hallauer A (ed) Specialty corn, 2nd edn. CRC, Boca Raton, FL, pp 85–129
Vasal SK, Villegas E, Bjarnason M et al (1980) Genetic modifiers and breeding strategies in developing hard endosperm *o2* materials. In: Pollmer WG and Phipps RH (eds) Improvement of quality traits of maize grain and silage use
Villegas E, Ortega E, Bauer R (1984) Chemical methods used at CIMMYT for determining protein quality in corn. CIMMYT, Mexico
Wang X, Larkins BA (2001) Genetic analysis of amino acid accumulation in *o2* maize endosperm. Plant Physiol 12:1766–1777
Wang X, Stumpf DK, Larkins BA (2001) Aspartate kinase 2-a candidate gene of quantitative trait locus influencing free amino acid content in maize endosperm. Plant Physiol 125:778–787

Wang X, Lopez-Valenzuela JA, Gibbon BC et al (2007) Characterization of monofunctional aspartate kinase genes in maize and their relationship with free amino acid content in the endosperm. J Exp Bot 58:2653–2660

Wu RL, Lou XY, Ma CX et al (2002) An improved genetic model generates high-resolution mapping of QTL for protein quality in maize endosperm. Proc Natl Acad Sci USA 99:11281–11286

Wu Y, Messing J (2011) Novel genetic selection system for Quantitative Trait Loci of Quality Protein Maize. Genetics 188:1019–1022

Xu Y, Crouch JH (2008) Marker assisted selection in plant breeding: From publications to practice. Crop Sci 48:391–407

Yang W, Zheng Y, Ni S, Wu J (2004) Recessive allelic variation of three microsatellite sites within the *o2* gene in maize. Plant Mol Biol Rep 22:361–374

Yang W, Zheng Y, Zheng W, Feng R (2005) Molecular genetic mapping of a high-lysine mutant gene (opaque-16) and the double recessive effect with *opaque-2* in maize. Mol Breed 15:257–269

Young ND, Tanksley SD (1989) RFLP analysis of the size of chromosomal segments retained around *Tm-2* locus of tomato during backcross breeding. Theor Appl Genet 77:353–359

Index

R. Tuberosa et al. (eds.), *Genomics of Plant Genetic Resources,*
DOI 10.1007/978-94-007-7575-6,