Pushpendra K. Gupta
Rajeev K. Varshney *Editors*

Cereal Genomics II

Springer

Cereal Genomics II

Pushpendra K. Gupta · Rajeev K. Varshney
Editors

Cereal Genomics II

Springer

Editors
Pushpendra K. Gupta
Molecular Biology Laboratory
Department of Genetics Plant Breeding
Chaudhary Charan Singh University
Meerut, India

Rajeev K. Varshney
Centre of Excellence in Genomics
International Crops Research Institute for
Semi-Arid Tropics (ICRISAT)
Patancheru
Hyderabad, India

ISBN 978-94-007-6400-2 ISBN 978-94-007-6401-9 (eBook)
DOI 10.1007/978-94-007-6401-9
Springer Dordrecht Heidelberg New York London

Library of Congress Control Number: 2013936960

Printed on acid-free paper

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

Preface

Cereals are the most important food crops of the world and make an important component of daily diet of a major section of human population. Cereals are also an important source of fat-soluble vitamin E (an essential antioxidant), and contribute 20–30 % of the daily requirement of minerals. Cereals provide >60 % of food requirement of growing human population. In view of this, it is necessary that the plant breeders continue to work toward increased crop productivity of cereals using the latest knowledge and genomics-based technologies that are becoming available at an accelerated and unprecedented pace.

Cereal production witnessed a significant and steady progress during the last few decades. This has been possible partly through the development of high-yielding and input-responsive cultivars during the so-called green revolution period. Introgression of alien genetic variation from related wild species also contributed to this revolution. During the last two decades, starting in mid-1990s, genomics approaches (including molecular marker technology) have been extensively used not only for understanding the structure and function of cereal genomes, but also for accelerated improvement of available cultivars in all major cereals. In our earlier edited volume "Cereal Genomics" published in 2004, we tried to collect information generated till then in the subject area of cereal genomics. That volume served a useful purpose and was well received by cereal workers globally. However, major advances in the field of cereal genomics have been made during the last 8 years (2004–2012), thus making our earlier 2004 volume out-of-date. This made it necessary for us to have a fresh look on the present status and future possibilities of cereal genomics research and hence this volume, "Cereal Genomics II". For instance, in 2004, except rice, no other cereal genome was sequenced, while now whole genome high quality or draft sequences of rice, maize, sorghum, barley, and that of the model grass species *Brachypodium distachyon* have become available and that of bread wheat should become available within the next 2–3 years. These genome sequences have become valuable resource for detailed analysis and improvement of cereals.

"Cereal Genomics II" has updated chapters on molecular markers, next generation sequencing platform and their use for QTL analysis, domestication studies, functional genomics, and molecular breeding. In addition, there are also chapters on computational genomics, whole genome sequencing and comparative genomics

of cereals. We believe that this book should prove useful to the students, teachers, and young research workers as a ready reference to the latest information on cereal genomics.

The editors are grateful to the authors of different chapters (Appendix I), who not only summarized the published research work in their area of expertize but also shared their unpublished results to make the articles up-to-date. We also appreciate their cooperation in meeting the deadlines and in revising their manuscripts, whenever required. While editing this book, the editors also received strong support from some colleagues (Appendix II), who willingly reviewed the manuscripts for their inclination toward the science of cereal genomics. Their constructive and critical suggestions have been very useful for improvement of the manuscripts.

The editors would like to extend their sincere thanks to colleagues and staff from their respective laboratories, who helped them to complete this important assignment. In particular, Manish Roorkiwal, B. Manjula, and Reyazul Rouf Mir helped RKV with the editorial work. The editors also recognize that the editorial work for this book has been quite demanding and snatched away from them some of the precious moments, which they should have spent together with their respective families. PKG is thankful to his wife Sudha Gupta and to the families of his son (Ankur) and daughter (Ritu) and RKV is thankful to his wife Monika Varshney and two kids (Prakhar and Preksha) for their support and help. RKV is also grateful to Dr. William Dar, Director General, ICRISAT for his guidance and support to complete the "Cereal Genomics II" volume. The cooperation and help received from Ineke Ravesloot and Jacco Flipsen of Springer during various stages of the development and completion of this project is duly acknowledged.

The book was edited during the tenure of PKG as NASI Senior Scientist at CCS University, Meerut (India) and that of RKV as Director, Center of Excellence in Genomics (CEG), ICRISAT, Hyderabad (India) and Theme Leader—Comparative and Applied Genomics (CAG), Generation Challenge Programme (GCP). The editors hope that the book will prove useful for the targeted audience and that the errors, omissions, and suggestions, if any, will be brought to their notice, so that a future revised and updated edition, if planned, may prove more useful.

Meerut, India — P. K. Gupta
Hyderabad, India — R. K. Varshney

Contents

Chapter 1
Cereal Genomics: Excitements, Challenges and Opportunities

Pushpendra K. Gupta and Rajeev K. Varshney

1.1 Introduction

Cereals constitute the most important food crops of the world, occupying ~680 million hectares of land and producing ~2,295 million tonnes of food grain globally (October 4, 2012; http://www.fao.org/worldfoodsituation/wfs-home/csdb/en/), even though this production is lower than that for the year 2011 (2,340 million tonnes). Cereals are also an excellent source of fat-soluble vitamin E (an essential antioxidant) and contain 20–30 % of our daily mineral requirement (including selenium, calcium, zinc, and copper). Among all crops, cereals also provide 60 % of calories and proteins for the growing human population, which is estimated to reach 9.2 billion level in the year 2050. The projected need for annual cereal production in 2050 is ~3,000 million tonnes, so that at least a 30 % increase in the annual cereal grain production would be needed to meet this demand; this translates into an annual growth rate of ~0.70 %, which should not be difficult to achieve if the present growth rate of ~1.0 % is maintained. It has also been noticed that although during the last 4–5 decades the annual production of cereals (sum of wheat, milled rice and coarse grains) has been steadily increasing, the rate of growth in this production has shown a fatigue, with the growth rate falling from 3.7 % p.a. in the 1960s, to 2.5 % in 1970s, 1.4 % in 1980s and 1.1 % in 1990s, this growth rate sometimes also being negative (years 2006–2007, 2010–2011; see Fig. 1.1). The rate of growth in

P. K. Gupta (✉)
Department of Genetics and Plant Breeding, CCS University, Meerut 250004, India
e-mail: pkgupta36@gmail.com

R. K. Varshney (✉)
Center of Excellence in Genomics (CEG), International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru 502324, India
e-mail: r.k.varshney@cgiar.org

R. K. Varshney
CGIAR Generation Challenge Programme (GCP), c/o CIMMYT, 06600 Mexico, DF, Mexico

P. K. Gupta and R. K. Varshney (eds.), *Cereal Genomics II*,
DOI: 10.1007/978-94-007-6401-9_1,

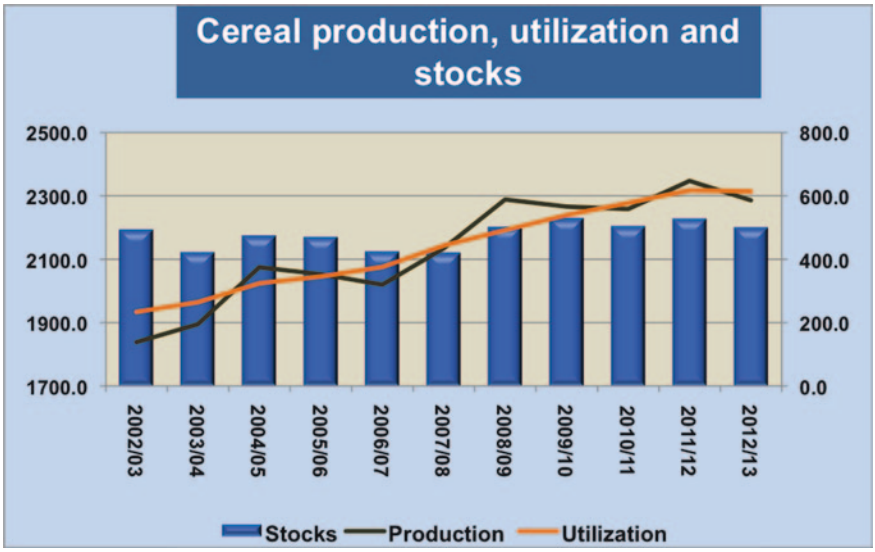

Fig. 1.1 Production, utilization and stocks of cereal grains globally (*source* http://www.fao.org/worldfoodsituation/wfs-home/csdb/en/)

yield (production per unit area) has also slowed down in recent years, so that there is a need to improve further the yield per unit area also. While total annual production of leading cereal crops such as rice, maize, and wheat has increased, the annual production of other cereal crops like barley, oats and rye has declined (Table 1.1). Since the productivity and production of cereals are not stable and may even decrease, depending on weather condition, and because the demand will increase in future due to population pressure, cereal workers can not be complacent and will need to keep on working for enhancing both, yield and production. It is anticipated that this may not be possible with the conventional plant breeding, and genomics-based technologies will have to supplement to meet this challenge. With the shrinking land area, water shortage and the projected climate change, the task is certainly not going to be easy.

The volume '*Cereal Genomics*' that was edited by the authors and published in 2004 contained useful articles written by eminent scientists in different areas of cereal genomics research. It served a useful purpose of making available all information on cereal genomics at one place (Gupta and Varshney 2004), and was well received by cereal workers globally. However, this volume has become out-of-date, since during the last eight years, cereal genomics research progressed at unprecedented pace (Table 1.1). Sufficient additional information has become available making it necessary to have another fresh look on the present status and future possibilities of cereal genomics research. For instance, whole genome sequences became available not only for rice (Goff et al. 2002; Yu et al. 2002), maize (Schnable et al. 2009) and sorghum (Paterson et al. 2009), the three major cereal crops, but also for *Brachypodium distichum* (TIBI 2010), a newly identified model grass species. Significant progress has also been made in sequencing

Table 1.1 A comparative overview of the crops discussed in cereal genomics volume I and II

Cereal species	Biological name	Chromsome number	Genome size (Mbp)[a]	Status year[b]	Yield (tonnes/ hectare)[c]	Production (million tonnes)[c]	Number of ESTs available in public domain[d]	Availability of resources
Barley	*Hordeum vulgare*	$2n = 2x = 14$	5,000	2012	2.6	124	5,01,838	Extensive genetic and physical maps; first glimpse of genome sequence is now available (the International Barley Genome Sequencing Consortium (IBGSC) 2012)
				2004	2.5	132	3,56,848	Both genetic and physical maps (based on translocation breakpoints) were available
Maize	*Zea mays*	$2n = 2x = 20$	2,500	2012	5.2	840	20,19,137	Genome sequencing and re-sequencing for several hundred accessions have become available, hapmap available; GWAS analysis undertaken; GS in progress
				2004	4.3	603	3,93,719	Extensive genetic (including transcript map) and BAC-based physical maps were available
Oats	*Avena sativa*	$2n = 6x = 42$	11,400	2012	2.2	20	25,344	Saturated genetic maps have become available
				2004	1.9	27	574	Genetic maps (but not saturated) were available
Rice	*Oryza sativa*	$2n = 2x = 24$	430	2012	4.4	696	12,52,989	Several drafts for whole genome sequence available, resequencing of several thousand accessions has been completed/in progress, GWAS undertaken; GS in progress
				2004	3.9	580	2,83,935	Extensive genetic (including transcript) and YAC/BAC-based physical maps, as well as 4 drafts of complete genome sequences available
Rye	*Secale cereale*	$2n = 2x = 14$	8,400	2012	2.3	12	9,298	Extensive genetic and physical maps available
				2004	2.2	21	9,194	Genetic maps (but not saturated) available

(continued)

Table 1.1 (continued)

Cereal species	Biological name	Chromsome number	Genome size (Mbp)[a]	Status year[b]	Yield (tonnes/ hectare)[c]	Production (million tonnes)[c]	Number of ESTs available in public domain[d]	Availability of resources
Sorghum	*Sorghum bicolor*	$2n = 2x = 20$	750	2012	1.4	56	2,09,835	Genome sequence has become available and re-sequencing of several accessions initiated
				2004	1.3	55	1,61,813	Integrated cytogenetic, genetic and physical maps were available
Wheat	*Triticum aestivum*	$2n = 6x = 42$	16,000	2012	3.0	654	12,86,173	Extensive genetic and physical map available, Gene space of individual chromosomes captured and genome sequencing in process; GWAS and GS conducted
				2004	2.7	568	5,49,926	Extensive genetic as well as deletion lines-based physical maps were available

[a]As per Bennett and Leitch (1995)
[b]Comparative status of productivity, production and genomic resources between 2004 and 2012 has been shown
[c]As per FAO wesbsite http://apps.fao.org; accessed in 2004 and 2012
[d]As per dbEST release 030504 and 181012- http://www.ncbi.nlm.nih.gov/dbEST/dbEST summary.html

genome or gene space in wheat (Paux et al. 2008; Berkman et al. 2011) and barley (Mayer et al. 2011; IBGSC 2012).

During recent years, another major technology development is the availability of next generation sequencing (NGS), which included the second and third generation high throughput and cost-effective sequencing systems (Thudi et al. 2012). These NGS platforms revolutionized genomics research not only in cereals, but in all living systems including humans, and other higher animals/plants and the microorganisms. By using NGS technologies, genomes of hundreds or thousands of accessions of an individual crop like rice have been generated, thus providing estimates of genome wide diversity and making genome wide association studies (GWAS) more meaningful (Huang et al. 2010; Tian et al. 2011; Zhao et al. 2011; Chia et al. 2012; Hufford et al. 2012). All these developments also created a demand for computational tools to analyse the massive data that was generated at an unprecedented pace. This challenge was met successfully by parallel growth in the field of bioinformatics. These developments have been briefly described in this volume, which is appropriately titled as "*Cereal Genomics II*", so that it supplements our earlier edited volume "*Cereal Genomics*" (Gupta and Varshney 2004). The different aspects covered in this volume are briefly summarized in this introductory chapter.

1.2 Molecular Markers in Cereal Genomics

Although cereal genomics had its birth during 1980s with the development and use of restriction fragment length polymorphism (RFLP) markers, it gained momentum with the development and use of simple sequence repeat (SSR) and amplified fragment length polymorphism (AFLP) markers during 1990s and single nucleotide polymorphism (SNP) and diversity array technology (DArT) markers during the first decade of the present century. However, wide-spread use of these markers in crop breeding programs was not possible due to low throughput and expensive genotyping involved in using these markers. With the availability of microarray technology, during 1990s and early years of the present century, an increased use of microarray-based marker genotyping (particularly for SNPs) was witnessed (Gupta et al. 2008). An updated account of these array-based markers for cereal genomics research is presented by Pushpendra K Gupta (CCS University, Meerut, India) and his former students by (Sachin Rustgi and Reyaz Mir) in Chap. 2 of this volume.

In parallel and following the development of microarray technology, another major development has been the availability of a number of NGS platforms (as mentioned above), which facilitated development and use of high throughput and cost-effective markers like SNPs, SSRs, Insertion Site-Based Polymorphism (ISBPs), Restriction-site Associated DNAs (RADs), Copy-Number Variations (CNVs)/Presence-Absence Variations (PAVs), etc. Dave Edward of the University of Queensland, Australia and Pushpendra Gupta of CCS University, Meerut, India discussed these aspects in Chap. 3 of this volume. The different NGS technologies and their use for study of genetic diversity in cereals are discussed in Chap. 4 by

Eduard Akhunov and colleagues from Kansas State University, USA. This chapter partly overlaps the contents of Chap. 3 in dealing with markers developed using NGS, this overlap being unavoidable in an edited volume.

1.3 Organization and Evolution of Cereal Genomes

In true sense, plant genomics research had its beginning in December 2000, with the publication of the whole genome sequence of the model plant species *Arabidopsis thaliana* (AGI 2000). This was followed by publication of the whole genome sequences of more than a dozen plant species, which included some cereals such as rice (Yu et al. 2002; Goff et al. 2002; IRGSP 2005), maize (Schnable et al. 2009), sorghum (Paterson et al. 2009) and foxtail millet (Bennetzen et al. 2012; Zhang et al. 2012) and a model grass species, *Brachypodium distichum* (TIBI 2010). Available genome sequences of several other plant species (http://genomevolution.org/wiki/index.php/Sequenced_plant_genomes; Plant GDB), which also became available in parallel, were also compared with available cereal genome sequences, thus facilitating further progress in cereal genomics research. In Chap. 5 of this volume, Xi-Yin Wang and Andrew H. Paterson from University of Georgia, USA utilize this information and discuss comparative genomics in cereals. During the study of genomic sequences of cereals, in particular those of corn, it has been recognized that transposable elements constitute a major part of cereal genomes. In Chap. 6 of this volume, Beat Keller and his coworkers from University of Zurich, Switzerland, have discussed the role of transposable elements in shaping cereal genomes.

1.4 Functional Genomics of Cereals

Cereal have also been subjected to functional genomics research, which during the last two decades covered both basic and applied aspects. As a result, not only we understand better the genomes of major cereals and the mechanisms involved in the function of different cereal genes, but we have also utilized information generated from genomics research in producing better transgenic crops, which will give higher yields, sometimes with value addition. In Chap. 7 of this volume, Pankaj Jaiswal and colleagues from Oregon State University, USA have discussed the techniques and bioinformatics involved in functional annotation of cereal genomes. In Chap. 8, Nese Sreenivasulu and his coworkers from IPK, Gatersleben, Germany discussed the different 'omics' approaches involved in functional genomics and their implications for developing a system biology approach for study of the mechanism involved in tolerance against abiotic stress. In Chap. 9, Peter Langridge and his coworkers from Australian Centre for Plant Functional Genomics (ACPFG), Australia discuss the functional genomics of seed development and in Chap. 10, Bikram Gill and his coworkers from Kansas State University, USA discuss the genomics of cereal based functional foods.

1.5 QTL Analysis, Domestication and Molecular Breeding

Another important development in cereal genetics and genomics during the last two decades is the availability of approaches for genetic dissection of complex quantitative traits. This became possible due to the availability of DNA-based molecular markers and statistical tools for analysis of complex traits. These aspects have been discussed in Chap. 11 by Pushpendra K Gupta and his two former students (Pawan Kulwal and Reyaz Mir). QTL analysis also facilitated the study of domestication process, so that domestication syndromes involving the selection of a set of genes have been discovered in all major cereals. These aspects have been discussed in Chap. 12 by Tao Sang from Key Laboratory of Plant Resources, China and Beijing Botanical Garden, China and Jiayang Li from National Center for Plant Gene Research, China. It has also been recognized that precision in phenotyping has been a limitation in studying the genetic architecture of cereal crops, and is absolutely necessary in order to improve the power and resolution of genetic approaches available for genetic dissection of complex traits. This has led to the development of a new discipline called phenomics, which is gaining momentum, so that phenotyping platforms are being established in several countries to facilitate precision in phenotyping. BM Prasanna and his coworkers from International Maize and Wheat Improvement Center (CIMMYT), Mexico discussed the subject of high throughput precision phenotyping for cereal breeding in Chap. 13 of this volume. Chapter 14 of this volume is devoted to molecular breeding written by Yunbi Xu from CIMMYT and his co workers from Chinese Academy of Agricultural Sciences, China and CIMMYT.

1.6 Summary and Outlook

In summary, this volume *Cereal Genomics II* with 14 chapters (including this introductory chapter) provides a glimpse of the advances in cereals genomics research made during the last eight years, that elapsed between now and the year 2004, when our earlier volume, *Cereal Genomics* was published. This volume presents state-of-art of cereal genomics and its utilization in both basic studies such as comparative genomics and functional genomics as well as applied aspects like QTL mapping and molecular breeding. Keeping in view the information that became available during the last one decade, one can certainly foresee an exciting period that lies ahead for cereal researchers globally, particularly because the large and complex cereal genomes of barley and wheat will also be fully sequenced within the next 2–3 years. Molecular mapping and breeding approaches will be shifting from marker-based genotyping to sequencing-based genotyping. Comparative genomics will be moving from comparison of genomes of two or more species to comparison of genomes of hundreds to thousands accessions of the same species. While data generation for even complex genomes of cereal species is expected to become routine, analysis and interpretation of data will be a challenge for both cereal biologists as well as for those involved in applied cereal genomics research. This will be facilitated through advances in

bioinformatics including high-throughput data analysis and cloud computing, which will help make further advances in this fascinating area of cereal genomics.

References

Arabidopsis Genome Initiative (AGI) (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796–815

Bennett MD, Leitch (1995) Nuclear DNA amounts in angiosperms. Ann Bot 76:113–176

Bennetzen JL, Schmutz J, Wang H, Percifield R, Hawkins J, Pontaroli AC, Estep M, Feng L, Vaughn JN, Grimwood J, Jenkins J, Barry K, Lindquist E, Hellsten U, Deshpande S, Wang X, Wu X, Mitros T, Triplett J, Yang X, Ye CY, Mauro-Herrera M, Wang L, Li P, Sharma M, Sharma R, Ronald PC, Panaud O, Kellogg EA, Brutnell TP, Doust AN, Tuskan GA, Rokhsar D, Devos KM (2012) Reference genome sequence of the model plant *Setaria*. Nat Biotechnol 30:555–561

Berkman BJ, Skarshewski A, Lorenc MT, Lai K, Duran C, Ling EYS, Stiller J, Smits L, Imelfort M, Manoli S, McKenzie M, Kubalakova M, Simkova H, Batley J, Fleury D, Dolezel J, Edwards D (2011) Sequencing and assembly of low copy and genic regions of isolated *Triticum aestivum* chromosome arm 7DS. Plant Biotechnol J 9:768–775

Chia JM, Song C, Bradbury PJ, Costich D, de Leon N, Doebley J, Elshire RJ, Gaut B, Geller L, Glaubitz JC, Gore M, Guill KE, Holland J, Hufford MB, Lai J, Li M, Liu X, Lu Y, McCombie R, Nelson R, Poland J, Prasanna BM, Pyhäjärvi T, Rong T, Sekhon RS, Sun Q, Tenaillon MI, Tian F, Wang J, Xu X, Zhang Z, Kaeppler SM, Ross-Ibarra J, McMullen MD, Buckler ES, Zhang G, Xu Y, Ware D (2012) Maize hapmap2 identifies extant variation from a genome in flux. Nat Genet 44:803–807

Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, Hadley D, Hutchison D, Martin C, Katagiri F, Lange BM, Moughamer T, Xia Y, Budworth P, Zhong J, Miguel T, Paszkowski U, Zhang S, Colbert M, Sun WL, Chen L, Cooper B, Park S, Wood TC, Mao L, Quail P, Wing R, Dean R, Yu Y, Zharkikh A, Shen R, Sahasrabudhe S, Thomas A, Cannings R, Gutin A, Pruss D, Reid J, Tavtigian S, Mitchell J, Eldredge G, Scholl T, Miller RM, Bhatnagar S, Adey N, Rubano T, Tusneem N, Robinson R, Feldhaus J, Macalma T, Oliphant A, Briggs S (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. japonica). Science 296:92–100

Gupta PK, Varshney RK (2004) *Cereal Genomics*. Kluwer Academic Publishers, The Netherlands, pp 1–369

Gupta PK, Rustgi S, Mir RR (2008) Array-based high-throughput DNA markers for crop improvement. Heredity 101:5–18

Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z, Li M, Fan D, Guo Y, Wang A, Wang L, Deng L, Li W, Lu Y, Weng Q, Liu K, Huang T, Zhou T, Jing Y, Li W, Lin Z, Buckler ES, Qian Q, Zhang QF, Li J, Han B (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. Nat Genet 42:961–967

Hufford MB, Xun X, van Heerwaarden J, Pyhäjärvi T, Chia JM, Cartwright RA, Elshire RJ, Glaubitz JC, Guill KE, Kaeppler SM, Lai J, Morrell PL, Shannon LM, Song C, Springer NM, Swanson-Wagner RA, Tiffin P, Wang J, Zhang G, Doebley J, McMullen MD, Ware D, Buckler ES, Yang S, Ross-Ibarra J (2012) Comparative population genomics of maize domestication and improvement. Nat Genet 44:808–811

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

Mayer KF, Martis M, Hedley PE, Simková H, Liu H, Morris JA, Steuernagel B, Taudien S, Roessner S, Gundlach H, Kubaláková M, Suchánková P, Murat F, Felder M, Nussbaumer T, Graner A, Salse J, Endo T, Sakai H, Tanaka T, Itoh T, Sato K, Platzer M, Matsumoto T, Scholz U, Dolezel J, Waugh R, Stein N (2011) Unlocking the barley genome by chromosomal and comparative genomics. Plant Cell 23:1249–1263

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T,

Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, Mehboob-ur-Rahman Ware D, Westhoff P, Mayer KF, Messing J, Rokhsar DS (2009) The *Sorghum bicolor* genome and the diversification of grasses. Nature 457:551–556

Paux E, Sourdille P, Salse J, Saintenac C, Choulet F, Leroy P, Korol A, Michalak M, Kianian S, Spielmeyer W, Lagudah E, Somers D, Kilian A, Alaux M, Vautrin S, Bergès H, Eversole K, Appels R, Safar J, Simkova H, Dolezel J, Bernard M, Feuillet C (2008) A physical map of the 1-gigabase bread wheat chromosome 3B. Science 322:101–104

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, Minx P, Reily AD, Courtney L, Kruchowski SS, Tomlinson C, Strong C, Delehaunty K, Fronick C, Courtney B, Rock SM, Belter E, Du F, Kim K, Abbott RM, Cotton M, Levy A, Marchetto P, Ochoa K, Jackson SM, Gillam B, Chen W, Yan L, Higginbotham J, Cardenas M, Waligorski J, Applebaum E, Phelps L, Falcone J, Kanchi K, Thane T, Scimone A, Thane N, Henke J, Wang T, Ruppert J, Shah N, Rotter K, Hodges J, Ingenthron E, Cordes M, Kohlberg S, Sgro J, Delgado B, Mead K, Chinwalla A, Leonard S, Crouse K, Collura K, Kudrna D, Currie J, He R, Angelova A, Rajasekar S, Mueller T, Lomeli R, Scara G, Ko A, Delaney K, Wissotski M, Lopez G, Campos D, Braidotti M, Ashley E, Golser W, Kim H, Lee S, Lin J, Dujmic Z, Kim W, Talag J, Zuccolo A, Fan C, Sebastian A, Kramer M, Spiegel L, Nascimento L, Zutavern T, Miller B, Ambroise C, Muller S, Spooner W, Narechania A, Ren L, Wei S, Kumari S, Faga B, Levy MJ, McMahan L, Van Buren P, Vaughn MW, Ying K, Yeh CT, Emrich SJ, Jia Y, Kalyanaraman A, Hsia AP, Barbazuk WB, Baucom RS, Brutnell TP, Carpita NC, Chaparro C, Chia JM, Deragon JM, Estill JC, Fu Y, Jeddeloh JA, Han Y, Lee H, Li P, Lisch DR, Liu S, Liu Z, Nagel DH, McCann MC, SanMiguel P, Myers AM, Nettleton D, Nguyen J, Penning BW, Ponnala L, Schneider KL, Schwartz DC, Sharma A, Soderlund C, Springer NM, Sun Q, Wang H, Waterman M, Westerman R, Wolfgruber TK, Yang L, Yu Y, Zhang L, Zhou S, Zhu Q, Bennetzen JL, Dawe RK, Jiang J, Jiang N, Presting GG, Wessler SR, Aluru S, Martienssen RA, Clifton SW, McCombie WR, Wing RA, Wilson RK (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112–1115

The International Barley Genome Sequencing Consortium (IBGSC) (2012) A physical, genetic and functional sequence assembly of the barley genome. Nature doi:10.1038/nature11543

The International Brachypodium Initiative (TIBI) (2010) Genome sequencing and analysis of the model grass *Brachypodium distachyon*. Nature 463:763–768

Thudi M, Li Y, Jackson SA, May GD, Varshney RK (2012) Current state-of-art of sequencing technologies for plant genomics research. Brief Funct Genomics 11:3–11

Tian F, Bradbury PJ, Brown PJ, Hung H, Sun Q, Flint-Garcia S, Rocheford TR, McMullen MD, Holland JB, Buckler ES (2011) Genome-wide association study of leaf architecture in the maize nested association mapping population. Nat Genet 43:159–162

Yu J, Hu S, Wang J, Wong GK, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, Cao M, Liu J, Sun J, Tang J, Chen Y, Huang X, Lin W, Ye C, Tong W, Cong L, Geng J, Han Y, Li L, Li W, Hu G, Huang X, Li W, Li J, Liu Z, Li L, Liu J, Qi Q, Liu J, Li L, Li T, Wang X, Lu H, Wu T, Zhu M, Ni P, Han H, Dong W, Ren X, Feng X, Cui P, Li X, Wang H, Xu X, Zhai W, Xu Z, Zhang J, He S, Zhang J, Xu J, Zhang K, Zheng X, Dong J, Zeng W, Tao L, Ye J, Tan J, Ren X, Chen X, He J, Liu D, Tian W, Tian C, Xia H, Bao Q, Li G, Gao H, Cao T, Wang J, Zhao W, Li P, Chen W, Wang X, Zhang Y, Hu J, Wang J, Liu S, Yang J, Zhang G, Xiong Y, Li Z, Mao L, Zhou C, Zhu Z, Chen R, Hao B, Zheng W, Chen S, Guo W, Li G, Liu S, Tao M, Wang J, Zhu L, Yuan L, Yang H (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). Science 296:79–92

Zhang G, Liu X, Quan Z, Cheng S, Xu X, Pan S, Xie M, Zeng P, Yue Z, Wang W, Tao Y, Bian C, Han C, Xia Q, Peng X, Cao R, Yang X, Zhan D, Hu J, Zhang Y, Li H, Li H, Li N, Wang J, Wang C, Wang R, Guo T, Cai Y, Liu C, Xiang H, Shi Q, Huang P, Chen Q, Li Y, Wang J, Zhao Z, Wang J (2012) Genome sequence of foxtail millet (*Setaria italica*) provides insights into grass evolution and biofuel potential. Nat Biotechnol 30:549–554

Zhao K, Tung CW, Eizenga GC, Wright MH, Ali ML, Price AH, Norton GJ, Islam MR, Reynolds A, Mezey J, McClung AM, Bustamante CD, McCouch SR (2011) Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. Nat Commun 2:467

Chapter 2
Array-Based High-Throughput DNA Markers and Genotyping Platforms for Cereal Genetics and Genomics

Pushpendra K. Gupta, Sachin Rustgi and Reyazul R. Mir

2.1 Introduction

During the last three decades, DNA-based molecular markers have become indispensible tools for detailed genetic analysis and molecular breeding in crop plants. Newer marker systems have been developed at regular intervals during last more than 20 years, and have been discussed by us in a series of articles (Gupta et al. 1996, 1999a, b; Gupta and Varshney 2000; Gupta et al. 2001, 2002; Gupta and Rustgi 2004; Gupta et al. 2008). For cereals, these marker systems and the corresponding molecular maps developed until eight years ago were described in an article included in our earlier edited volume '*Cereal Genomics*' (Gupta and Varshney 2004). A number of other reviews on molecular markers and their uses also appeared after the publication of our book '*Cereal Genomics*'. In particular, a book "*Molecular Marker Systems in Plant Breeding and Crop Improvement*" edited by Lörz and Wenzel (2005) contained a number of useful articles. Based on these earlier reviews, it may be recalled that the marker systems used in 1980s, 1990s and in the early years of the present century were largely either hybridization-based (without PCR; e.g., RFLPs) or PCR-based (involving slab-gel or capillary-electrophoresis based separation of PCR products). In most cases, genotyping was performed

P. K. Gupta (✉)
Department of Genetics and Plant Breeding, Ch Charan Singh University, Meerut 250004, U.P, India
e-mail: pkgupta36@gmail.com

S. Rustgi
Department of Crop and SoilSciences, Washington State University, Pullman 99164, W.A, USA

R. R. Mir
Division of Plant Breeding and Genetics, Shere-Kashmir University of Agricultural Sciences and Technology of Jammu (SKUAST-J), Chatha 180 009, Jammu, India

P. K. Gupta and R. K. Varshney (eds.), *Cereal Genomics II*,
DOI: 10.1007/978-94-007-6401-9_2,

for individual markers, although multiple loading, multiplexing and multi mixing did improve the speed of genotyping. But this status of molecular marker technology was still far short of the kind of high-throughput that is required for genotyping of either thousands of plants with few markers or a limited number of plants with millions of individual markers. This level of high throughput is needed not only for high precision in the detection of QTLs/genes involving linkage-cum-association mapping, but also for their use in improving the efficiency of large plant breeding programs. Map-based cloning of genes/QTLs involving use of large segregating populations is another area of research, where this high-throughput is required.

During the last eight years, i.e. after the publication of the edited volume "*Cereal Genomics*" in 2004, there has actually been a revolution in the development of not only the new marker systems, but also in the development of corresponding high throughput-genotyping platforms. This led to the extensive use of molecular markers in gene discovery/cloning and molecular breeding studies in cost/time-effective fashion, thus making them routine in most laboratories around the world. Further, this has become possible particularly due to the adoption of microarrays/chips, real-time detection methods and next generation sequencing (NGS) technologies for discovery and detection of markers, which is a pre-requisite for identification of genes/QTLs associated with specific traits of interest. The marker-trait associations (MTAs) identified for a variety of simple and complex traits in a number of crops have also been effectively utilized for molecular breeding leading to the release of improved cultivars, particularly in cereals including wheat (Gupta et al. 2010a, b), rice (Singh et al. 2011) and maize (Prasanna and Hoisington 2003). The importance of molecular breeding in developing countries has also been repeatedly emphasized (Ribault et al. 2010; Anthony and Ferroni 2011).

The choice for marker systems that are now being increasingly utilized has also shifted from the first and second generation marker systems including RFLPs, RAPDs, SSRs and AFLPs to the third and the fourth generation marker systems, which include SNPs, DArT, TDMs (including SFPs), ISBP markers (Gupta et al. 2008; Potokina et al. 2008; Paux et al. 2008, 2010, 2012), and CNVs/PAVs (Springer et al. 2009; Belo et al. 2010). Most of these latest marker systems make use of microarrays, which is the subject of this brief review. In fact, with the availability of NGS technology, and due to the possible low-cost resequencing of whole genomes in crops like rice (with a small genome), it is now possible to resolve recombination breakpoints within an average length of 40 kb. It has been estimated thus that in comparison with the PCR-based markers, map construction using sequencing based methods is now 20 × faster in data collection and has 35 × higher precision in determination of recombination breakpoints (Huang et al. 2009; Wang et al. 2011; also see, Seifollah et al.2013 in this volume).

The array-based marker systems became available more than five years ago, and were discussed by us in an earlier review (Gupta et al. 2008). In this earlier review, we had briefly described four types of marker systems (SNPs, SFPs, DArT and RAD), the corresponding genotyping platforms and their effective use in a variety of plant species. However, during the last few years, after the publication of our above review, the array-based genotyping platforms have been increasingly used in plants including cereals, thus establishing their utility in molecular breeding programs.

Array-based comparative genomic hybridization (aCGH) has also been used for detection of structural variations (SVs) including copy number variations (CNVs), presence-absence variations (PAVs) and insertions/deletions (InDels). Newer and improved array-based genotyping platforms with enhanced throughput and reduced genotyping cost are also being regularly developed (Fig. 2.1). Thus, these array-based platforms will perhaps continue to remain important for high-throughput marker discovery and genotyping, although low-cost NGS technologies are also being used now in parallel with the development of these array-based marker systems (see Seifollah et al.2013 in this volume). In this chapter, we wish to describe first the different array types and the principles/methods involved in these array-based marker systems, then discuss the results obtained in cereals during the last few years using the array-based marker genotyping platforms, and finally discuss the utility of these markers and the corresponding genotyping platforms in future molecular crop breeding programs. In order to give a comprehensive picture of the array-based marker systems, some of the information available in our earlier review will also be included in this chapter, this repetition being unavoidable. The sequencing-based marker systems and the genotyping platforms involving NGS technologies will not be included in this chapter, since these are being covered in the next chapter of this book.

2.2 Array Types for Marker Development and Genotyping

As mentioned above, microarrays were initially developed to facilitate large-scale screening of whole genomes/transcriptomes or a few genes/proteins in any living system including crop plants like cereals. These microarrays have undergone a variety of modifications in the original design and concept depending upon the aims and

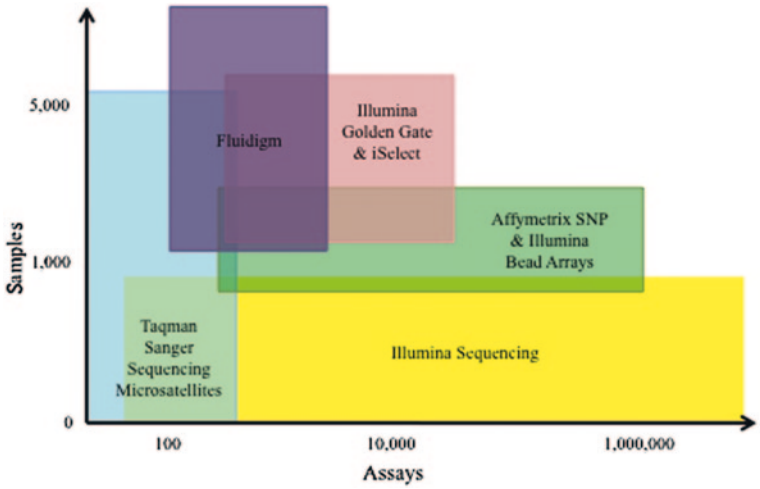

Fig. 2.1 Different platforms for genotyping, showing their relative high-throughput in terms of number of *samples* and *assays* that can be used in a single run (modified from BioScience LifeSciences)

objectives of using them, and also to achieve high throughput and cost-effectiveness. These microarrays were initially developed and continue to be used as planar arrays, where nucleic acid oligomers/probes or beads with oligonucleotides were immobilized on a flat solid surface mostly represented by a glass slide, a silicon wafer or a patterned planar substrate (for reviews, see Gupta et al. 1999a, b; Venkatasubbarao 2004). Later, microsphere-based suspension arrays were developed, where encoded microspheres (with tens of millions of particles per milliliter of suspension) with distinct optical properties were used as solid supports for biomolecules and the suspension arrays analyzed through flow cytometry; this also includes Luminex xMAP™ technology developed for rapid and high-throughput multiplexed genotyping (for reviews, see Nolan and Sklar 2002; Dunbar 2006).

Fluidigm Dynamic Arrays™ is another technology, where microfluidics with a miniaturized complex fluid-handling system is used on the chip, making it a 'lab-on-a-chip device'. This is actually a 'reagent in-data out' platform, where PCR reactions also take place on the chip, followed by identification of PCR products for genotyping. As a proof of concept, using Fluidigm's new SNPtype™ system a genotyping panel each of 48 SNPs was designed and validated respectively on a set of 94 rice (Ilic et al. 2011) and 46 cocoa (Ilic et al. 2012) accessions. Similarly, a panel of 96 SNPs was validated on a set of 23 *Bromus tectorum* accessions using Fluidigm and KASPar SNP genotyping platforms (Merrill et al. 2011).

Ultra-high-throughput nano-arrays/nano-chips were also developed and used for screening human genomes (Chen and Li 2007). Biomolecule conjugated quantum dots (QDs), which are semiconductor fluorescent nanocrystals, were also used for the assembly of microarrays to improve sensitivity of these microarrays (Chan and Nie 1998; Ioannou and Griffin 2010). Latest in the series are the Ion Torrent Chips, which employ a unique combination of fluidics, micromachining, and semiconductor technology to increase the high throughput further. For instance, recently an Ion Torrent Chip with a capacity to support up to 1.2 million DNA-testing wells was launched and used successfully to demonstrate sequencing of a bacterial genome in amazingly short period of time, within just two hours (Zakaib 2011). A next generation 11 million well Ion Torrent Chip was also under testing, which will allow the use of this ultra-high throughput technology in marker development and genotyping in foreseeable future.

However, not all of the above array technologies have been used in crop plants (particularly in cereals), so that we will restrict our discussion in this chapter to only those array technologies, which have been used for cereal genomes. However, a brief reference to other technologies will also be made, whenever possible and necessary.

2.2.1 DNA Chips/Microarrays/High-Density Oligonucleotide Arrays (Ordered Arrays)

The high-density oligonucleotide arrays (also described as DNA chips or microarrays) differ either in their design (featured chips *vs* tiling arrays) or in their production (spotting *vs* in-situ synthesis). Therefore, these arrays could be

classified into following two types: (1) featured chips (carrying gene sequences, or other polymorphic sequences known for an organism), and (2) tiling arrays that virtually represent the whole genome of an organism.

Among featured chips, the gene chips carry either cDNA-PCR products or long/short oligonucleotide probes (70–100 mers or 25 mers) printed on coated glass slides or silicon wafers. In featured chips with short oligonucleotide probes, for each of a number of genes, 20 pairs of oligonucleotide probes are selected from the exons located near the 3′-ends. The number of features on a single featured chip can also vary from 10,000 to > 6 million (Liu 2007). The tiling arrays, on the other hand, virtually cover the entire genome.

Although the above DNA chips/microarrays have generally been extensively used for expression analysis, but emphasis of the current review is specifically on their use in the detection/genotyping of SNPs/InDels, TDMs (GEMs and SFPs), ELPs, and CNVs/PAVs at genome-wide scale. Other featured arrays (e.g., TAM, DArT, RAD, aCGH) were also used to identify DNA polymorphisms.

Whole Genome High Density Resequencing Microarrays for SNP Discovery

A variety of whole genome microarrays have been developed that are suitable for both discovery of markers (e.g., SNPs) and genotyping. These microarrays are often based on oligonucleotides, which may be (1) partially overlapping or non-overlapping and tiled end to end, or (2) spaced at regular intervals to interrogate the entire genome without annotation bias. The whole genome microarrays also include biased expression arrays, splice-junction arrays, or exon-scanning arrays, when one wishes to scan specific regions at the whole genome level. Lastly, these microarrays include tiling resequencing arrays, where each nucleotide of the reference genomic DNA sequence is represented by a set of eight oligonucleotide probes (four possible nucleotide for each strand). Some of these whole genome microarrays are shown in Fig. 2.2.

High-density, oligonucleotide microarrays are often used for detection of genome-wide DNA polymorphisms (e.g., Chee et al. 1996; Patil et al. 2001; Hinds et al. 2005). These microarrays, follow a 1-bp tiling path to query each base of the genome relative to a known reference sequence, and are therefore, described as resequencing arrays. Each base is interrogated with eight features that consist of forward and reverse strand 25-mer oligonucleotide quartets. Within a quartet, oligonucleotides are identical to the reference sequence except at the central position, where each of the four possible bases is represented in four oligonucleotides. When hybridized to labeled genomic DNA, the highest signal intensity is expected for the perfect match, thereby predicting the base in the corresponding target DNA sample. Large-scale polymorphism discovery using such resequencing arrays was first performed in humans, identifying a large fraction of common single nucleotide polymorphisms (SNPs) in the global population (Patil et al. 2001; Hinds et al. 2005). These resequencing arrays have also been used in plants including

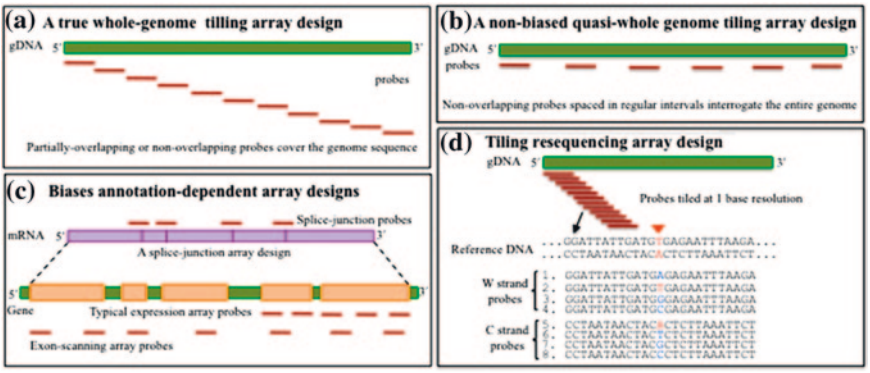

Fig. 2.2 A comparison of different whole-genome array designs. Unbiased whole-genome tiling array designs (**a** and **b**) contain oligonucleotide probes representing the entire genomic sequence. Probes may be **a** partially overlapping or nonoverlapping and tiled end to end or may be (**b**) spaced at regular intervals to interrogate the entire genome without annotation bias. **c** Other biased whole-genome array designs such as typical expression arrays, splice-junction arrays, and exon-scanning arrays contain only oligonucleotide probes for the known and predicted features of a genome. **d** Tiling resequencing arrays represent each nucleotide of the reference genomic DNA sequence with a set of eight oligonucleotide probes

Arabidopsis thaliana (Clark et al. 2007) and *Oryza sativa* (McNally et al. 2009). Re-sequencing arrays are now available for several plant systems, although high error rate (~ 50 %) makes them unreliable for identification of individual SNPs.

Gene-Based Microarrays (GeneChips) for TDMs (ELPs/GEM and SFPs)/CNVs

GeneChips have been developed by Affymetrix, Nimbelgen and Agilent in several plant species. In Affymetrix chips, a large number of genes are each represented by 11 perfect match (PF) and 11 mismatch (MM) 25-mer oligonucleotides constituting a probe set, so that thousands of such probe sets are included in a GeneChip. In contrast, Nimbelgen and Agilent chips include two to seven of 60-mer oligonucleotide probes representing each gene, although these chips have limited application in polymorphism survey in the form of markers like SFPs. The two oligonucleotides (PM and MM) in a pair differ in only one base. In wheat, as many as 55,052 transcripts (spanning all the 21 chromosomes) were used for this purpose (Bernardo et al. 2009). Genomic DNA is often hybridized to such GeneChip arrays, and a difference in hybridization intensity with two oligos (PM and MM) is recorded as a DNA polymorphism often described as an SFP. If cRNA is used for hybridization (in order to minimize the problem of large genome size and repetitive DNA), the polymorphisms recorded are described as transcript-derived markers (TDMs). Such GeneChips have been developed and used in maize, barley, wheat and rice (see Gupta et al. 2008 for details).

Microarrays for SNP Genotyping (SNP Chips)

SNP chips, sometimes also described as 'variant detector arrays' (VDAs) have been developed using different approaches. In one of these approaches, four oligos that differ only at the last position are used per SNP. To determine which alleles are present, genomic DNA from an individual is isolated, fragmented, tagged with a fluorescent dye, and applied to the chip. The genomic DNA fragments anneal only to those oligos to which they are perfectly complementary, which allows SNP genotyping through computer-aided identification of the position of fluorescent tags.

In another approach, the oligonucleotide on the chip may stop one base before the variable site, and typing relies on allele-specific primer extension. A DNA sample is added, which gets stuck onto the chip through base-pairing and is used as a template for DNA synthesis, with the immobilized oligonucleotide being used as a primer. The four nucleotides, containing different fluorescent labels, are added along with DNA polymerase. The incorporated base, which is inserted opposite to the polymorphic site on the template, is identified by the nature of its fluorescent signal. In a variation of this technique, the added nucleotide is identified not by a fluorescent label but by mass spectrometry, as done in MassARRAY used in Sequenom platform, launched by SEQUENOM. This platform makes use of a 384 Spectro CHIP, where PCR products are automatically transferred and utilized for mass spectrometry. However, Mass Array technology has not been used for cereal genomics on any large scale.

Illumina's SNP chip is based on Bead Array Technology (utilized in Illumina's iScan System), where silica beads (3-micron in size) self assemble in micro wells on either fiber optic bundles or planar silica slides. Each silica bead is covered with hundreds of thousands of copies of a specific oligonucleotide acting as the capture sequences in one of the several available Illumina's assays.

2.2.2 Diversity Arrays for DArT Markers

A diversity array is necessarily crop-specific and consists of a large number of diverse anonymous clones. It is prepared by a proprietary method used for selection of diverse clones of genomic DNA from a sample of pooled DNA derived from a number of diverse accessions of the crop under study. These clones are characterized through sequencing and are mapped on the genome. The arrays are used for development of markers, where polymorphic DArT fragments between any two or more genotypes can be identified and used for a variety of studies.

2.2.3 Arrays of PCR Products Spotted on a Slide to be Scanned by Probes: Tagged Array Markers (TAMs)

In the approaches discussed above, the 'target' sequence is interrogated by each of a number of 'probes' arrayed on the chip. In tagged array markers (TAMs),

the situation is reversed, where the target sequences in the form of PCR products (obtained using one biotin labeled primer shared by both alleles, and two unique primers carrying allele-specific oligonucleotide tags) are spotted directly from 96 or 384-well PCR plates onto streptavidin-coated glass microarray slides. The tags attached to the primers remain single stranded due a C-18 linker between the allele specific sequence and the tag. The target sequences (allele-specific PCR products carrying unique tags) are then hybridized with fluorescent-labeled detector probes to identify alleles represented by each of the targeted sequences that are arrayed. In this way, thousands of tagged samples that are arrayed can be genotyped for a few markers in a single experiment, making it particularly useful for screening large populations for few important markers. The method was developed to score retrotransposon-based insertion polymorphism (RBIP) markers, but can also be used to score SNPs (Flavell et al. 2003).

2.2.4 Arrays for Comparative Genomic Hybridization (aCGH)

The technique of comparative genomic hybridization (CGH) was originally developed in early 1990s and involved competitive hybridization of two differentially labeled genomic DNA samples (a test and a control) on to metaphase chromosomes. The fluorescent signal intensity of the labeled test DNA relative to that of the reference DNA could then be linearly plotted across each chromosome, facilitating identification of copy number variations (Kallioniemi et al. 1992). However, the resolution of this CGH technology was limited to alterations of approximately 5–10 Mb (Lichter et al. 2000; Kirchhoff et al. 1998), and could identify only microscopic structural and numerical alterations in chromosomes (including duplications/deletions and aneuploidy).

In order to improve resolution, microarrays were designed and used for the so-called array comparative genomic hybridization (aCGH), which allowed resolution in the range of 1 Kb–3 Mb (Lucito et al. 2003). These microarrays resemble those generally designed for the detection of SFPs, except that the oligonucleotides used for aCGH are longer (generally 50–85 bp, but could be hundreds of kb as in the BACs). The spotted oligos in microarrays used for aCGH generally represent the genomic regions of interest. Digital imaging systems are used to capture and quantify the relative fluorescence intensities of the labeled DNA that is hybridized to each target. The fluorescence ratio of the test and reference hybridization signals is determined at different positions along the genome, and it provides information on the relative copy number of sequences in the test genome as compared to the reference genome.

Currently, Roche NimbleGen and Agilent Technologies are the major suppliers of whole-genome array CGH platforms. The oligonucleotides in a CGH microarray may number up to 2–4 million (2–4 M) in case of Roche/NimbleGen and up to 1 M in case of Agilent Technologies (Alkan et al. 2011).

2.3 Array-Based Molecular Markers: Classification

In our previous review on array-based markers, we grouped array-based marker systems into four classes including SNP, SFP, DArT and RAD markers (Gupta et al. 2008). To these four classes, we may add structural variations (SVs) including copy number variations (CNVs), presence-absence variations (PAVs) and insertions/deletions (InDels). At the molecular level, each of these individual array-based marker types represents either nucleotide substitutions or duplications/InDels. In view of the nature of genotyping/detection platforms developed during the last few years, and their throughput and suitability for various applications, we classified array-based markers into two classes, (1) those based on genotyping procedure and (2) those based on the level of throughput. Later in this chapter, for a more detailed account of array-based markers, we will follow the classification based on the technique (procedure) involved.

In the first classification, which is based on the technology involved, the array-based marker platforms are placed into the following three major groups: (1) hybridization-based platforms, where SNP-specific chips or microarrays, developed for CGH, are used for hybridization; (2) real-time detection-based platforms, where an array of samples to be genotyped is used to provide templates for real-time PCR, and (3) the platforms involving both hybridization and real-time detection.

The markers involving hybridization based platforms for genotyping can be further classified as follows: (1) transcript derived markers [TDMs, including expression level polymorphisms (ELPs) or gene expression markers (GEMs) and single feature polymorphisms (SFPs)] (Potokina et al. 2008), (2) diversity array technology (DArT) markers (Jaccoud et al. 2001; Wenzl et al. 2004), (3) tagged microarray markers (TAMs) (Jing et al. 2007), (4) restriction site-associated DNA (RAD) markers (Miller et al. 2007a, b), (5) GoldenGate SNP genotyping assays (Steermers and Gunderson 2007); and (6) copy number variations (CNVs) and presence-absence variations (PAVs). Similarly, the real-time detection-based platforms include (a) *KASPar SNP* genotyping system (http://www.kbioscience.co.uk/reagents/KASP.html), and (b) high-resolution melting (HRM) curve analysis (Hoffmann et al. 2007). Illumina's GoldenGate and Infinium assays, on the other hand, rely on a 'hybrid technology' involving both hybridization and real-time detection (Steermers and Gunderson 2007).

In the second classification that is based on the level of throughput and suitability for different applications, the assays involving array-based markers are placed into the following three groups: (a) assays involving genotyping an individual with an array of marker-specific probes, (b) assays involving genotyping an array of individuals with a single marker, and (c) the flexible platforms involving both of the above situations. Based on this classification, hybridization-based platforms (including SFPs, DArT and RAD) represent the first group, real-time detection-based platforms (viz. KASPar and HRM) represent the second group, and Illumina's GoldenGate and Infinium assays represent the third group (for details, see later).

2.4 Array-Based Molecular Markers: Principles and Methods

As mentioned above, a majority of array-based markers are hybridization-based (sometimes including real-time detection method), which will be described in relatively greater detail in this section.

2.4.1 Microarray-Based Markers Involving Hybridization

The microarray-based markers rely on hybridization of genomic DNA, cDNA, cRNA and/or mRNA to GeneChips, oligonucleotide tiling arrays, diversity arrays, and/or glass microarray slides. Even though, these methods were quite successful in detecting/determining polymorphism at the genome-wide scale, they suffer from several associated limitations including the following: (1) High background noise owing to cross-hybridization, and a limited dynamic range of detection because of both background and saturation of signals. (2) High experimental cost, since many replications are required per experiment to increase statistical confidence of the observations. (3) Dependence upon the existing sequence (genome/transcriptome) information. (4) Requirement of comparing expression levels across different experiments, which is often difficult (Wang et al. 2009). These hybridization-based markers (except CNVs and PAVs) were described in greater detail in our previous review (Gupta et al. 2008) and therefore, will be only briefly summarized here.

Transcript Derived Markers (TDMs)

The TDMs constitute a major class of microarray-based markers, detected upon hybridization of transcripts/cDNAs on microarrays or GeneChips. These markers were successfully used for cereals including maize, barley and wheat (for details, see Sect. 2.6.3). The TDMs are gene-specific markers (due to SNPs/InDels) and include both ELPs/GEMs and SFPs. ELPs/GEMs represent expression level differences (total absence to differences in transcript abundance) recorded on chips in the form of difference in signal intensity observed for different samples under study. But the observed signal intensity for a particular sample is consistently shown by all the features (oligos) representing a particular gene on the chip. In contrast, SFPs represent differences in hybridization intensity observed between two samples, denoted by only one of the many features representing a gene on the microarray (c.f. Gupta et al. 2008). Both kinds of molecular markers (ELPs/GEMs and SFPs), if applied on a mapping population, will allow grouping a population into two discrete classes (presence or absence of expression), in contrast to the cases where the population shows a continuous distribution for the transcript abundance of a particular gene, which can be recorded as an e-trait to be mapped as expression QTL(s).

Diversity Array Technology (DArT)

DArT is a high-throughput microarray hybridization-based technique that allows simultaneous typing of several hundred polymorphic loci spread over the entire genome without any prior sequence information about these loci (Jaccoud et al. 2001; Wenzl et al. 2004). DArT is an extension of 'garden blots' prepared using the genomic DNA of different plant species. DArT involves development of a 'discovery array', which is developed from the metagenome (pool of genomes representing the diverse germplasm of interest), that was subjected to complexity reduction to reduce the level of repetitive DNA, since repetitive sequences interfere with DArT assays (Kilian et al. 2003, 2005). For a 'discovery array', individual clones from a genomic representation library are amplified and spotted onto glass slides (www.diversityarrays.com). Labelled genomic representations of individual genomes that were earlier included in the metagenome pool, were then hybridized to this 'discovery array', and polymorphic clones (called DArT markers) thus detected are assembled into a 'genotyping array' for routine genotyping work. These markers are biallelic and dominant (presence vs absence) or co-dominant (two doses vs one dose vs absent) in nature, and were successfully used in rice, barley, wheat and maize (see Sect. 2.6.2 for details).

Tagged Array Markers (TAMs)

TAMs allow high-throughput distinction between predicted alternative PCR products. Typically, the method is used as a molecular marker approach for determining the allelic states of individual single nucleotide polymorphisms (SNPs) or insertions/deletions (InDels) in multiple individuals. Biotin-labeled PCR (unpurified), products are spotted, onto a streptavidin-coated glass slide and the alternative products are distinguished by hybridization to fluorescent detector oligonucleotides that recognize corresponding allele-specific tags on the PCR primers. There are several advantages of this method, which include high throughput (thousands of PCRs are analyzed per slide), flexibility of scoring (any combination, from a single marker in thousands of samples to thousands of markers in a single sample, can be analyzed) and flexibility of scale (any experimental scale, from a small lab setting to a large project). The TAM technology was initially adopted in pea to genotype retrotransposon-based insertional polymorphism (RBIP) markers on a dot assay, which was later made fully automated for handling thousands of samples. The basic RBIP method has been developed for high-throughput applications by replacing gel electrophoresis with array hybridization to a filter (Flavell et al. 1998; Jing et al. 2007).

CNVs and PAVs

Copy number variations (CNVs) and presence-absence variations (PAVs) are the latest markers developed recently. They have been extensively used in humans and are now being used in plants also. These markers are detected through the use of

microarrays, which are specially developed for each individual plant species, and are then used for comparative genomic hybridization (CGH). The genomic DNAs of test sample and the reference sample are differentially labeled (with C3 and C5) and are then hybridized on to the CGH microarray. The ratio of the fluorescent signal intensity of the labeled test DNA to that of the reference DNA is used to detect CNVs, PAVs and InDels (Schrider and Hahn 2010).

2.4.2 Real-Time Detection Based Markers: KASPar Genotyping System

The KBioscience competitive allele-specific PCR (KASPar) genotyping system is a modification of TAM technology that does not require a hybridization step; instead it involves real-time detection of the product, giving it an advantage in terms of steps and time involved in the detection process. This makes KASPar a simple, cost-effective and flexible way for determining both SNP and InDel genotypes, since the assays can be adjusted according to the needs to 48, 96, 384 and 1536-well plate formats. The technology utilizes a unique form of allele specific PCR that is different from the conventional amplification refractory mutation system (*ARMS)*, which makes use of four primers, two allele specific and two locus specific primers. The KASPar chemistry involves two competitive allele specific tailed forward primers and one common reverse primer. The KASPar® assay system relies on the discrimination power of a novel form of competitive allele specific PCR to determine the alleles at a specific locus. To improve the performance of the detection platform, KBioscience perfected this technique by incorporating (1) a 5′–3′ exonuclease cleaved *Taq* DNA polymerase (the engineered *Taq* increases its discrimination power) and (2) a homogeneous Fluorescence Resonance Energy Transfer (FRET) detection system. The two allele-specific primers of a SNP are designed such that they incorporate a unique 18 bp tail to the respective allele specific products, which in later cycles allow incorporation of allele specific fluorescent labels to the PCR products (with the help of corresponding labelled primers). The dual emission modules of the detection system offer great advantage to read the internal standard (ROX) and the allele specific dyes (FAM or VIC) together, making it a qualitative SNP genotyping assay technique.

2.4.3 High-Resolution Melting (HRM) Curve Analysis

The HRM analysis is a recently developed promising technology used for the detection of variations in DNA. The thermal stability of a DNA fragment is determined by its base sequence. When the DNA fragment contains an altered sequence, the duplex stability is changed, leading to different melting behavior, which can be identified with HRM analysis. During HRM analysis, melting curves are produced

using intercalating DNA dyes (*SYBR Green*) that fluoresce in the presence of double-stranded DNA and a specialized instrument designed to monitor fluorescence during heating. When the temperature increases, the DNA-intercalating dye is released from the DNA and the fluorescence decreases. This process produces a characteristic melting profile that can be monitored with precision. Changes in the sequence within the DNA fragment, as in SNPs or InDels, alter the melting profile. The HRM assay has successfully been used to detect point mutations in crop plants. The technique has also served as an alternative to agarose gel electrophoresis to score for the presence/absence of amplicons to detect insertion/deletion polymorphisms (IDPs) with or without prior knowledge of the presence of a polymorphism. For instance in bread wheat insertion site-based polymorphism (ISBP) were amplified by PCR in the presence of SYBR Green I and subsequently used for melting curve analysis on an ABI_PRISM 7900HT. The HRM analysis of 711 ISBP markers allowed assignment of these markers to deletion bins by scoring for the presence⁄absence of the amplicon in chromosome 3B aneuploid lines and also allowed evaluation of polymorphism between the parents of five mapping populations (Paux et al. 2010).

2.5 High Density Arrays-Based Resequencing for SNP Discovery

High-density whole genome oligonucleotide tiling arrays have been used for resequencing whole genomes of model systems like Arabidopsis and rice, leading to discovery of millions of SNPs. *Arabidopsis* was the first plant to enjoy the advent of these whole genome approaches, where high-density microarrays were used to describe sequence diversity in the entire Arabidopsis genome (Clark et al. 2007). In this study, hybridizing genomic DNA from 20 divergent Arabidopsis strains to tiling arrays with almost one billion different oligonucleotides increased the number of known SNPs to 1,074,055, and provided the foundation for the first haplotype map among organisms outside mammals. The extensive number of polymorphisms identified in this study eliminated the need for polymorphism discovery, and made mapping and tracking of genes controlling complex traits feasible by the development of very high-density genetic linkage maps. A similar study was also conducted in rice (*O. sativa*), where genomes of 50 accessions of cultivated rice allowed detection of 6.5 million SNPs (Xu et al. 2012); however both of these studies were quite successful but marked the end of an era of microarray-based resequencing, when NGS technologies became available.

2.6 Array-Based Genotyping Platforms

2.6.1 Array-Based High-throughput SNP Genotyping Platforms

Microarray based genotyping platforms are increasingly becoming popular for genome-wide genotyping since they offer highly multiplexed assays at a relatively

low cost per data point. These high-throughput platforms offer large-scale genotyping for dozens to thousands of SNPs in one or more genomic DNA samples (see Syvanen 2005; Fan et al. 2006a, b; Gupta et al. 2008; McCouch et al. 2010). Both low to high resolution platforms are available to meet the different needs of research communities in different crop plants. Some of the important SNP genotyping platforms reported for low to high through-put SNP genotyping include the following: (1) Illuminas GoldenGate platform (Fan et al. 2003), (2) Illumina's BeadChip™ based Infinium platform (Steermers and Gunderson 2007), (3) GenomeLab™ SNPstream Genotyping System, (4) MegAllele genotyping system based on Affymetrix ParAllele's Molecular Inversion Probe (MIP) Technology, (5) GeneChip™ technology and ASO tiling arrays based on Affymetrix GeneChip platform, (6) TaqMan by Life Technologies (Livak et al. 1995), (7) OpenArray platform (TaqMan OpenArray Genotyping System, Product Bulletin), and (8) Competitive Allele Specific PCR (KASPar) by KBiosciences (http://www.kbioscience.co.uk/index.html).

Based on the methodology involved, majority of the above SNP genotyping assays have been classified into the following groups: (1) allele-specific hybridization, (2) primer extension, (3) oligonucleotide ligation (4) invasive cleavage, (5) allele-specific PCR amplification, (6) DNA conformation methods, and (7) enzymatic cleavage method to include the invader assay (for details, see Xu 2010). The details of these platforms have been described elsewhere in several earlier reviews (see Fan et al. 2003; Syvanen 2005; Gundersson et al. 2006; Steemers and Gundersson 2007; Gupta et al. 2008; Appleby et al. 2009; Ragoussis 2009). However, among all these genotyping platforms, the most popular high-throughput genotyping assays among researchers working on cereal crops included Illumina's GoldenGate and Infinium assays, and KBiosciences KASPar assay. Therefore, in this section, we will briefly discuss the development and use of Illumina's GoldenGate and Infinium assays and KBioSciences' KASPar assay for world's major cereal crops.

Illumina's GoldenGate Assay

Illuminas GoldenGate assay, which makes use of customized oligonucleotide pool assays (OPAs), is one of the most widely used genotyping platforms for cereals at present. It provides low to mid-plex genotyping for genome profiling and validation studies. This genotyping platform is extremely flexible and allows researchers to select for number of SNPs (for each of the samples to be genotyped) and the throughput level that best suit their experimental requirements. The system can be utilized for any crop species using either Bead Array, or VeraCode technology (http://www.illumina.com/applications/detail/snp_genotyping and_ cnv_analysis/custom_low_to_mid_plex_genotyping.ilmn). Based on the level of multiplexing and throughput, GoldenGate assays can be classified into: (1) GoldenGate Bead Array (2) GoldenGate Veracode and (3) GoldenGate Indexing. GoldenGate assay, which is common to all the three technologies involves use of two allele specific

oligonucleotides (ASOs) and a locus specific oligonucleotide (LSO) for each SNP. All the three oligonucleotides are supplemented with non-template specific universal primer sites; the LSO also carries an anti-tag sequence corresponding with a particular bead type on the BeadArray. The specific primers (ASOs and LSO) bordering each SNP allow allele specific primer extension and universal primers allow labeling and detection of the product (for details, see Fan et al. 2006a, b). A comparison of three assays has been presented in Table 2.1.

1. **GoldenGate BeadArray Assays (Bead Array, iScan).** This assay allows simultaneous genotyping of 96–3,072 (96-, 192-, 384, 768-, 1536- and 3,072) SNP loci in a fairly large collection of samples (up to 384 samples) in parallel. This is one of the most popular SNP genotyping platforms providing cost effective assays (per genotype cost $0.03). These assays are now becoming available in all major cereals including wheat (Akhunov et al. 2009; Chao et al. 2010), rice (McCouch et al. 2010), barley (Rostocks et al. 2006; Close et al. 2009; Druka et al. 2011) and maize (Mammadov et al. 2012). Among cereals, barley is the first crop where GoldenGate assay for 1,536 SNPs (selected on the basis of EST mining) was developed and used for the study of population structure and the level of LD exhibited in elite Northwest European barley (Rostoks et al. 2006). Later, it was also used for molecular characterization, genetic diversity analysis, preparation of integrated maps, consensus maps, bulk segregant analysis (BSA), identification of QTL, linkage disequilibrium (LD) studies, association mapping, joint linkage–linkage disequilibrium (LD) mapping approaches, etc. A summary of SNP studies conducted in different cereals using Illumina's GoldenGate (GG) assays is presented in Table 2.2.
2. **GoldenGate VeraCode (VeraCode Bead Plate BeadXpress)**: BeadXpress involving Veracode technology is considered as most flexible and low- to mid-plex GoldenGate SNP genotyping assay. In this genotyping platform, custom SNP assays are ordered from Illumina in 48-, 96-, 192-, and 384-plex (GoldenGate Kits), 1-144-plex (Universal Capture Bead Sets) and 1–48-plex (Carboxyl Bead Sets) formats, and the DNA samples are processed in a 96-well format (Table 2.1). BeadXpress involving 384-SNP OPAs is very useful for cereal breeding/genetics

Table 2.1 Comparison of various GoldenGate SNP genotyping assays

Features	GoldenGate BeadArray	GoldenGate VeraCode	GoldenGate Indexing
Multiplexing	96–1536-plex	48–384-plex	96–384-plex
DNA needed	~250 ng	~250 ng	~250 ng
System used	iScan system	BeadXpress system	iScan system
Array type	Bead Array	VeraCode	Bead Array
Through-put	~288 samples/day	~288 samples/day	>2,000 samples/day
Reaction	ASPE	ASPE	ASPE
Suitability for MAS	Less	More	Less
Suitability for mapping	More	Less	More

ASPE = allele specific primer extension

community since this is reliable and requires little technical adjustments after their designing and optimization. As against, BeadArray technology, VeraCode makes use of a VeraCode Bead Plate, which carries addresses for SNP alleles to be detected by the anti-tags carried by LSO. The use of VeraCode Bead Plate in place of BeadArray reduces the cost per sample, when lower-plex genotyping is needed. This genotyping platform is suitable to assay hundreds or thousands of genotypes in a short span of time. Several 384-SNP BeadXpress assays have already been developed and used in cereals (see Table 2.2).

3. **GoldenGate Indexing (iScan):** Illumina's GoldenGate Indexing is a recent high-throughput, low cost technology involving low to mid-plex genotyping of 96–384 SNPs simultaneously. This will allow researchers to pool multiple samples, thus increasing the number of samples in a single run (http://www.illumina.com/documents/products/datasheets/datasheet_goldengate_indexing.pdf). One can screen up to 16 times more samples per reaction than one can do with the standard GoldenGate (GG-BeadArray) assay (see above) and therefore, increases throughput from 288 samples/day (in GG-BeadArray assay) to >2,000 samples/day, thereby decreasing cost (Table 2.1). This system has not been used in plants so far, but it is anticipated that this emerging technology will soon find its application in plants also.

Illumina's Infinium Assays (SNP-CGH)

Illumina's BeadChip^{TM} based Infinium assay, involving array-comparative genomic hybridization (aCGH), is a high-density SNP genotyping technology for whole-genome genotyping that allows genotyping of fixed sets of hundreds of thousands of SNPs simultaneously. It allows simultaneous measurement of both signal intensity variations and changes in allelic composition (Varshney 2010). In this assay, BeadChips with 12-, 24-, 48- or 96 sections can be used simultaneously with each section of a BeadChip containing 1.1 million beads carrying decoded oligonucleotides (for further details consult, Syvanen 2005; Gundersson et al. 2006; Steemers and Gundersson 2007; Gupta et al. 2008).

With the advent of next-generation sequencing technologies, high density SNPs have been discovered in all important crop plants including cereals; this facilitated the development of Infinium assays in these crops. For instance, in soybean, 44,299 informative SNPs were used to develop '*Illumina Infinium iSelect SoySNP50 chip*' that was later used to dissect and resolve the issue of origin of genomic heterogeneity in soybean cultivar, William 82. The CGH analysis for >2,03,000 loci revealed the consequences of this heterogeneity in terms of structural and gene content variants among individuals of the cultivar, Williams 82 (Haun et al. 2011). Similarly, efforts are being made to design a 50 K SNP Illumina Infinium assay and use it to analyze each of the 18,603 cultivated and 1,116 wild soybean accessions from the USDA soybean germplasm collection and 1,000 RILs from each of the two mapping populations of soybean (Williams

Table 2.2 A summary of SNP genotyping studies conducted in some important cereal crops

Crop and Platform	Silent features of the study	Reference
1. Wheat		
GG (1536 SNPs)	878 loci assigned to 7 linkage groups at a maximum resolution of 0.087 cM. Map comparisons with rice and sorghum revealed 50 inversions and translocations	Luo et al. (2009)
GG (96 SNPs)	53 tetraploid and 38 hexaploid wheat lines were genotyped at 96 SNP loci to demonstrate utility of GoldenGate assay for polyploids	Akhunov et al. (2009)
GG(1536 SNPs)	Studied LD and population structure in a panel of 478 spring and winter wheat cultivars from USA and Mexico	Chao et al. (2010)
GG	53 ISBP-derived SNPs markers were used to genotype 96 hexaploid wheat varieties, 96 individuals from a Chinese Spring x Renan F_2 population, and aneuploid lines	Paux et al. (2010)
Illumina beadexpress (768 SNPs)	275 new SNPs reported; 157 SNPs mapped in one of two mapping populations (Meridiano × Claudio and Colosseo × Lloyd) and integrated into a common genetic linkage map	Trebbi et al. (2011)
2. Rice		
(1,536 SNPs) GG	Captured variation within and between *O. sativa* subpopulations	Zhao et al. (2010)
	Captured variation within temperate *japonica* cultivars	Yamamoto et al. (2010); Nagasaki et al. (2010)
Illumina BeadXpress (384-plex)	Evaluated variation between various rice species/subspecies: *indica* and *japonica, indica* and *aus,* US tropical *japonica, indica* and *O. rufipogon, japonica* and *O. rufipogon*	Thomson et al. (2012)
Affy (44,100 SNPs)	Evaluate diversity within and between sub-populations of *O. sativa*	Tung et al. (2010)
Illumina BeadXpress (384-plex)	372 SNPs unraveled the *indica–japonica* subspecific differentiation and geographic differentiations within *Indica* and *Japonica* in 300 rice inbred lines	Chen et al. (2011)
Affy (1 M SNPs) Resequencing 6.5 M SNPs	Evaluate diversity within and between *O. sativa*, *O. rufipogon*/*O. nivara*, *O. glaberrima* and *O. barthii*	Xu et al. (2012)
3. Barley		
GG BeadArray (SNPs in 1,524 barley unigenes)	Studied genome-wide molecular diversity, population substructure, and LD in elite Northwest European barley cultivars	Rostocks et al. (2006)
GG (two 1,536-SNPs assays)	Use of high-throughput SNP genotyping platform for the development of a consensus map containing 2,943 SNP loci covering a genetic distance of 1,099 cM	Close et al. (2009)

(continued)

Table 2.2 (continued)

Crop and Platform	Silent features of the study	Reference
GG (1,536 SNPs)	Reported GWA mapping of 15 morphological traits across ~500 cultivars genotyped with 1,536 SNPs, and fine-mapped anthocyanin pigmentation to a 140-kb interval containing 3 genes	Cockram et al. (2010)
GG (1,536 SNPs)	Defined the genetic location of 426 morphological mutants using 3,072 SNPs	Druka et al. (2011)
GG	Diversity analysis using 1,301 SNPs on a set of 37 barley accessions revealed high polymorphism rate between 'Haruna Nijo' and 'Akashinriki'. A DH population was derived from them, and genotyped using 1,448 SNPs, of which 734 showed polymorphism and integrated into the linkage map. 98 RCSLs developed from the same cross were also genotyped using SNPs	Sato et al. (2011)
4. Maize		
GG (1536 SNPs)	Genotyping global maize collection of 632 inbred lines and estimation of genetic diversity, population structure, and LD	Yan et al. (2009)
GG (1,536 SNPs)	Molecular characterization of global maize breeding germplasm involving study of genetic diversity	Lu et al. (2009)
GG (1536 SNP)	Genotyping of Nested Association Mapping (NAM) population (4,699 RILs) with 1,106 SNPs to develop integrated linkage map	McMullen et al. (2009)
Illumina_BeadArray™ (768 SNPs)	Mapping of 591 markers on IBM2 genetic map covering ~88 % genome	Jones et al. (2009)
Sequenom-based typing of 1,359 SNPs	Mapping of phenotypic mutants using a combination of quantitative SNP-typing and bulked segregant analysis	Liu et al. (2010)
GG (1,536 SNPs)	Construction of a high-density linkage map containing 662 markers (1,673.7 cM)	Yan et al. (2010)
GG (two 1536 SNP assays)	Conducted linkage and LD based mapping for detection of drought tolerance QTLs	Lu et al. (2010)
GG assay (1,000 SNPs; ~700 loci)	Mapped 604 SNPs distributed on ten maize chromosomes	Mammadov et al. (2010)
GG (1,536 SNPs)	Genotyping of 2–3 comparable generations of twenty maize accessions conserved in five genebanks	Wen et al. (2011)
GG	Used 695 highly polymorphic SNPs for genotyping with Illumina's GoldenGate/Infinium, TaqMan and KASPar	Mammadov et al. (2012)

(continued)

Table 2.2 (continued)

Crop and Platform	Silent features of the study	Reference
GG (2 OPAs of 1,536 SNPs each)	An integrated map spanning 1,346 cM was constructed using 1,443 molecular markers, including 1,155 SNPs. A 100-fold difference in recombination frequency was observed between different chromosomal regions	Farkhari et al. (2011)
GG (1,536 SNPs)	1,006 polymorphic SNPs grouped 80 lines in 6 subgroups. Pairwise LD and association mapping with phenotypic traits investigated under water-stressed and well-watered conditions showed rapid LD decline within 100–500 kb	Hao et al. (2011)
Illumina MaizeSNP50 BeadChip (56,000 SNP)	Joint-linkage mapping and GWAS revealed that kernel composition traits are controlled by 21–26 QTLs. Numerous GWAS associations were detected, including several oil and starch associations in *acyl-CoA:diacylglycerol acyltransferase 1-2*, a gene that regulates oil composition and quantity	Cook et al. (2012)
5. Sorghum		
GG (384 SNPs)	Genotyped 125 sorghum genotypes to perform whole genome association mapping for height and brix (stem sugar)	http://maizeandgenetics.tamu.edu/presentations

GG = GoldenGate

82 × PI468916 and Essex × Williams 82) with the Illumina BeadStation 500 to obtain ultra-high resolution genetic maps of soybean (http://www.soybeanchecko ffresearch. org/DetailsbyPaperid.php?id_Paper = 991). These Illumina Infinium genotyping assays have now been used even in non-model plant species. For instance, using Infinium assays, 622 loblolly pine trees sampled from 167 locations were genotyped using SNPs across 3,059 functional genes. This allowed a study of population structure and environmental associations to aridity in loblolly pine (Eckert et al. 2010).

In cereals, some of the applications of Infinium assay included the following: (1) A 50 K SNP Infinium chip in maize covers approximately two-thirds of all maize genes and also includes additional SNPs spread over most of the remaining maize genome resulting in an average marker density of approximately one marker every 40 kb (Ganal et al. 2011). (2) A set of 618 gene-based SNPs were successfully converted into different genotyping assays including infinium assay in maize (Mammadov et al. 2012). The study also demonstrated the conversion of SNPs from GoldenGate assays into Infinium assays with a success rate of ~89 %. The commercial availability of these high-density SNP platforms will undoubtedly facilitate the application of SNP markers in molecular breeding (Mammadov et al. 2012). (iii) A pilot 9 K SNP Infinium assay (http://129.130.90.21/IWSWG/sites/default/files/9K_assay_available_updated.docx) was developed recently in a USA/Australia collaborative project and used to genotype tetraploid and hexaploid wheat lines and cultivars. The assay includes SNPs discovered from the transcriptomes generated from a set of 27 US/Australian lines. Preliminary results showed that more than 90 % of SNPs produce high-quality genotype calls.

Cost Effectiveness of GoldenGate (GG) Assays and Their Suitability for Molecular Breeding

In maize, genotyping using GoldenGate (GG) assay was found 100-fold faster than gel-based methods. When the cost of genotyping for preparation of two linkage maps was compared, it was found that there was a cost saving of ~75 % in GG-based SNP genotyping relative to gel based methods used for SSRs. In addition, SNP genotyping with GG assays allows development of molecular maps with 2–3 times higher density in a fraction of time required for the development of SSR-based maps (Yan et al. 2010). However, while comparing with DArT markers, GG assays were found to be 3 times more expensive (Mantovani et al. 2008).

However, GG assays have not been favored for molecular breeding, although both GG and Infinium assays have been applied for rapid construction of genetic linkage maps, gene/QTL mapping and GWAS. This may be due to the requirement for multiplexing to bring down the assay cost per data point, which will be a bottleneck for their use in molecular breeding involving use of only a few SNPs closely linked to the gene/QTL of interest. If an associated SNP belongs to the set of SNPs (OPA) included in a GG or Infinium assay, the breeder prefers to convert the desired SNP(s) into another user-friendly high-throughput assay (e.g. KASPar or TaqMan) that does

not require multiplexing and are still cost effective. However, several issues may crop up while converting one SNP assay into another and may jeopardize the application of a particular marker in MAS (Mammadov et al. 2012).

Fluidigm SNP Genotyping

In addition to several genotyping platforms by Illumina discussed above, BioScience LifeSciences™ "Fluidigm SNP genotyping system" is also one of the important SNP genotyping platforms gaining popularity in plant molecular breeding community. It uses the innovative integrated fluidic circuit (IFC) and is used for studies requiring ultra-low cost and high–sample throughput for low- to mid-multiplex SNP genotyping™ (http://www.lifesciences.sourcebioscience.com/genomic-services/genotyping/snp-genotyping-using-the-fluidigm-ep1-system.aspx). The system involves the use of 48.48, 96.96 and more recently 192.24 type of arrays. The 192.24 array is called "Dynamic Array™ IFC" because it is designed to genotype 192 samples against 24 SNP assays in a single run, thus greatly increasing the sample throughput with only few selected important SNPs for molecular breeding. The platform has already been used in rice (Ilic et al. 2011), cocoa (Ilic et al. 2012), grain amaranths (Maughan et al. 2011) and *Bromus tectorum* (Merrill et al. 2011).

Competitive Allele Specific PCR (KASPar) Assays

GoldenGate (GG) and/or Infinium assays have been widely used for rapid genotyping of a large number of SNP markers in all major crop species including cereals (see Table 2.2). However, for genotyping a population for few SNPs, where GG assays are not cost-effective (Chen et al. 2010), KASPar (**K**BioScience **A**llele-**S**pecific **P**olymorphism, KBioscience, UK) system provides a promising alternative. The method involves competitive allele-specific PCR, followed by SNP detection via Fluorescence Resonance Energy Transfer (FRET; for review see McCouch et al. 2010).

KASPar genotyping may be of particular interest to breeders and researchers who are interested in analyzing a small number of targeted SNPs in a large number of samples. Therefore, KASPar genotyping assays may be used for a variety of purposes including the following: (1) genetic diversity studies; (2) genetic mapping and saturation of already prepared maps; (3) fine-mapping of QTLs; (4) detection of functional SNPs within a subset of germplasm; (5) marker-assisted breeding, and (6) retaining target regions in NIL development (see McCouch et al. 2010). This genotyping system has already been used for a large number of species including cereal crops like rice, maize and wheat. In wheat, the technique has been used for rapid generation of a linkage map containing several hundred SNPs (Allen et al. 2011). Similarly, in maize, a set of 695 highly polymorphic gene-based SNPs from a total of 13,882 GG-validated SNPs were selected and converted into KASPar genotyping assay with a success rate of 98 % (Mammadov et al. 2012).

2.6.2 Diversity Array Technology (DArT) Markers in Cereal Crops

Diversity array technology (DArT) is a high throughput microarray hybridization-based technique that allows genotyping for several hundred polymorphic loci spread over the whole genome without any prior sequence information (Jaccoud et al. 2001). The technique is reproducible and cost-effective (Wenzl et al. 2004), and therefore, has been used in a number of crop species (including cereals) as evident from the trends in the number of papers published during the last decade (Fig. 2.3). This trend of papers published using DArT platforms is continuously increasing each year along with papers being published using GoldenGate assays. It is estimated that for the discovery of polymorphic markers ~5,000–8,000 genomic loci are typed in parallel in a single-reaction assay using a small quantity (50–100 ng) of highly purified genomic DNA. Polymorphic markers once discovered on a discovery array (prepared using metagenome of a crop species) are combined into a single array called "genotyping array" to be used for routine genotyping work (Huttner et al. 2005). The detailed method used for the development of DArT markers was described in our earlier review (Gupta et al. 2008). However, with the success of array based DArT markers over the past ~12 years, it was realized that the number of polymorphic markers can be increased by involving the use of next-generation sequencing (NGS), so that the cost of producing sufficient number of tag counts dropped to the commercially viable levels. In this platform, genome complexity reduction for genotyping has been combined to next

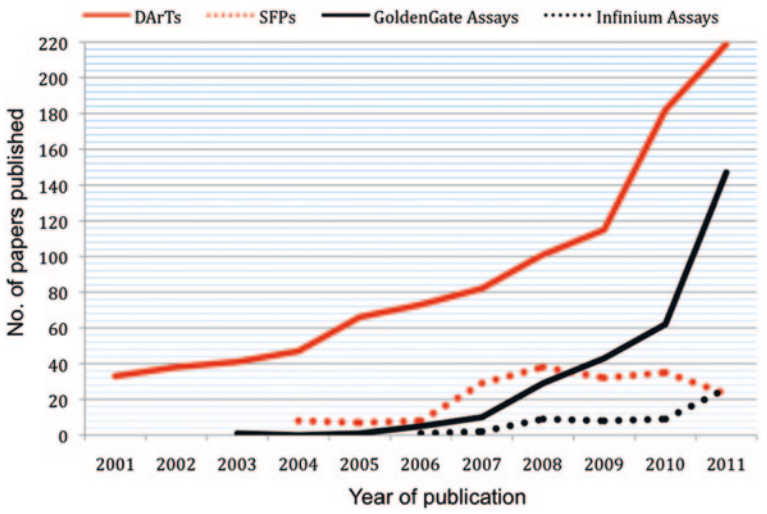

Fig. 2.3 Trends in publications related to GoldenGate assays, Infinium assays, DArT and SFPs in crop plants. The publications on DArT, GoldenGate and Infinium assays are increasing every year at fast rate, while as only a limited number of publications have become available on SFPs. (*Source* Google Scholar, March 2, 2011)

generation sequencing (NGS) technologies. Such a strategy has been used for rapid SNP discovery in different organisms. This was also proposed as a method for genotyping with RAD (Restriction-site Associated DNA) sequencing and by another similar method generally termed GbS (Genotyping-by-Sequencing). Therefore, a new platform was developed for >30 organisms now and it was shown that there is 3-fold (or more) increase in marker number on the new platform compared to arrays (Andrzej Kilian, personal communications; Sansaloni et al. 2011).

Nature of DArT Markers

DArT markers are biallelic and dominant in nature (presence *vs* absence); only rarely, these can be co-dominant (2 doses *vs* 1 dose *vs* absent). The software DArTsoft is used for the analysis of hybridization intensities. The efficiency of identification of polymorphic DArT markers depends on the level of genetic diversity within a crop species and the nature of diversity in metagenome constituting the discovery array. DArT markers usually detect polymorphisms due to single base-pair changes (SNPs) within the restriction sites recognized by endonucleases, or due to insertion/deletion (InDels)/rearrangements (Jaccoud et al. 2001). The type of polymorphism detected by DArT markers depends on the complexity reduction method applied to DNA samples of various genotypes/populations. For instance, a methylation sensitive restriction enzyme like *Pst*I will identify markers having both sequence variation (SNPs and InDels) and DNA methylation polymorphism (Kilian et al. 2005).

DArT Markers in Breeding for Cereal Crops

DArT assays are available for a fairly large number of plant species, including some orphan crops, for which no molecular information is available (Huttner et al. 2006; for details see Table 2.3). It is interesting to note that within a short span of time, DArT markers have become popular among researchers and have now become available for >70 species involving plants and animals (http://www.diversityarrays.com/genotypingserv.html). This technology provides a good alternative to currently available marker techniques including RFLP, AFLP, SSR and SNP in terms of cost, speed/amount of data generation. DArT markers are particularly useful for organisms, where no SNP arrays are publicly available (Mace et al. 2009). In addition, DArT markers are sequence-independent and non-gel based in nature. Also DArT assays involve automation and allow discovery of hundreds of high quality markers in a single assay making DArT markers the marker of choice for resource poor and underutilized orphan crops. The cost per data point (a few cents per marker assay) is reduced by at least an order of magnitude compared to gel-based technologies (Mace et al. 2009). The initial proof of concept of DArT technology was provided by using relatively simple genome of rice (Jaccoud et al. 2001). Later barley and other crops with more complex genomes were also used to demonstrate the

Table 2.3 A summary of DArT genotyping studies conducted in some important cereal crops

Crop involved & salient features of the study	Reference
1. Wheat	
First successful report of DarT in bread wheat (*T. aestivum*). Mapped DArT, AFLP, SSR and STM markers using 90 DH lines	Akbari et al. (2006)
Developed complexity reduction methods to generate large number of diverse clones for genotyping arrays. Constructed a framwork linkage map using 93 DH lines and DArT, AFLP and SSR markers	Semagn et al. (2006)
Used >1,000 DArT markers and developed a linkage map with 90 SSR and 543 DArT markers using 176 RILs	Wenzl et al. (2007b)
242 DArT markers used to study association with resistance against stem rust, leaf rust, yellow rust, powdery mildew, yield and yield contributing traits	Crossa et al. (2007)
Studied genetic diversity of UK, US and Australian wheat varieties	White et al. (2008)
Studied genetic diversity to elucidate the genetic relationships among the selected spring and winter wheat lines/cultivars using DArT and SSR markers	Badea et al. (2008)
A linkage map was prepared using SSRs and 209 DArT loci; QTL were identified for powdery mildew resistance and their correspondence with adult plant rust resistance loci *Lr34/Yr18* and *Lr46/Yr29* was established	Lillemo et al. (2008)
An integrated DArT-SSR map incuding 162 SSRs and 392 DArT loci (2,022 cM) developed; DArT markers were also used to profile a panel of durum accessions; the genetic relationships based on DArT and SSR markers were compared	Mantovani et al. (2008)
High density genetic map including 197 SSR and 493 DArT loci developed	Peleg et al. (2008)
Genetic map with SSR and DArT markers was used for identification of QTLs for grain fructan concentration	Huynh et al. (2008)
DArT markers were used to evaluate gene bank accessions of spelt for genetic diversity and for resistance to aluminium toxicity and for low *PPO* activity	Raman et al. (2009)
DArT markers were developed for *T. monococcum* to assess genetic diversity, compare relationships with hexaploid genomes, and construct a genetic linkage map integrating 274 DArT and 82 SSR markers	Jing et al. (2009)
Comparison of genetic and cytogenetic maps of *T. aestivum* using SSR and DArT markers. Extended number of DArT markers on the wheat array that can be used for mapping by determining their chromosomal location in reference to SSRs	Francki et al. (2009)
Developed high-density genetic map of chromosome 3B containing 939 markers (779 DArT and 160 other markers)	Wenzl et al. (2010)
Developed a genetic map with 429 DArT and few SSR markers. DArT markers *wPt-3049* (2.9 cM) and *wPt-0289* (4.6 cM) respectively showed associations with tan spot resistance genes *Tsr1* and *Tsr6*	Singh et al. (2010)
Used DArT and SSR markers for an updated map with higher marker density; QTL analysis for stripe, leaf and stem rust resistance	Prins et al. (2011)
Use of DArT in the study of population structure, LD and association with 20 agronomic traits (including grain yield, grain quality and disease resistance)	Neumann et al. (2011)

(continued)

Table 2.3 (continued)

Crop involved & salient features of the study	Reference
Studied genetic diversity and population structure using 1,637 DArT markers among 111 genotypes from northern China	Zhang et al. (2011)
567 spring wheat landraces were genotyped with 832 DArT markers to identify QTLs for resistance against *P. tritici-repentis* races 1 and 5	Gurung et al. (2011)
843 polymorphic DArT markers were used to genotype wheat lines, and genotypic data was used to assess genetic relationships among the accessions	Yu et al. (2010)
A whole genome map was developed using 676 polymorphic DArT markers. The map was used to find genomic regions associated with citrate efflux using single marker regression and interval mapping. A major QTL, *Qce$_c$-4BL* was identified on the long arm of chromosome 4B by both of the methods	Ryan et al. (2009)
195 Western European elite wheat varieties genotyped using 159 SSRs and 634 DArT markers to evaluate the effect of population structure in association tests for three major genes involved in plant height, heading date and awnedness	Le Couviour et al. (2011)
2. *Rice*	
Initial proof of concept and validation of DArT using AFLP like complexity reduction methods involving nine rice cultivars	Jaccoud et al. (2001)
Evaluated genetic diversity in a general purpose rice gene pool and validated DArT for rice genotyping	Xie et al. (2006)
3. *Barley*	
Constructed a genetic map with ~385 unique DArT markers (1,137 cM), using barley cultivars Steptoe and Morex	Wenzl et al. (2004)
Constructed high-density consensus linkage map with ~ 3,000 loci including 2,085 DArT loci	Wenzl et al. (2006)
Used DArT/SSR markers for QTL mapping of *Fusarium* head blight (FHB) resistence	Rheault et al. (2007)
Mapped >600 DArT markers covering ~2,000 cM; identified 15 clustered loci for multiple resistance	Aslop et al. (2007)
Mapped *Rsp1* on 3H and *Rsp2*, and *Rsp3* on 1H	Lee and Neate (2007)
Constructed high-density genetic map with 558 SSR and 442 DArT markers	Hearnden et al. (2007)
Tested suitability of DArT for bulk segregant analysis in barley; validated an aluminum tolerance locus on chromosome 4H	Wenzl et al. (2007a)
Detected associations of DArT markers (on 5H) with stem rust resistance in *Aegilops sharonensis* and wild barley	Steffenson et al. (2007)
Mapping of quantitative trait loci (QTL) associated with net blotch resistance in a DH population using DArT markers	Grewal et al. (2008)
Constructed a composite map using SSR, RFLP and DArT markers and used it for identification of QTLs for water logging tolerance	Li et al. (2008a)
QTL analysis for seedling and adult-plant resistance to spot and net blotch using an SSR, AFLP and DArT based map	Grewal et al. (2012)
1,130 DArT markers were used on a diverse barley collection to scan their genomes for associations with yield components	Comadran et al. (2009)

(continued)

Table 2.3 (continued)

Crop involved & salient features of the study	Reference
1,000 polymorphic DArT markers were used to study genetic diversity, population structure, and extent of LD in 170 Canadian barley genotypes	Zhang et al. (2009)
Preparation of integrated map with SSR, AFLP and DArT markers and identification of QTLs for β-glucan content	Li et al. (2008)
Association mapping for malting quality traits using 91 elite two-rowed malting barleys; 27 DArT markers were found to be associated with malting quality	Beattie et al. (2010)
Association mapping for spot blotch resistance using Wild Barley Diversity Collection (WBDC) using 558 DArT and 2,878 SNP markers; 13 QTLs identified	Roy et al. (2010)
DArT markers were used to find out if the accessions with multiple resistance (MR) from the Vavilov nursery were genetically related to accessions with MR from Ethiopia	Bonman et al. (2011)
Genetic maps were developed for four populations and 607 new DArT markers were integrated in a consensus map with 3,542 markers	Alsop et al. (2011)
253 DArT markers were used on a set of 183 varieties, which clearly distinguished between spring and winter types and classified them into five subgroups	Matthies et al. (2012)
4. Sorghum	
Genotyped of two mapping populations of sorghum	Mace et al. (2007)
Mapped 330 non-redundant DArT markers covering whole genome	Bouchet et al. (2007)
DArT markers were used to identify genomic regions associated with yield and adaptation	Jordan et al. (2007)
DArT markers were used for genetic diversity study and for construction of a genetic linkage map	Mace et al. (2008)
A genetic linkage map was constructed using 36 SSR, 117 AFLP and 148 DArT markers. QTLs for ergot resistance and two pollen traits were identified	Parh et al. (2008)
Developed consensus map including 1190 DArT loci (58.6 % of total mapped loci)	Mace et al. (2009)
5. Rye	
Determined genetic relationships between rye varieties and inbred lines using 1,022 DArT markers. Chromosomal location of 1,872 DArT markers was also determined, providing an average density of one unique marker every 2.68 cM	Bolibok-Brągoszewska et al. (2009)
564 RILs from 5 mapping populations were genotyped using DArT markers and subjected to linkage analysis. A consensus map was constructed using a total of 9,703 segregating markers. The average chromosome map length ranged from 199.9 cM (2R) to 251.4 cM (4R) and the average map density was 1.1 cM	Milczarski et al. (2011)

(continued)

Table 2.3 (continued)

Crop involved & salient features of the study	Reference
6. Triticale	
21 linkage groups assigned to the A, B, and R genomes using 155 SSR, 1,385 DArT, and 28 AFLP markers	Tyrka et al. (2011)
Evaluated DArT markers for transferability from rye and wheat to triticale; DArT technology used for diversity analyses on a set of 144 triticale accessions	Badea et al. (2011)
Consensus map, constructed out of six segregating populations, incorporated 2,555 DArT markers (2,602 loci) spanning 2,309.9 cM with an average number of 123.9 loci per chromosome and an average marker density of 1 locus/1.2 cM	Alheit et al. (2011)
7. Oat	
1,010 new DArT markers were used to saturate earlier genetic map. A set of 1,295 markers was used to analyze genetic diversity	Tinker et al. (2009)
Seedling crown rust resistance gene "*Pc91*" was mapped to a linkage group with DArT markers. Five robust SCARs were developed from three non-redundant DArTs that co-segregated with *Pc91*	McCartney et al. (2011)
1,205 lines and 402 DArT markers were used for population structure and LD studies in oat germplasm and determine their implications for GWAS	Newell et al. (2011)
A high density map was prepared using 974 DArT, 26 SSR, 13 SNP, and 4 phenotypic markers	Oliver et al. (2011)

usefulness of DArT technology (Wenzl et al. 2004, 2006, 2007; Hearnden et al. 2007; also see Table 2.3). For instance, DArT markers are now available for diploid wheat (Jing et al. 2009), tetraploid wheat (Peleg et al. 2008), hexaploid wheat (Akbari et al. 2006; Semagn et al. 2006; White et al. 2008), sorghum (Mace et al. 2008, 2009), rye (Bolibok-Brągoszewska et al. 2009) and more than 30 other plant species (Jing et al. 2009). In case of wheat alone, more than 50,000 samples (>95 % as service at ~1 cent per marker assay) involving >350 mapping populations were processed, which resulted in preparation of >100 maps having ~7,000 markers assigned to specific chromosomes (A. Kilian, personal communication). Choromosome specific (3B) and individual chromosome arm (1BS)-specific DArT markers have also been developed using flow sorted chromosome/chromosome arm. A total of 553 of the 711 polymorphic 3B-derived markers (78 %) were mapped on chromosome 3B, and 59 of the 68 polymorphic 1BS-derived markers (87 %) were mapped to 1BS, thus confirming the efficiency of the chromosome-sorting approach in DArT technology. A consensus map of chromosome 3B using 19 mapping populations, including some that were genotyped with the 3B-enriched array was also prepared and the map is probably the densest genetic map of 3B available to date; the map contains 939 markers including 779 DArT markers and 160 other markers (Wenzl et al. 2010). Also, DArT markers are now available on large scale for all major cereal crops and were extensively utilized for the study of genetic diversity, preparation of integrated framework linkage maps, QTL interval mapping, association mapping, etc. (Rheault et al. 2007; Wenzl et al. 2007a). The physical map of wheat genome is also being constructed using a chromosome-by-chromosome approach, where individual laboratories developed maps for individual chromosomes (Feuillet and Eversole 2008). These laboratories can now develop saturated DArT maps for their specific chromosomes in an affordable and targeted manner.

Cost-Effectiveness of DArT Assays and Their Suitability for Molecular Breeding

DArT marker assays have shown to be cheaper than any other marker system available at present. The cost per marker assay in commercial service offered by Triticarte P/L is ~US$ 0.02 (or approximately US$ 50 per genotype for ~2,500 DArT markers; Mantovani et al. 2008), which is >6 times cheaper than the cost of SSR genotyping, and ~3 times cheaper than Illumina GG assay (Yan et al. 2010).

The DArT markers on the wheat array are now being assigned to chromosomal bins by deletion mapping. This deletion mapping of DArT markers will provide a reference to align genetic and cytogenetic maps and estimate the coverage of DArT markers across genome (Francki et al. 2009). However, the associated DArT markers identified through QTL interval mapping or association mapping cannot be directly used in marker-assisted selection for the improvement of a desired trait in a crop species. In order to overcome this limitation, the sequences of DArT markers (usually those linked to traits of interest) can be obtained from Triticarte service and can be converted into user-friendly PCR-based markers. For instance,

five robust SCARs were developed from three non-redundant DArT markers that co-segregated with crown rust resistance gene "*Pc91*" in oats. These SCAR markers were developed for different assay platforms: agarose gel electrophoresis, capillary electrophoresis, and TaqMan single nucleotide polymorphism detection (McCartney et al. 2011).

2.6.3 Single Feature Polymorphisms (SFPs) Genotyping in Cereal Crops

Single feature polymorphisms (SFPs) represent another high-throughput array-based genotyping technology. It involves use of oligonucleotides (features), which represent segments of individual genes. Affymetrix (http://www.affymetrix.com) GeneChips or Nimblegen (http://www.nimblegen.com) arrays with small probes (25 bp) capable of detecting sequence polymorphism are the most widely used arrays for SFP genotyping (see Table 2.4). Majority of studies involving discovery and genotyping for SFPs have been conducted in model organisms like yeast, mouse and *Arabidopsis*, whose genomes have been sequenced and characterized (Brem et al. 2002; Borevitz et al. 2003; Kumar et al. 2007). In plants, SFP technology was first applied for *A. thaliana* using Affymetrix expression array (Borevitz et al. 2003). The technique was later used for all important crops including cereals. However, in crops with complex large genomes, a suitable complexity reduction method is used for sample preparation and replicating arrays are used for hybridization. Therefore, SFPs became available for all major cereal crops including barley (Cui et al. 2005; Rostoks et al. 2005), rice (Kumar et al. 2007), maize (Kirst et al. 2006; Gore et al. 2007) and wheat (Coram et al. 2008; Banks et al. 2009; Bernardo et al. 2009). SFPs have actually been used for a variety of studies including the following: (1) genetic mapping (Zhu et al. 2006; Somers et al. 2008; Banks et al. 2009; Bernardo et al. 2009), and (2) QTL interval mapping leading to detection of main effect QTLs and eQTLs (Potokina et al. 2008; Kim et al. 2009). A summary of SFP studies conducted in some of the major cereal crop species is presented in Table 2.4.

2.6.4 Use of Sequenom MassARRAY System for SNP Genotyping in Cereals

The MassARRAY platform has successfully been used to genotype SNPs in mammals (Vogel et al. 2009). More recently, it has also been used in cereals including rice, maize, barley and wheat. For instance, in wheat, SNPs identified in homoeologues of gene for acetohydroxyacid synthase conferring resistance against imidazolinone herbicides were successfully converted, and used on Sequenom MassARRAY system (Dr. Divya Neelam, BASF personal communication). Sequenom-based SNP-typing assays were also developed for 1,359

Table 2.4 A summary of SFP genotyping studies conducted in some important cereal crops

Plant species and resources used	Salient features of study	References
1. Wheat		
Aff ymetrix GeneChip Wheat Genome Array and cRNA	More than 1,500 SFPs were placed on genetic maps using 64 DH individual lines	Banks et al. (2009)
Affymetrix GeneChip (Probe sets from 55,052 transcripts) and cRNA	Microarray analysis of 71 RILs identified 955 SFPs; 877 were mapped with 269 SSR markers	Bernardo et al. (2009)
Affymetrix GeneChip® Genome Array	Identified 100's of SFP markers and integrated them into an existing SSR map	Somers et al. (2008)
55 K Affymetrix Wheat GeneChip (61,127 probe sets) and cRNA	297 SFPs were identified between NILs for stripe rust resistance	Coram et al. (2008)
Affymetrix Genome Array (61,127 probe sets) and cRNA	208 high variance probe sets (HVPs) assigned to wheat chromosome arm 1BS	Bhat et al. (2007)
Wheat oligonucleotide array and cRNA	44 SFPs for 7E (*Thinopyrum* and *Lophopyrum*) identified using alien substitution/addition lines	Buescher et al. (2007)
Affymetrix GeneChip (38,577 probe sets) and cRNA	SFPs identified in 948 genes using two wheat varieties ('Eltan' and 'Oregon feed wheat')	Ling et al. (2006)
2. Rice		
Affymetrix Genome Array (55,515 probe sets) and cRNA	5,376 SFPs in 'LaGrue' (*japonica*), and 25,325 SFPs in 'RT0034' (*indica*), when compared with Cypess (*japonica*)	Kumar et al. (2007)
Affymetrix rice Genome Array and cRNA	1208 SFP probes were detected between two presumed parental genotypes of a RIL population segregating for salt tolerance	Kim et al. (2009)
GeneChip Genome Array (57,381 probe sets) and cRNA	6,655 SFPs between two rice varieties representing 3,131 rice unique genes	Xie et al. (2009)
GeneChip Genome Array (57,381 probe sets) and cRNA	1,632 SFPs and 23 markers were placed into 601 recombinant bins, spanning 1,459 cM. Map was used to identify 26,051 eQTLs assigned to 171 eQTL hotspots for 16,372 e-traits	Wang et al. (2010)
3. Barley		
Barley1 GeneChip (22,840 probe sets) and cRNA	64/46 SFPs detected from shoot/root datasets, when comparisons were made among 'Golden Promise' and 'Maythorpe'	Walia et al. (2007)
Barley1 GeneChip and cRNA	924 of 1,257 genes assigned to chromosomes with the help of SFPs	Bilgic et al. (2007)

(continued)

Table 2.4 (continued)

Plant species and resources used	Salient features of study	References
Barley1 GeneChip and cRNA	Mapped >2,000 transcript derived markers (TDMs; including both SFPs and GEMs) and 23,738 eQTLs	Potokina et al. (2008)
Barley1 GeneChip and cRNA	>4,000 SFPs identified between 'Steptoe' and 'Morex', and segregation studied using DH population (Steptoe × Morex)	Luo et al. (2007)
Barley 1 GeneChip and cRNA	10,504 SFPs identified between 'Golden Promise' and 'Morex'; also compared with known SNPs	Rostoks et al. (2005)
Barley 1 GeneChip and cRNA	2007 SFPs identified between 'Steptoe' vs 'Morex', 'Morex' vs 'Barke', and Oregon Wolfe Barley Dominant vs Recessive; 80 % were confirmed by direct sequencing	Cui et al. (2005)
4. Maize		
GeneChip Genome Array (17,555 probe sets; 17,477 probe sets with 15 probe pairs, and 78 with 14 or less probe pairs)	Assessment of various target preparation and hybridization methodologies (e.g., cRNA, methylation filtration, high C_0t and AFLP) using three diverse maize inbred lines	Gore et al. (2007)
GeneChip Genome Array	34,034 SFPs identified and mapped using an intervarietal mapping population; mapped loci validated through sequencing	Zhu et al. (2006)
Affymetrix CornChip0 (8403 probe sets) and cRNA	36,196 SFPs identified among 'B73' (reference genotype) and three US maize lines: Mo17, Wf9-BG and W23	Kirst et al. (2006)

maize SNPs identified via comparative next-generation transcriptome sequencing. Approximately 75 % of these SNPs were successfully converted into genetic markers that can be scored reliably and used to generate a SNP- based genetic map by genotyping recombinant inbred lines derived from the popular cross B73 × Mo17 (Liu et al. 2010). In barley, in order to determine identity of 60 Australian varieties, a high-throughput multiplexed SNP genotyping assay was developed using Sequenom MassARRAY and iPLEX™ Gold genotyping systems. As a result, a unique identifier (barcode) of up to 20 SNPs was established for each of the 60 studied varieties (Pattemore and Henry 2008). Similarly, in rice identity of the functional polymorphism in genes influencing different aspects of salt tolerance was determined by combining genetic mapping and transcriptome profiling of bulked RILs (having extreme phenotypes) using Sequenom MALDI-TOF MassARRAY system (Pandit et al. 2010). In wheat, iPlex has been used to genotype 47 wheat SNPs on 1,314 lines (Berard et al. 2009). The agreement of the genotypes obtained by iPlex with the results obtained by different validation methods (sequencing or SNPlex™) was 96 % showing that it can be used successfully in polyploid plants. Mass spectrometry was also used to evaluate the SNP diversity within genes related to bread making quality (*Glu* and *SPA*) on a set 113 lines (Ravel et al. 2007). Similarly, iPlex was tested in a tetraploid wheat, *Triticum durum* x *T. dicoccoides* F_2 population, and was shown to be efficient even for discrimination of heterozygotes (Paux et al. 2012).

2.6.5 Restriction-Site Associated DNA (RAD) Markers in Cereals

RAD markers have witnessed a switch from the low cost microarray-based genotyping platforms to next-generation sequencing based detection procedures. This shift is mainly attributed to the drop in sequencing cost, ease and time for genotyping. RAD sequencing is a form of genotyping by sequencing method, which has recently been put to a variety of applications including genetic mapping and QTL analysis in wide range of organisms (Rowe et al. 2011). The RAD sequencing has provided a method for the discovery of thousands of SNPs. For instance, in barley, a total of 530 SNPs were identified from initial scans of the Oregon Wolf Barley parental inbred lines, and scored in a 93 member doubled haploid (DH) mapping population. RAD sequence data from the DH population was used for genetic map construction. The assembled RAD-only map consists of 445 markers with an average interval length of 5 cM. Sequenced RAD markers are distributed across all seven chromosomes, with polymorphic loci originating from both coding and noncoding regions in the genome (Chutimanitsakun et al. 2011). Similarly, in *Lolium perenne* SSR and STS markers were combined with the RAD markers to produce maps for the female (738 cM) and male (721 cM) parents, and QTLs were identified for resistance to stem rust caused by *Puccinia graminis* subsp. *graminicola* (Pfender et al. 2011). RAD tags were also generated from the

genomic DNA of a pair of eggplant mapping parents. The resulting non-redundant genomic sequence dataset consisted of ~45,000 sequences, of which ~29 % were putative coding sequences and ~70 % were common between the mapping parents. The shared sequences allowed the discovery of ~10,000 SNPs and nearly 1,000 indels, equivalent to a SNP frequency of 0.8 per Kb and an indel frequency of 0.07 per Kb (Barchi et al. 2011).

2.6.6 Use of CNVs and PAVs as Markers in Cereals

CNVs, PAVs and InDels are a new class of markers that are based on microarrays (making use of array-based CGH) and have been extensively used in humans. These relatively new marker types are now being increasingly used in cereals also, and are likely to be preferred over other marker systems in future. Some of the studies already conducted involving these new marker types are briefly reviewed in this section.

CNVs and PAVs in Rice

A high-density oligonucleotide aCGH microarray (containing 7,18,256 oligonucleotide probes) was used in rice to estimate the number of CNVs between the genomes of two cultivars, Nipponbare and Guang-lu-ai4. These CNVs involved known genes, and may be linked to variation among rice varieties, thus contributing to species-specific characteristics (Yu et al. 2011).

CNVs and PAVs in Maize

Whole-genome aCGH was also used for the analysis of CNVs and PAVs in maize. In one study, Mo17 was compared with B73 (Springer et al. 2009) using a microarray with 2.1 million probes developed by Roche NimbleGen, and in another study, 13 inbred lines were compared with the same standard genotype B73 (Belo et al. 2010). For this purpose, high-density microarrays developed by Roche NimbleGen (2.1 million probes) and Agilent (~60,500 probes) were utilized. The two studies revealed a fairly high level of structural diversity between the inbred lines. Several hundred CNVs and thousands of PAVs, distributed over all the chromosome arms, were identified. In yet another study, aCGH was used to compare gene content and CNVs among 19 diverse maize inbreds and 14 genotypes of the wild ancestor of maize, teosinte. CNVs in hundreds of genes were identified, and it was shown that no strong selection for or against CNVs/PAVs accompanied domestication (Swanson-Wagner et al. 2010), although these were shown to contribute to significant quantitative variations.

2.7 Summary and Outlook

The use of array-based genotyping platforms and next-generation sequencing methods for the development and use of third and fourth generation markers has already overwhelmed the plant breeding programs. This became possible due to the generation of data by these platforms in a cost and time-effective manner. These methods have revolutionized plant biology, by bringing precision to whole genome association and linkage studies. These advances in marker-technology equipped the plant breeders with tools to engineer cultivar genotypes with the desired attributes following the concept of 'Breeding by Design' (Peleman and van der Voort 2003). The array-based markers have been put to a variety of applications including genetic linkage and association mapping, diversity and LD studies, gene/QTLs cloning studies, and more recently to anchor BAC-based physical maps with the genetic linkage maps.

The use of array-based platforms for physical and comparative mapping has also improved our understanding of gene and genome organization in major cereals. For instance, Rustenholz et al. (2010) explored the possibility of using barley transcript genetic maps as a surrogate to anchor and order the wheat physical contigs by hybridizing 60 three-dimensional (plate, row, column) BAC pools representing the minimal tiling path (MTP) of wheat chromosome 3B onto barley Agilent 15 K unigene microarray. This has allowed localization of genes along chromosome 3B. The results showed that such barley-wheat cross-hybridizations represent a high throughput cost-efficient approach for anchoring genes on wheat physical maps and for performing comparative genomics studies between wheat and other grass genomes. This study has also led to fine mapping of 738 barley orthologous genes on wheat chromosome 3B. In addition, comparative analyses revealed that 68 % of the genes identified were syntenic between the wheat chromosome 3B and barley chromosome 3H and 59 % were syntenic between wheat chromosome 3B and rice chromosome 1. Later, a subset of 9,216 BACs representing the MTP of the new version of the 3B physical map was pooled into 64 three-dimensional (plate, row, and column) pools and hybridized onto a newly developed wheat NimbleGen 40 K unigene microarray. This not only improved the physical map of chromosome 3B but also allowed mapping of almost 3,000 genes on this chromosome. The expression pattern of these genes was also studied in 15 different conditions. This transcription map of chromosome 3B confirmed that 70 % of the genes are organized in islands that are responsible for an increasing gradient of gene density observed from the centromere to the telomeres. By studying their expression, and putative function, it has been concluded that the gene islands are enriched significantly in genes sharing the same function or expression profile, thereby suggesting that genes in islands acquired shared regulation during evolution (Rustenholz et al. 2011). Similarly, a Morex BAC library in barley consisting of 147,840 clones was pooled into 55 Super Pools (SPs) having seven 384-well plates per SP. The plate, row, and column pools from each SP were further pooled, respectively, into five Matrix Plate Pools (MPPs), eight Matrix Row Pools (MRPs), and 10 Matrix Column Pools (MCPs), giving a

total of 23 Matrix Pools (MPs), which were hybridized on to an Agilent 44 K barley microarray, representing 42,302 expressed genes (Liu et al. 2011). These BACs from multidimensional pools of BAC clones were also incorporated into the HICF physical map of barley. By using array hybridization in combination with next-generation sequencing, and systematic exploitation of conserved synteny with model grasses (rice, sorghum and *Brachypodium distachyon*) 21,766 of the estimated 32,000 barley genes were assigned to individual chromosome arms and their linear order was determined (Mayer et al. 2011). More recently, the array-based markers have also facilitated fine mapping and cloning of several genes contributing to a number of morphological traits in barley. For instance, cloning of *Mat-A* (Zakhrabekova et al. 2012) and *Intermedium-C* (Ramsay et al. 2011) genes respectively responsible for short-season adaptation and spikelet fertility, respectively and fine mapping of *ANT2* (Cockram et al. 2010), *TRD1, VRS1, UZU, NUD1* and *WAXY* genes (Druka et al. 2011) facilitated by the use of GG technology.

References

Akbari M, Wenzl P, Caig V, Carlig J, Xia L, Yang S, Uszynski G, Mohler V, Lehmensiek A, Kuchel H et al (2006) Diversity arrays technology (DArT) for high-throughput profiling of the hexaploid wheat genome. Theor Appl Genet 113:1409–1420

Seifollah K, Alina A, Eduard A (2013) Application of next-generation sequencing technologies for genetic diversity analysis in cereals. In: Gupta PK, Varshney RK (eds) Cereal Genomics-II, Springer

Akhunov E, Nicolet C, Dvorak J (2009) Single nucleotide polymorphism genotyping in polyploid wheat with the Illumina GoldenGate assay. Theor Appl Genet 119:507–517

Alheit KV, Reif JC, Maurer HP, Hahn V, Weissmann EA, Miedaner T, Würschum T (2011) Detection of segregation distortion loci in triticale (x *Triticosecale* Wittmack) based on a high-density DArT marker consensus genetic linkage map. BMC Genomics 12:380

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

Allen AM, Barker GLA, Berry ST, Coghill JA, Gwilliam R, Kirby S, Robinson P, Brenchley RC, D'Amore R, McKenzie N et al (2011) Transcript-specific, single-nucleotide polymorphism discovery and linkage analysis in hexaploid bread wheat (*Triticum aestivum* L.). Plant Biotechnol J 9:1–14

Alsop BP, Farre A, Wenzl P, Wang JM, Zhou MX, Romagosa I, Kilian A, Steffenson BJ (2011) Development of wild barley-derived DArT markers and their integration into a barley consensus map. Mol Breed 27:77–92

Anthony VM, Ferroni M (2011) Agricultural biotechnology and smallholder farmers in developing countries. Curr Opin Plant Biotechnol 23:1–8

Appleby N, Edwards D, Batley J (2009) New technologies for ultra-high throughput genotyping in plants. In: Somers DJ et al (eds) Methods in Molecular Biology, Plant Genomics, Humana Press, New York, 2008, 19–39

Aslop BP, Kilian A, Carling J, Pickering RA, Steffenson BJ (2007) DArT marker-based linkage analysis and inheritance of multiple disease resistance in a wild x cultivated barley population. In: Plant and Animal Genome XV Conference. San Diego, CA, P333

Badea A, Eudes F, Graf RJ, Laroche A, Gaudet DA, Sadasivaiah RS (2008) Phenotypic and marker-assisted evaluation of spring and winter wheat germplasm for resistance to *Fusarium* head blight. Euphytica 164:803–819

Badea A, Eudes F, Salmon D, Tuvesson S, Vrolijk A, Larsson C-T, Caig V, Huttner E, Kilian A, Laroche A (2011) Development and assessment of DArT markers in triticale. Theor Appl Genet 122:1547–1560

Banks TW, Jordan MC, Somers DJ (2009) Single feature polymorphism mapping in bread wheat (*Triticum aestivum* L.). Plant Genome 2:167–178

Barchi L, Lanteri S, Portis E, Acquadro A, Valè G, Toppino L, Rotino GL (2011) Identification of SNP and SSR markers in eggplant using RAD tag sequencing. BMC Genomics 12:304

Beattie AD, Edney MJ, Scoles GJ, Rossnagel BG (2010) Association mapping of malting quality data from western Canadian two-row barley cooperative trails. Crop Sci 50:1649–1663

Beló A, Beatty MK, Hondred D, Fengler KA, Li B, Rafalski A (2010) Allelic genome structural variations in maize detected by array comparative genome hybridization. Theor Appl Genet 120:355–367

Berard A, Le Paslier MC, Dardevet M, Exbrayat-Vinson F, Bonnin I, Cenci A et al (2009) High-throughput single nucleotide polymorphism genotyping in wheat (*Triticum spp.*). Plant Biotechnol J 7:364–374

Bernardo AN, Bradbury PJ, Ma H, Hu S, Bowden RL, Buckler ES, Bai G (2009) Discovery and mapping of single feature polymorphisms in wheat using Affymetrix arrays. BMC Genomics 10:251

Bhat PR, Lukaszewski A, Cui X, Xu J, Svensson JT, Wanamaker S, Waines JG, Close TJ (2007) Mapping translocation breakpoints using a wheat microarray. Nucl Acids Res 35:2936–2943

Bilgic H, Cho S, Garvin DF, Muehlbauer GJ (2007) Mapping barley genes to chromosome arms by transcript profiling of wheat-barley ditelosomic chromosome addition lines. Genome 50:898–906

Bolibok-Brągoszewska H, Heller-Uszyńska K, Wenzl P, Uszyński G, Kilian A, Rakoczy-Trojanowska M (2009) DArT markers for the rye genome: genetic diversity and mapping. BMC Genomics 10:578

Bonman JM, Gu Y, Coleman-Derr D, Jackson EW, Bockelman HE (2011) Inferring geographic origin of barley (*Hordeum vulgare* L. subsp. *vulgare*) accessions using molecular markers. Genet Resour Crop Evol 58:291–298

Borevitz JO, Liang D, Plouffe D, Chang H-S, Zhu T, Weigel D, Berry CC, Winzeler E, Chory J (2003) Large-scale identification of single-feature polymorphisms in complex genomes. Genome Res 13:513–523

Bouchet S, Billot C, Deu M, Rami JF, Xia L, Kilian A, Glaszmann J-C (2007) Whole genome scan and linkage disequilibrium evaluation on a sorghum core collection. In: Plant and Animal Genome XV Conference. San Diego, CA, P365

Brem RB, Yvert G, Clinton R, Kruglyak L (2002) Genetic dissection of transcriptional regulation in budding yeast. Sci 296:752–755

Buescher E, Cui X, Anderson JM (2007) Detecting single-feature polymorphisms on the 7e Thinopyrum chromosome using the wheat oligonucleotide array. In: Plant and Animal Genome XV Conference. San Diego, CA, P185

Chan WC, Nie S (1998) Quantum dot bioconjugates for ultrasensitive nonisotopic detection. Sci 281:2016–2018

Chao S, Dubcovsky J, Dvorak J, Luo M-C, Baenziger SP, Matnyazov R, Clark DR, Talbert LE, Anderson JA, Dreisigacker S et al (2010) Population- and genome-specific patterns of linkage disequilibrium and SNP variation in spring and winter wheat (*Triticum aestivum* L.). BMC Genomics 11:727

Chee M, Yang R, Hubbell E, Berno A, Huang XC, Stern D, Winkler J, Lockhart DJ, Morris MS, Fodor SP (1996) Accessing genetic information with high-density DNA arrays. Sci 274:610–614

Chen H, He H, Zou Y, Chen W, Yu R, Liu X, Yang Y, Gao Y-M, Xu J-L, Fan L-M 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

Chen H, Li J (2007) Nanotechnology: moving from microarrays toward nanoarrays. Methods Mol Biol 381:411–436

Chen W, Mingus J, Mammadov J, Backlund JE, Greene T, Thompson S, Kumpatla S (2010) KASPar: a simple and cost-effective system for SNP genotyping. In: Proceedings of Plant and Animal Genome XVIII conference, San Diego, US, P194

Chutimanitsakun Y, Nipper RW, Cuesta-Marcos A, Cistué L, Corey A, Filichkina T, Johnson EA, Hayes PM (2011) Construction and application for QTL analysis of a restriction site associated DNA (RAD) linkage map in barley. BMC Genomics 12:4

Clark RM, Schweikert G, Ossowski S, Zeller G, Shinn P, Rätsch G, Warthmann N, Fu G, Hinds D, Chen H-M et al (2007) Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*. Sci 317:338–342

Close TJ, Bhat PR, Lonardi S, Wu Y, Rostoks N, Ramsay L, Druka A, Stein N, Svensson JT, Wanamaker S et al (2009) Development and implementation of high-throughput SNP genotyping in barley. BMC Genomics 10:582

Cockram J, White J, Zuluaga DL, Smith D, Comadran J, Macaulay M, Luo Z, Kearsey MJ, Werner P, Harrap D, Tapsell C, Liu H, Hedley PE, Stein N, Schulte D, Steuernagel B, Marshall DF, Thomas WT, Ramsay L, Mackay I, Balding DJ, The AGOUEB Consortium, Waugh R, O'Sullivan DM (2010) Genome-wide association mapping to candidate polymorphism resolution in the unsequenced barley genome. Proc Natl Acad Sci USA 107: 21611–21616

Comadran J, Thomas WT, Eeuwijk FA, Ceccarelli S, Grando S, Stanca AM, Pecchioni N, Akar T, Al-Yassin A, Benbelkacem A et al (2009) Patterns of genetic diversity and linkage disequilibrium in a highly structured *Hordeum vulgare* association-mapping population for the Mediterranean basin. Theor Appl Genet 119:175–187

Cook JP, McMullen MD, Holland JB, Tian F, Bradbury P, Ross-Ibarra J, Buckler ES, Flint-Garcia SA (2012) Genetic architecture of maize kernel composition in the nested association mapping and inbred association panels. Plant Physiol 158:824–834

Coram TE, Settles ML, Wang M, Chen X (2008) Surveying expression level polymorphism and single-feature polymorphism in near-isogenic wheat lines differing for the *Yr5* stripe rust resistance locus. Theor Appl Genet 117:401–411

Crossa J, Burgueno J, Dreisigacker S, Vargas M, Herrera-Foessel SA, Lillemo M, Singh RP, Trethowan R, Warburton M, Franco J et al (2007) Association analysis of historical bread wheat germplasm using additive genetic covariance of relatives and population structure. Genet 177:1889–1913

Cui A, Xu J, Asghar R, Condamine P, Svensson JT, Wanamaker A, Stein N, Roose M, Close TJ (2005) Detecting single-feature polymorphisms using oligonucleotide arrays and robustified projection pursuit. Bioinform 21:3852–3858

Druka A, Franckowiak J, Lundqvist U, Bonar N, Alexander J, Houston K, Radovic S, Shahinnia F, Vendramin V, Morgante M et al (2011) Genetic dissection of barley morphology and development. Plant Physiol 155:617–627

Dunbar SA (2006) Applications of LuminexR xMAPi technology for rapid, high-throughput multiplexed nucleic acid detection. Clin Chim Acta 363:71–82

Eckert AJ, Bower AD, González-Martínez SC, Wegrzyn JL, Coop G, Neale DB (2010) Back to nature: ecological genomics of loblolly pine (*Pinus taeda*, Pinaceae). Mol Ecol 19:3789–3805

Fan JB, Chee MS, Gunderson KL (2006a) Highly parallel genomic assays. Nat Rev Genet 7:632–644

Fan JB, Gunderson KL, Bibikova M, Yeakley JM, Chen J, Wickham Garcia E, Lebruska LL, Laurent M, Shen R, Barker D (2006b) Illumina universal bead arrays. Methods Enzymol 410:57–73

Fan JB, Oliphant A, Shen R, Kermani BG, Garcia F, Gunderson KL, Hansen M, Steemers F, Butler SL, Deloukas P et al (2003) Highly parallel SNP genotyping. Cold Spring Harb Symp on Quant Biol 2003; 67: 69–78

Farkhari M, Lu Y, Shah T, Zhang S, Naghavi MR, Rong T, Xu Y (2011) Recombination frequency variation in maize as revealed by genome wide single-nucleotide polymorphisms. Plant Breed 130:533–539

Feuillet C, Eversole K (2008) Physical mapping of the wheat genome: A coordinated effort to lay the foundation for genome sequencing and develop tools for breeders. Isr J Plant Sci 55:307–313

Flavell AJ, Bolshakov VN, Booth A, Jing AR, Russell J, Ellis THN, Isaac P (2003) A microarray-based high throughput molecular marker genotyping method—the tagged microarray marker (TAM) approach. Nucl Acids Res 31:e115

Flavell AJ, Knox MR, Pearce SR, Ellis TH (1998) Retrotransposon-based insertion polymorphisms (RBIP) for high throughput marker analysis. Plant J 16:643–650

Francki MG, Walker E, Crawford AC, Broughton S, Ohm HW, Barclay I, Wilson RE, McLean R (2009) Comparison of genetic and cytogenetic maps of hexaploid wheat (*Triticum aestivum* L.) using SSR and DArT markers. Mol Genet Genomics 281:181–191

Ganal MW, Durstewitz G, Polley A, Bérard A, Buckler ES, Charcosset A, Clarke JD, Graner E-M, Hansen M, Joets J et al (2011) A large maize (*Zea mays* L.) SNP genotyping array: Development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. PLoS One 6:e28334

Gore M, Bradbury P, Hogers R, Kirst M, Verstege E, van Oeveren J, Peleman J, Buckler E, Eijk MV (2007) Evaluation of target preparation methods for single-feature polymorphism detection in large complex plant genomes. Crop Sci 47:135–148

Grewal TS, Rossnagel BG, Pozniak C, Scoles GJ (2008) Mapping quantitative trait loci associated with barley net blotch resistance. Theor Appl Genet 116:529–539

Grewal TS, Rossnagel BG, Scoles GJ (2012) Mapping quantitative trait loci associated with spot blotch and net blotch resistance in a doubled-haploid barley population. Mol Breed 30:267–279

Gunderson KL, Kuhn KM, Steemers FJ, Ng P, Murray SS, Shen R (2006) Whole-genome genotyping of haplotype tag single nucleotide polymorphisms. Pharmacogenomics 7:641–648

Gupta PK, Balyan HS, Sharma PC, Ramesh B (1996) Microsatellite in plants-a new class of molecular markers. Curr Sci 70:45–54

Gupta PK, Kumar J, Mir RR, Kumar A (2010a) Marker-assisted selection as a component of conventional plant breeding. Plant Breed Rev 33:145–217

Gupta PK, Roy JK, Prasad M (1999a) DNA chips, microarrays and genomics. Curr Sci 77:875–884

Gupta PK, Roy JK, Prasad M (2001) Single nucleotide polymorphisms: a new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. Curr Sci 80:524–535

Gupta PK, Rustgi S (2004) Molecular markers from the transcribed/expressed region of the genome in higher plants. Funct Integr Genomics 4:139–162

Gupta PK, Rustgi S, Mir RR (2008) Array-based high-throughput DNA markers for crop improvement. Heredity 101:5–18

Gupta PK, Varshney RK (2000) The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica 113:163–185

Gupta PK, Varshney RK, Prasad M (2002) Molecular Markers: Principles and methodology. In: Jain SM, Ahloowalia BS and Brar DS (eds), "Molecular Techniques in Crop Improvement" Kluwer Academic Publishers. Netherlands 2002:9–54

Gupta PK, Varshney RK, Sharma PC, Ramesh B (1999b) Molecular markers and their applications in wheat breeding. Plant Breed 118:369–390

Gupta PK, Vrashney RK (2004) Cereal genomics: an overview. In: Gupta PK, Varshney RK (eds) Cereal genomics. Kluwer Academic Publishers, The Netherlands, pp 1–18

Gupta PK, Langridge P, Mir RR (2010b) Marker-assisted wheat breeding: present status and future possibilities. Mol Breed 26:145–161

Gurung S, Mamidi S, Bonman JM, Jackson EW, del Rıo LE, Acevedo M, Mergoum M, Adhikari TB (2011) Identification of novel genomic regions associated with resistance to *Pyrenophora tritici-repentis* races 1 and 5 in spring wheat landraces using association analysis. Theor Appl Genet 123:1029–1041

Hao Z, Li X, Xie C, Weng J, Li M, Zhang D, Liang X, Liu L, Liu S, Zhang S (2011) Identification of functional genetic variations underlying drought tolerance in maize using SNP markers. J Integr Plant Biol 53:641–652

Haun WJ, Hyten DL, Xu WW, Gerhardt DJ, Albert TJ, Richmond T, Jeddeloh JA, Jia G, Springer NM, Vance CP et al (2011) The composition and origins of genomic variation among individuals of the soybean reference cultivar Williams 821. Plant Physiol 155:645–655

Hearnden PR, Eckermann PJ, McMichael GL, Hayden MJ, Eglinton JK, Chalmers KJ (2007) A genetic map of 1,000 SSR and DArT markers in a wide barley cross. Theor Appl Genet 115:383–391

Hinds DA, Stuve LL, Nilsen GB, Halperin E, Eskin E, Ballinger DG, Frazer KA, Cox DR (2005) Whole-genome patterns of common DNA variation in three human populations. Sci 307:1072–1079

Hoffmann M, Hurlebaus J, Weilke C (2007) Novel methods for high-performance melting curve analysis using the Light Cycler® 480 system. Biochemica 1:17–19

Huang X, Feng Q, Qian Q, Zhao Q, Wang L, Wang A, Guan J, Fan D, Weng Q, Huang T et al (2009) High-throughput genotyping by whole-genome resequencing. Genome Res 19:1068–1076

Huttner E, Caig V, Carling J, Evers M, Howes N, Uszynski G, Wenzl P, Xia L, Yang S, Risterucci A-M et al (2006) New plant breeding strategies using an affordable and effective whole-genome profiling method. BioVis Alex 26–29:P73

Huttner E, Risterucci AM, Hippolyte I, Caig V, Carling J, Evers M, Uszynski G, Wenzl P, Glaszmann J-C, Kilian A (2007) Establishment of diversity arrays technology for whole-genome profiling of banana. In: Plant and Animal Genome XV Conference,. San Diego, CA, W34

Huttner E, Wenzl P, Akbari M, Caig V, Carling J, Cayla C, Evers M, Jaccoud D, Peng K, Patarapuwadol S et al (2005) Diversity arrays technology: a novel tool for harnessing the genetic potential of orphan crops. In: Serageldin I, Persley GJ (eds) Discovery to delivery: BioVision Alexandria 2004, Proceedings of the 2004 conference of the world biological forum. CABI Publishing, UK, pp 145–155

Huynh B-L, Wallwork H, Stangoulis JCR, Graham RD, Willsmore KL, Olson S, Mather DE (2008) Quantitative trait loci for grain fructan concentration in wheat (*Triticum aestivum* L.). Theor Appl Genet 117:701–709

Ilic K, Thomson MJ, Virk P, Meyers SN, Yi Y, Wang A, Unger MA, Jones RC, McNally KL, Wang J (2011) Low-cost, high-throughput genotyping of rice Germplasm accessions with fluidigm SNPtype™ assays. http://www.fluidigm.com/home/fluidigm/Posters/IRRI_2011_Genotyping_of_Rice.pdf

Ilic K, Zhang D, Wang X, Jones RC, Meinhardt LW, Wang J (2012) Cacao tree Germplasm characterization with 48-SNP genotyping panel using fluidigmSNPtypeTM Assays and dynamic arraytmintegrated fluidic circuits. Plant and Animal Genome XX Conference, San Diego, CA, USA, P01

Ioannou D, Griffin DK (2010) Nanotechnology and molecular cytogenetics: the future has not yet arrived. Nano Rev 1:5117

Jaccoud D, Peng K, Feinstein D, Kilian A (2001) Diversity arrays: a solid state technology for sequence information independent genotyping. Nucl Acids Res 29:e25

Jing H-C, Bayon C, Kanyuka K, Berry S, Wenzl P, Huttner E, Kilian A, Hammond-Kosack KE (2009) DArT markers: diversity analyses, genomes comparison, mapping and integration with SSR markers in *Triticum monococcum*. BMC Genomics 10:458

Jing R, Bolshakov VI, Flavell AJ (2007) The tagged microarray marker (TAM) method for high throughput detection of single nucleotide and indel polymorphisms. Nat Protoc 2:168–177

Jones E, Chu W-C, Ayele M, Ho J, Bruggeman E, Yourstone K, Rafalski A, Smith OS, McMullen MD, Bezawada C et al (2009) Development of single nucleotide polymorphism (SNP) markers for use in commercial maize (*Zea mays* L.) germplasm. Mol Breed 24:165–176

Jordan DR, Hammer GL, Rodgers D, Butler DG, Hunt CH, Collard B, Mace ES (2007) Multi-population to mapping to increase genetic diversity and grain yield in sorghum. In: Plant and Animal Genomes XV Conference. San Diego, CA, P398

Kallionienil A, Kallioniemi O-P, Sudar D, Rutovitz D, Gray JW, Waldman F, Pinkel D (1992) Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. Sci 258:818–821

Kilian A, Huttner E, Wenzl P, Jaccoud D, Carling J, Caig V, Evers M, Heller-Uszynska K, Cayla C, Patarapuwadol S et al (2005) The fast and the cheap: SNP and DArT-based whole genome profiling for crop improvement. In: Tuberosa R, Phillips RL, Gale M (eds). Proceedings of the International Congress in the Wake of the Double Helix: from the Green Revolution to the Gene Revolution, May 27–31, Avenue Media: Bologna, Italy, 2003, pp 443–461

Kim S-H, Bhat PR, Cui X, Walia H, Xu J, Wanamaker S, Ismail AM, Wilson C, Close TJ (2009) Detection and validation of single feature polymorphisms using RNA expression data from a rice genome array. BMC Plant Biol 9:65

Kirchhoff M, Gerdes T, Rose H, Maahr J, Ottesen AM, Lundsteen C (1998) Detection of chromosomal gains and losses in comparative genomic hybridization analysis based on standard reference intervals. Cytometry 31:163–173

Kirst M, Caldo R, Casati P, Tanimoto G, Walbot V, Wise RP, Buckler ES (2006) Genetic diversity contribution to errors in short oligonucleotide microarray analysis. Plant Biotechnol J 4:489–498

Kumar R, Qiu J, Joshi T, Valliyodan B, Xu D, Nguyen HT (2007) Single feature polymorphism discovery in rice. PLoS One 2:e284

Le Couviour F, Faure S, Poupard B, Flodrops Y, Dubreuil P, Praud S (2011) Analysis of genetic structure in a panel of elite wheat varieties and relevance for association mapping. Theor Appl Genet 123:715–727

Lee SH, Neate SM (2007) Molecular mapping of *Rsp1, Rsp2*, and *Rsp3* genes conferring resistance to *Septoria* speckled leaf blotch in barley. Phytopathol 97:155–161

Li HB, Vaillancourt R, Mendham NJ, Zhou MX (2008a) Comparative mapping of quantitative trait loci associated with waterlogging tolerance in barley (*Hordeum vulgare* L.). BMC Genomics 9:401

Li J, Båga M, Rossnagel BG, Legge WG, Chibbar RN (2008b) Identification of quantitative trait loci for β-glucan concentration in barley grain. J Cereal Sci 48:647–655

Lichter P et al (2000) Comparative genomic hybridization: uses and limitations. Semin Hematol 37:348–357

Lillemo M, Asalf B, Singh RP, Huerta-Espino J, Chen XM, He ZH, Bjørnstad Å (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

Ling P, Campbell KG, Little LM, Skinner DZ (2006) Service and research for molecular markers development in USDA-ARS western-regional small grain genotyping laboratory. In: Plant & Animal Genome Conference XIV. Town & Country Convention Center: San Diego, CA, 2006, P203

Liu S, Chen HD, Makarevitch I, Shirmer R, Emrich SJ, Dietrich CR, Barbazuk WB, Springer NM, Schnable PS (2010) High-throughput genetic mapping of mutants via quantitative single nucleotide polymorphism typing. Genetics 184:19–26

Liu XS (2007) Getting started in tiling microarray analysis. PLoS Comput Biol 3:1842–1844

Liu H, McNicol J, Bayer M, Morris JA, Cardle L, Marshall DF, Schulte D, Stein N, Shi B-J, Taudien S, Waugh R, Hedley PE (2011) Highly parallel gene-to-BAC addressing using microarrays. BioTechniques 50:165–174

Livak KJ, Marmaro J, Todd JA (1995) Towards fully automated genome-wide polymorphism screening. Nat Genet 9:341–342

Lörz H, Wenzel G (2005) Molecular Marker Systems in Plant Breeding and Crop Improvement. Series: Biotechnology in Agriculture and Forestry, 55:478. Springer-Verlag, New York

Lu Y, Yan J, Guimarães CT, Taba S, Hao Z, Gao S, Chen S, Li J, Zhang S, Vivek BS et al (2009) Molecular characterization of global maize breeding germplasm based on genome-wide single nucleotide polymorphisms. Theor Appl Genet 120:93–115

Lu Y, Zhang S, Shah T, Xie C, Hao Z, Li X, Farkhari M, Ribaut J-M, Cao M, Rong T, Xu Y (2010) Joint linkage–linkage disequilibrium mapping is a powerful approach to detecting quantitative trait loci underlying drought tolerance in maize. Proc Natl Acad Sci USA 107:19585–19590

Lucito R, Healy J, Alexander J, Reiner A, Esposito D, Chi M, Rodgers L, Brady A, Sebat J, Troge J et al (2003) Representational oligonucleotide microarray analysis: a high-resolution method to detect genome copy number variation. Genome Res 13:2291–2305

Luo MC, Deal KR, Akhunov ED, Akhunova AR, Anderson OD, Anderson JA, Blake N, Clegg MT, Coleman-Derr D, Conley EJ et al (2009) Genome comparisons reveal a dominant mechanism of chromosome number reduction in grasses and accelerated genome evolution in Triticeae. Proc Natl Acad Sci USA 106:15780–15785

Luo ZW, Potokina E, Druka A, Wise R, Waugh R, Kearsey MJ (2007) SFP genotyping from affymetrix arrays is robust but largely detects cis-acting expression regulators. Genetics 176:789–800

Mace ES, Kilian E, Halloran K, Xia L, Collard B, Jordan DR (2007) Application of diversity arrays technology (DArT) for sorghum mapping, diversity analysis and breeding. In: Plant and Animal Genome XV Conference. San Diego, CA, P366

Mace ES, Rami JF, Bouchet S, Klein PE, Klein RR, Kilian A, Wenzl P, Xia L, Halloran K, Jordan DR (2009) A consensus genetic map of sorghum that integrates multiple component maps and high-throughput Diversity Array Technology (DArT) markers. BMC Plant Biol 9:13

Mace ES, Xia L, Jordan DR, Halloran K, Parh DK, Huttner E, Wenzl P, Kilian A (2008) DArT markers: diversity analyses and mapping in *Sorghum bicolor*. BMC Genomics 9:26

Mammadov J, Chen W, Mingus J, Thompson S, Kumpatla S (2012) Development of versatile gene-based SNP assays in maize (*Zea mays* L.). Mol Breed 29:779–790

Mammadov JA, Chen W, Ren R, Pai R, Marchione W, Yalcin F, Witsenboer H, Greene TW, Thompson SA, Kumpatla SP (2010) Development of highly polymorphic SNP markers from the complexity reduced portion of maize (*Zea mays* L.) genome for use in marker-assisted breeding. Theor Appl Genet 121:577–588

Mantovani P, Maccaferri M, Sanguineti MC, Tuberosa R, Catizone I, Wenzl P, Thomson B, Carling J, Huttner E, DeAmbrogio E et al. (2008) An integrated DArT–SSR linkage map of durum wheat. Mol Breed 22: 629–648

Matthies IE, van Hintum T, Weise S, Roder MS (2012) Population structure revealed by different marker types (SSR or DArT) has an impact on the results of genome-wide association mapping in European barley cultivars. Mol Breed 30:951–966

Maughan PJ, Smith S, Fairbanks D, Jellen E (2011) Development, characterization, and linkage mapping of single nucleotide polymorphisms in the grain amaranths (*Amaranthus* sp.). Plant Genome 4:92–101

Mayer KF, Martis M, Hedley PE, Simková H, Liu H, Morris JA, Steuernagel B, Taudien S, Roessner S, Gundlach H et al (2011) Unlocking the barley genome by chromosomal and comparative genomics. Plant Cell 23:1249–1263

McCartney CA, Stonehouse RG, Rossnagel BG, Eckstein PE, Scoles GJ, Zatorski T, Beattie AD, Chong J (2011) Mapping of the oat crown rust resistance gene *Pc91*. Theor Appl Genet 122:317–325

McCouch SR, Zhao K, Wright M, Tung C, Ebana K, Thomson M, Reynolds A, Wang D, DeClerck G, Ali ML et al (2010) Development of genome-wide SNP assays for rice. Breed Sci 60:524–535

McMullen MM, Kresovich S, Villeda HS, Bradbury P, Li H, Sun Q, Flint-Garcia S, Thornsberry J, Acharya C, Bottoms C et al (2009) Genetic properties of the maize nested association mapping population. Sci 325:737–740

McNally KL, Childs KL, Bohnert R, Davidson RM, Zhao K, Ulat VJ, Zeller G, Clark RM, Hoen DR, Bureau TE et al (2009) Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. Proc Natl Acad Sci USA 106:12273–12278

Merrill KR, Coleman CE, Ghimire S, Meyer SE (2011) High throughput single nucleotide polymorphism (SNP) Development and genotyping In: Bromus tectorum. Plant and Animal Genome XIX Conference, San Diego, CA, USA, P171

Milczarski P, Bolibok-Brągoszewska H, Myśków B, Stojałowski S, Heller-Uszyńska K, Góralska M, Brągoszewski P, Uszyński G, Kilian A, Rakoczy-Trojanowska M (2011) A high density consensus map of rye (*Secale cereale* l.) based on DArT markers. PLoS One 6:e28495

Miller MR, Atwood TS, Eames BF, Eberhart JK, Yan YL, Postlethwait JH, Johnson EA (2007a) RAD marker microarrays enable rapid mapping of zebrafish mutations. Genome Biol 8:R105

Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA (2007b) Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome Res 17:240–248

Nagasaki H, Ebana K, Shibaya T, Yonemaru J, Yano M (2010) Core single-nucleotide polymorphisms: a tool for genetic analysis of the Japanese rice population. Breed Sci 60:648–655

Neumann K, Kobiljski B, Dencic S, Varshney RK, Borner A (2011) Genome-wide association mapping: a case study in bread wheat (*Triticum aestivum* L.). Mol Breed 27:37–58

Newell MA, Cook D, Tinker NA, Jannink J-L (2011) Population structure and linkage disequilibrium in oat (*Avena sativa* L.): implications for genome-wide association studies. Theor Appl Genet 122:623–632

Nolan JP, Sklar LA (2002) Suspension array technology: evolution of the flat-array paradigm. Trends Biotechnol 20:9–12

Oliver RE, Jellen EN, Ladizinsky G, Korol AB, Kilian A, Beard JL, Dumlupinar Z, Wisniewski-Morehead NH, Svedin E, Coon M et al. (2011) New diversity arrays technology (DArT) markers for tetraploid oat (*Avena magna* Murphy et Terrell) provide the first complete oat linkage map and markers linked to domestication genes from hexaploid *A. sativa* L. Theor Appl Genet 123:1159–1171

Pandit A, Rai V, Bal S, Sinha S, Kumar V, Chauhan M, Gautam RK, Singh R, Sharma PC, Singh AK et al (2010) Combining QTL mapping and transcriptome profiling of bulked RILs for identification of functional polymorphism for salt tolerance genes in rice (*Oryza sativa* L.). Mol Genet Genomics 284:121–136

Parh DK, Jordan DR, Aitken EAB, Mace ES, Jun-ai P, McIntyre CL, Godwin ID (2008) QTL analysis of ergot resistance in sorghum. Theor Appl Genet 117:369–382

Patil N, Berno AJ, Hinds DA, Barrett WA, Doshi JM, Hacker CR, Kautzer CR, Lee DH, Marjoribanks C, McDonough DP et al (2001) Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. Sci 294:1719–1723

Pattemore J, Henry RJ (2008) Sequenom MassARRAY® iPLEX™ Gold SNP genotyping for high throughput variety identification. In: Plant and Animal Genome XVI Conference, Sequenome Workshop, San Diego CA, USA, 12–6

Paux E, Faure S, Choulet F, Roger D, Gauthier V, Martinant J-P, Sourdille P, Balfourier F, Le Paslier M-C, Cakir CM et al (2010) Insertion site-based polymorphism markers open new perspectives for genome saturation and marker-assisted selection in wheat. Plant Biotechnol J 8:196–210

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

Paux E, Sourdille P, Salse J, Saintenac C, Choulet F, Leroy P, Korol A, Michalak M, Kianian S, Spielmeyer W et al (2008) A physical map of the 1-gigabase bread wheat chromosome 3B. Sci 322:101–104

Peleg Z, Saranga Y, Suprunova T, Ronin Y, Röder MS, Kilian A, Korol AB, Fahima T (2008) High-density genetic map of durum wheat × wild emmer wheat based on SSR and DArT markers. Theor Appl Genet 117:103–115

Peleman JD, van der Voort JR (2003) Breeding by design. Trends Plant Sci 8:330–334

Pfender WF, Saha MC, Johnson EA, Slabaugh MB (2011) Mapping with RAD (restriction-site associated DNA) markers to rapidly identify QTL for stem rust resistance in *Lolium perenne*. Theor Appl Genet 122:1467–1480

Potokina E, Druka A, Luo Z, Wise R, Waugh R, Kearsey M (2008) Gene expression quantitative trait locus analysis of 16,000 barley genes reveals a complex pattern of genome-wide transcriptional regulation. Plant J 53:90–101

Prasanna BM, Hoisington D (2003) Molecular breeding for maize improvement: an overview. Ind J Biotechnol 2:85–98

Prins R, Pretorius ZA, Bender CM, Lehmensiek A (2011) QTL mapping of stripe, leaf and stem rust resistance genes in a Kariega × Avocet S doubled haploid wheat population. Mol Breed 27:259–270

Ragoussis J (2009) Genotyping technologies for genetic research. Annu Rev Genomics Hum Genet 10:117–133

Raman H, Rahman R, Luckett D, Raman R, Bekes F, Láng L, Bedo Z (2009) Characterisation of genetic variation for aluminium resistance and polyphenol oxidase activity in genebank accessions of spelt wheat. Breed Sci 59:373–381

Ramsay L, Comadran J, Druka A, Marshall DF, Thomas WT, Macaulay M, MacKenzie K, Simpson C, Fuller J, Bonar N et al (2011) *INTERMEDIUM-C*, a modifier of lateral spikelet fertility in barley, is an ortholog of the maize domestication gene *TEOSINTE BRANCHED 1*. Nat Genet 43:169–172

Ravel C, Praud S, Canaguier A, Dufour P, Giancola S, Balfourier F et al (2007) DNA sequence polymorphisms and their application to bread wheat quality. Euphytica 158:331–336

Rheault ME, Dallaire C, Marchand S, Zhang L, Lacroix M, Belzile F (2007) Using DArT and SSR markers for QTL mapping of *Fusarium* head blight resistance in six-row barley. In: Plant and Animal Genome XV Conference. San Diego, CA, P335

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

Rostocks N, Ramsay L, MacKenzie K, Cardle L, Bhat PR, Roose ML, Svensson JT, Stein N, Varshney RK, Marshall DF, Graner A, Close TJ, Waugh R (2006) Recent history of artificial outcrossing facilitates whole genome association mapping in elite crop varieties. Proc Natl Acad Sci USA 103:18656–18661

Rostoks N, Borevitz JO, Hedley PE, Russell J, Mudie S, Morris J, Cardle L, Marshall DF, Waugh R (2005) Single-feature polymorphism discovery in the barley transcriptome. Genome Biol 6:R54

Rowe HC, Renaut S, Guggisberg A (2011) RAD in the realm of next-generation sequencing technologies. Mol Ecol 20:3499–3502

Roy JK, Smith KP, Muehlbauer GJ, Chao S, Close TJ, Steffenson BJ (2010) Association mapping of spot blotch resistance in wild barley. Mol Breed 26:243–256

Rustenholz C, Choulet F, Laugier C, Safar J, Simkova H, Dolezel J, Magni F, Scalabrin S, Cattonaro F, Vautrin S et al (2011) A 3,000-loci transcription map of chromosome 3B unravels the structural and functional features of gene islands in hexaploid wheat. Plant Physiol 157:1596–1608

Rustenholz C, Hedley PE, Morris J, Choulet F, Feuillet C, Waugh R, Paux E (2010) Specific patterns of gene space organization revealed in wheat by using the combination of barley and wheat genomic resources. BMC Genomics 11:714

Ryan PR, Raman H, Gupta S, Horst WJ, Delhaize E (2009) A second mechanism for aluminum resistance in wheat relies on the constitutive efflux of citrate from roots. Plant Physiol 149:340–351

Sansaloni C, Petroli C, Jaccoud D, Carling J, Detering F, Grattapaglia D (2011) Diversity Arrays Technology (DArT) and nextgeneration sequencing combined: genome-wide, high throughput, highly informative genotyping for molecular breeding of Eucalyptus. BMC Proceedings 5(Suppl 7):P54

Sato K, Close TJ, Bhat P, Munoz-Amatriain M, Muehlbauer GJ (2011) Single nucleotide polymorphism mapping and alignment of recombinant chromosome substitution lines in barley. Plant Cell Physiol 52:728–737

Schrider DR, Hahn MW (2010) Gene copy-number polymorphism in nature. Proceedings of the Royal Society B: Biological Sciences. 277: 3213–3221

Semagn K, Bjornstad A, Skinnes H, Maroy AG, Tarkegne Y, William M (2006) Distribution of DArT, AFLP, and SSR markers in a genetic linkage map of a doubled-haploid hexaploid wheat population. Genome 49:545–555

Singh D, Kumar A, Sirohi A, Kumar P, Singh J, Kumar V, Jindal A, Kumar S, Kumar N, Kumar V et al (2011) Improvement of Basmati rice (*Oryza sativa* L.) using traditional breeding technology supplemented with molecular markers. African J Biotechnol 10:499–506

Singh PK, Mergoum M, Adhikari TB, Shah T, Ghavami F, Kianian SF (2010) Genetic and molecular analysis of wheat tan spot resistance effective against *Pyrenophora tritici-repentis* races 2 and 5. Mol Breed 25:369–379

Somers DJ, Jordan MC, Banks TW (2008) Single feature polymorphism discovery using the affymetrix wheat Gene-Chip. In: Plant and Animal Genome XVI Conference. San Diego, CA, P272

Springer NM, Ying K, Fu Y, Ji T, Yeh C-T, Jia Y, Wu W, Richmond T, Kitzman J, Rosenbaum H et al (2009) Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. PLoS Genet 5:e1000734

Steemers FJ, Gunderson KL (2007) Whole genome genotyping technologies on the BeadArray platform. Biotechnol J 2:41–49

Steffenson BJ, Oliver P, Roy JK, Jin Y, Smith KP, Muehlbauer GJ (2007) A walk on the wild side: mining wild wheat and barley collections for rust resistance genes. Aus J Agric Res 58:1–13

Swanson-Wagner RA, Eichten SR, Kumari S, Tiffin P, Stein JC, Ware D, Springer NM (2010) Pervasive gene content variation and copy number variation in maize and its undomesticated progenitor. Genome Res 20:1689–1699

Syvanen AC (2005) Toward genome-wide SNP genotyping. Nat Genet 37:S5–S10

Thomson MJ, Zhao K, Wright M, McNally KL, Rey J, Tung C-W, Reynolds A, Scheffler B, Eizenga G, McClung A et al (2012) High-throughput single nucleotide polymorphism genotyping for breeding applications in rice using the BeadXpress platform. Mol Breed 29:875–886

Tinker NA, Kilian A, Wight CP, Heller-Uszynska K, Wenzl P, Rines HW, Bjørnstad Å, Howarth CJ, Jannink J-L, Anderson JM et al (2009) New DArT markers for oat provide enhanced map coverage and global germplasm characterization. BMC Genomics 10:39

Trebbi D, Maccaferri M, de Heer P, Sørensen A, Giuliani S, Salvi S, Sanguineti MC, Massi A, van der Vossen EAG, Tuberosa R (2011) High-throughput SNP discovery and genotyping in durum wheat (*Triticum durum* Desf.). Theor Appl Genet 123:555–569

Tung CW, Zhao K, Wright K, Ali L, Jung J, Kimball J, Tyagi W, Thomson M, McNally KL, Leung H et al (2010) Development of a research platform for dissecting phenotype-genotype associations in rice (*Oryza* spp.). Rice 23:205–217

Tyrka M, Bednarek PT, Kilian A, Wędzony M, Hura T, Bauer E (2011) Genetic map of triticale compiling DArT, SSR, and AFLP markers. Genome 54:391–401

Varshney RK (2010) Gene-based marker systems in plants: High throughput approaches for marker discovery and genotyping. In: Molecular techniques in crop improvement. Jain SM, Brar DS (eds) 2nd edn, Springer, New York

Venkatasubbarao S (2004) Microarrays: status and prospects. Trends Biotechnol 22:630–637

Vogel N, Schiebel K, Humeny A (2009) Technologies in the whole-genome age: MALDI-TOF_ based genotyping. Transfus Med Hemother 36:253–262

Walia H, Wilson C, Condamine P, Ismail AM, Xu J, Cui X, Close TJ (2007) Array-based genotyping and expression analysis of barley cv, Maythorpe and Golden Promise. BMC Genomics 8:87

Wang J, Kong L, Zhao S, Zhang H, Tang L, Li Z, Gu X, Luo J, Gao G (2011) Rice-Map: a new-generation rice genome browser. BMC Genomics 12:165

Wang J, Yu H, Xie W, Xing Y, Yu S, Xu C, Li X, Xiao J, Zhang Q (2010) A global analysis of QTLs for expression variations in rice shoots at the early seedling stage. Plant J 63:1063–1074

Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10:57–63

Wen W, Araus JL, Shah T, Cairns J, Mahuku G, Bänziger M, Torres JL, Sánchez C, Yan J (2011) Molecular characterization of a diverse maize inbred line collection and its potential utilization for stress tolerance improvement. Crop Sci 51:2569–2581

Wenzl P, Carling J, Kudrna D, Jaccoud D, Huttner E, Kleinhofs A, Kilian A (2004) Diversity arrays technology (DArT) for whole-genome profiling of barley. Proc Natl Acad Sci USA 101:9915–9920

Wenzl P, Catizone I, Thomson B, Huttner E, Mantovani P, Maccaferri M, DeAmbrogio E, Corneti S, Sanguineti MC, Tuberosa R et al (2007) A DArT platform for high throughput profiling of durum wheat. Plant and Animal Genome XV Conference,. San Diego, CA, P263

Wenzl P, Li H, Carling J, Zhou M, Raman H, Paul E, Hearnden P, Maier C, Xia L, Caig V 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

Wenzl P, Raman H, Wang J, Zhou M, Huttner E, Kilian A (2007b) A DArT platform for quantitative bulked segregant analysis. BMC Genomics 8:196

Wenzl P, Suchankova P, Carling J, Simkova H, Huttner E, Kubalakova M, Sourdille P, Paul E, Feuillet C, Kilian A et al (2010) Isolated chromosomes as a new and efficient source of DArT markers for the saturation of genetic maps. Theor Appl Genet 121:465–474

White J, Law JR, MacKay I, Chalmers KJ, Smith JSC, Kilian A, Powell W (2008) The genetic diversity of UK, US and Australian cultivars of *Triticum aestivum* measured by DArT markers and considered by genome. Theor Appl Genet 116:439–453

Xie W, Chen Y, Zhou G, Wang L, Zhang C, Zhang J, Xiao J, Zhu T, Zhang Q (2009) Single feature polymorphisms between two rice cultivars detected using a median polish method. Theor Appl Genet 119:151–164

Xie Y, McNally K, Li C-Y, Leung H, Zhu Y–Y (2006) A High-throughput Genomic Tool: diversity array technology complementary for rice genotyping. J Integr Plant Biol 48:1069–1076

Xu X, Liu X, Ge S, Jensen JD, Hu F, Li X, Dong Y, Gutenkunst RN, Fang L, Huang L et al (2012) Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. Nat Biotechnol 30:105–111

Xu Y (2010) Molecular breeding tools: markers and maps. In: Molecular Plant Breeding. CAB International, Oxford, pp 21–58

Yamamoto T, Nagasaki H, Yonemaru J-I, Ebana K, Nakajima M, Shibaya T, Yano M (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 J, Shah T, Warburton ML, Buckler ES, McMullen MD, Crouch J (2009) Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. PLoS One 4:e8451

Yan J, Yang X, Shah T, Sanchez-Villeda H, Li J, Warburton M, Zhou Y, Crouch JH, Xu Y (2010) High-throughput SNP genotyping with the GoldenGate assay in maize. Mol Breed 25:441–451

Yu L-X, Liu S, Anderson JA, Singh RP, Jin Y, Dubcovsky J, Brown-Guidera G, Bhavani S, Morgounov A, He Z et al (2010) Haplotype diversity of stem rust resistance loci in uncharacterized wheat lines. Mol Breed 26:667–680

Yu P, Wang C, Xu Q, Feng Y, Yuan X, Yu H, Wang Y, Tang S, Wei X (2011) Detection of copy number variations in rice using array-based comparative genomic hybridization. BMC Genomics 12:372

Zakaib GD (2011) Chip chips away at the cost of a genome, Ion-sensing method offers cheap sequencing in record time. Nat 475:278

Zakhrabekova S, Gough SP, Braumann I, Muller AH, Lundqvist J, Ahmann K, Dockter C, Matyszczak I, Kurowska M, Druka A et al (2012) Induced mutations in circadian clock regulator *Mat-a* facilitated short-season adaptation and range extension in cultivated barley. Proc Natl Acad Sci USA 109:4326–4331

Zhang L, Liu D, Guo X, Yang W, Sun J, Wang D, Sourdille P, Zhang A (2011) Investigation of genetic diversity and population structure of common wheat cultivars in northern China using DArT markers. BMC Genet 12:42

Zhang LY, Marchand S, Tinker NA, Belzile F (2009) Population structure and linkage disequilibrium in barley assessed by DArT markers. Theor Appl Genet 119: 43–52

Zhao K, Wright M, Kimball J, Eizenga G, McClung A, Kovach M, Tyagi W, Md Ali L, Tung C-W, Reynolds A et al (2010) Genomic diversity and introgression in *O. sativa* reveal the impact of domestication and breeding on the rice genome. PLoS One 5: e10780

Zhu T, Xia Y, Chilcott C, Dunn M, Dace G, Sessions A, Gayle D, Jon R, John A, Gilles G et al (2006) Maize ultra high-density gene map for genome assisted breeding. In: 48th Annual Maize Genetics Conference, March 9–12, Asilomar Conference Grounds, Pacific Grove, CA, P181

Chapter 3
Sequence Based DNA Markers and Genotyping for Cereal Genomics and Breeding

David Edwards and Pushpendra K. Gupta

3.1 Introduction

The application of molecular markers to advance cereal breeding is now well established, and molecular markers are now used routinely for crop improvement. Applications include: (1) rapid and precise characterization of germplasm; (2) variety distinctiveness, uniformity and stability (DUS) assessment; (3) the selection of parental genotypes within a breeding program; (4) the characterisation of marker-trait associations (MTAs); (5) marker assisted selection (MAS) and the elimination of linkage drag in back-crossing programs; (6) studies of population history, including domestication; (7) the construction of genetic linkage maps; and (8) the analysis of synteny, collinearity and genome rearrangements across species.

During the past three decades, several molecular marker technologies have been developed and applied for plant genome analysis and crop breeding. However, due to a relatively high cost associated with the development of markers, the technology was only applied to a limited number of crop species and for a relatively small number of high value traits (Edwards and Batley 2010). The development of technologies that increase marker throughput and reduce cost will broaden the application of marker technology to more diverse crops and for a greater variety of traits, including complex polygenic traits such as resistance to abiotic stresses. The application of genome-wide association studies (GWAS), linkage disequilibrium (LD), and genomic selection (GS) through prediction of genomic

D. Edwards (✉)
Australian Centre for Plant Functional Genomics and School of Agriculture and Food Sciences, University of Queensland, Brisbane, QLD 4072, Australia
e-mail: Dave.Edwards@uq.edu.au

P. K. Gupta (✉)
Ch Charan Singh University, Meerut, U.P 250004, India
e-mail: pkgupta36@gmail.com

P. K. Gupta and R. K. Varshney (eds.), *Cereal Genomics II*,
DOI: 10.1007/978-94-007-6401-9_3,

estimated breeding values (GEBV) of marker alleles is also being increasingly applied. These activities demand that in any crop, one should be able to develop large number of robust markers, rapidly and at a low cost.

Since 2005, DNA sequencing technology has undergone a revolution termed next or second generation sequencing (NGS). With the rapid growth and plummeting cost of sequence data generation, the discovery and application of sequence-based molecular markers is becoming increasingly common (Berkman et al. 2012a; Imelfort et al. 2009b). The increasing availability and advances in NGS have brought down by several orders of magnitude the cost of sequencing, and it is expected that the volume and quality of sequence data will continue to improve in the years to come. One of the initial concerns with the growth in sequence data production was whether bioinformatics analysis capabilities could match this growth. While there remains a huge potential for advances in bioinformatics analysis, the initial concerns were unfounded, and bioinformatics research, supported by advances in computer hardware, continue to manage and analyse the data flood (Batley and Edwards 2009a; Duran et al. 2009b; Lai et al. 2012a, c; Marshall et al. 2010). There is an increasing number of bioinformatics tools being developed to process, analyse and visualise DNA sequence data (Duran et al. 2009b, 2010a; Edwards 2007, 2011; Edwards and Batley 2004, 2008; Love et al. 2003). Thus, NGS data mining is becoming the default approach for molecular marker discovery in a range of species (Appleby et al. 2009; Edwards et al. 2012a, b; Imelfort et al. 2009b). Large-scale sequence data can also be assembled *de novo* for marker identification (Berkman et al. 2012a; Edwards and Batley 2010; Imelfort et al. 2009a; Imelfort and Edwards 2009).

The methods for high-throughput sequencing can be broadly classified into two groups, the one based on Sanger's dideoxynucleotide synthetic method, which was automated and was extensively used for whole genome sequencing during 1995–2005, and NGS methods. The NGS platforms that are available now for generating sequence data for marker development and genotyping broadly include the following: (1) Roche 454 GS-FLX+, (2) Illumina HiSeq/MiSeq, (3) ABI SOLiD-5500xl (4 hq), (4) Pacific Biosciences, SMART, (5) Helicos Helicoscope; and (6) Ion Torrent (318 chip). Nanopore sequencing machines may also become available (Eisenstein 2012; Pennisi 2012) along with a range of other technologies that are yet to be commercialized.

In this chapter, we will not describe in detail the different sequencing technologies as these have been described in several reviews (Glenn 2011; Gupta 2008; Liu et al. 2012; Metzker 2010; Shendure and Ji 2008) and are also described in another chapter of this book (Seifollah et al. 2013). Several international conferences have also been organized on NGS technologies, so that details of the latest developments are easily accessible. In this chapter, we will focus on the strategies that have been developed for the discovery and use of markers for each of the following marker type: (1) single nucleotide polymorphisms (SNPs), (2) simple sequence repeats (SSRs), (3) insertion site-based polymorphisms (ISBPs), (4) structural variants (SVs) including copy number variations (CNVs) and presence and absence variations (PAVs), and (5) recombination bins as markers.

3.2 Single Nucleotide Polymorphisms

SNPs are biallelic and co-dominant markers and represent the most abundant and high-density DNA-based markers (Batley and Edwards 2007; Edwards et al. 2007a). A variety of methods are available for SNP discovery, and more than 30 different methods have been applied for SNP detection or genotyping (Batley et al. 2007; Chagné et al. 2007; Edwards et al. 2007b; Gupta et al. 2001). The high-density of SNPs makes them valuable for the generation of high-density genetic maps, haplotyping genes or regions of interest, and for map-based positional cloning of QTL/genes. SNPs are now used routinely for genetic diversity analysis, cultivar identification, phylogenetic analysis, characterisation of genetic resources and for detecting marker-trait associations (Gupta et al. 2001; Landjeva et al. 2007; Rafalski 2002), although their use in cereal breeding programs has been relatively slow. As more plant genomes are sequenced and NGS technology becomes routine for SNP discovery and genotyping, these markers will be widely used in almost all plant systems.

3.2.1 Array-Based SNPs versus NGS Based-SNPs

Most of the first and second generation markers (e.g., RFLPs, SSRs and AFLPs) were expensive, laborious and time-consuming. The advent of high-throughput SNP arrays allowed genotyping individuals with hundreds and thousands of SNPs (the third generation markers) in a time and cost-effective manner (Gupta 2008; Gupta et al. 2013). These array-based markers had some limitations, since this technology did not allow the discovery of new SNPs on a large scale. The production of a high-quality array is also relatively expensive and arrays may be biased towards SNPs discovered in the original survey (ascertainment bias).

3.2.2 SNPs Based on First Generation Sequencing (Including Sanger's Method and Sequencing by Hybridization or SBH)

Whole genome sequencing using Sanger technology was applied for several higher plants including Arabidopsis, rice, poplar, grapevine, sorghum, soybean, maize. This provided reference genome sequences for each of these species that could be used to identify DNA polymorphisms. In addition, several studies were also conducted to discover SNPs from EST data or by sequencing PCR amplified products (Table 3.1).

SNPs from Genome Sequencing

The sequencing of rice genomes for Nipponbare and 93–11 led to the identification of 1,703,176 SNPs and 479 indels (Shen et al. 2004), and a set of 384,431

high-quality SNPs and 24,557 single-base indels (Feltus et al. 2004) using the same dataset. The sequences used in these two studies were originally generated using Sanger sequencing technology. A more recent study used re-sequencing microarrays for sequencing by hybridization (SBH) to characterise 20 diverse rice varieties and landraces (McNally et al. 2009). They reported the presence of 160,000 non-redundant SNPs. A major re-sequencing effort, mainly using Sanger technology, was also undertaken by the Oryza Map Alignment Project (OMAP). In this project, BAC endsequences from 11 wild species of rice were aligned to the Nipponbare sequence to identify SNPs (Ammiraju et al. 2006, 2010; Wing et al. 2007). In this manner, even without the use of NGS technology, a comprehensive set of rice SNPs became available for analysis of rice diversity and marker-trait associations.

SNPs Developed from ESTs and Sequencing of PCR Products

In wheat, barley and maize, large numbers of SNPs were discovered by mining Sanger ESTs (Barker et al. 2003; Batley et al. 2003a; Ching et al. 2002; Close et al. 2009; Duran et al. 2009a, c; Kota et al. 2003; Rustgi et al. 2009; Trebbi et al. 2011). The PCR amplification of targeted genome regions has also been used to identify markers; in wheat, the sequences of 21 genes, amplified in 26 wheat lines, were used for the discovery of 64 SNPs (Ravel et al. 2009).

3.2.3 SNPs Based on Second or Next Generation Sequencing

NGS methods have been applied for SNP discovery in a number of cereals and their wild relatives including rice (*Oryza sativa*, *O. rufipogon, O. nivara*), maize (*Zea mays*), wheat (*Tritcum aestivum*, *T. turgidum*) and goat grass (*Aegilops tauschii*) (Huang et al. 2010; Lai et al. 2010; Mammadov et al. 2010; You et al. 2011) (Table 3.1). For rice, the NGS-based discovery of SNPs has been reviewed by McCouch et al. (2010).

In order to discover genomic SNPs in a species, one may either de novo assemble genomes or re-sequence genomes, mapping reads to a reference using either high coverage for SNP discovery or low coverage for genotyping (Edwards and Wang 2012; Imelfort et al. 2009a). In addition, non-reference based approaches have been developed (Azam et al. 2012). Methods have been used to achieve genome reduction based on several approaches including the following: (1) use of restriction enzymes; (2) high-C_0t selection; (3) methylation filtering; (4) sequence capture, (5) RNA-Seq. Among these methods, the more commonly used methods include use of restrictions enzymes, targeted genome capture and RNA-seq.

Restriction Enzyme Based NGS for SNP Discovery and Genotyping

The most common NGS method for SNP genotyping involves restriction digestion of genomic DNA followed by sequencing the ends of the restriction fragments. By

Table 3.1 A summary of studies for sequencing-based development of SNPs in cereals

Crop and species	Genotypes used	Sequencing method	SNPs discovered	References
I. Studies conducted using Sanger's method of sequencing or SBH				
Rice (*Oryza sativa*)	2 subspecies (indica, japonica)	Sanger	2,95,633	Feltus et al. (2004)
	20 varieties and landraces	High-density array (SBH)	1,60,000	McNally et al. (2009)
II. Studies conducted using different NGS platforms				
Maize (*Zea mays*)	B73, Mo17	Roche 454 transcriptome	7,000	Barbazuk et al. (2007)
	B73, Mo17	Roche 454	>2,500	Fu et al. (2010)
	Six inbred lines	Illumina	>1 million	Lai et al. (2010)
	Two inbred lines	Illumina GA IIx (CRoPS)	1,123	van Orsouw et al. (2007), Mammadov et al. (2010)
Rice (*Oryza sativa*)	Two strains	Illumina	67,051	Yamamoto et al. (2010)
	517 landraces		~3.6 million	Huang et al. (2010)
	50 cultivars + 10 wild accessions	Illumina GAIIx	~6.5 million	Xu et al. (2012)
Wheat (*Triticum aestivum*)	Five lines	Illuimina GAIIx	14,078	Allen et al. (2011)
	Four cultivars	Illumina	800,000	Lorenc et al. (2012)
	Eight cultivars	Illumina	99,945	Winfield et al. (2012)
Wheat (*T. durum*)	Four cultivars		2,659	Trebbi et al. (2011)
	Two NILs	Illumina	6,035	Trick et al. (2012)
Wheat (*Ae tauschii*)	Two genotypes	Roche 454, ABI SOLiD	195,631	You et al. (2011)
Rye (*Secale cereale*)	Five inbred lines	Roche 454	5,234	Haseneyer et al. (2011)
Barley (*Hordeum vulgare*)	82 RILs OWB pop^n	Illumina GA IIx (GBS)	34,000	Poland et al. (2012)

selecting methylation sensitive restriction enzymes, it is possible to reduce the representation of repetitive regions of the genome. These methods include: (1) reduced representation sequencing (including reduced representation libraries or RRLs and complexity reduction of polymorphic sequences or CRoPS), and (2) restriction-site associated DNA sequencing (RAD-seq).

Reduced Representation Sequencing

RRLs and CRoPS are two methods of sampling and sequencing a subset of genomic regions, without sampling the entire genome.

1. *Reduced representation libraries* (*RRLs* and HMPR libraries). In this approach, genomic DNA from multiple individuals is digested with a frequent cutter enzyme, and the restricted fragments are pooled. The restriction fragments are selected by size and either end-sequenced or sequenced in their entirety. If a reference genome sequence is available, the reads from RRLs can be mapped to this reference and SNPs can be called. RRLs were first applied to produce a SNP map for the human genome using capillary sequencing (Altshuler et al. 2000). Later, NGS technology was used for sequencing RRLs in a variety of animals and plant systems, including soybean (Hyten et al. 2010a, b).

One of the methods for constructing RRLs involves use of 5-methyl-cytosene sensitive (MCS) restriction enzymes with 4 bp site such as *Hpa*II or *ApeK*1. After partial digestion, it is possible to separate small gene-enriched fragments (<1,000 bp) and eliminate larger fragments (20–150 kb) that contain methylated repeat sequences. These hypomethylated partial restriction (HMPR) libraries were first prepared in maize (Emberton et al. 2005) and exhibited more than 6-fold enrichment for genes. A modified method was used for the construction of HMR libraries of B73 and M017. The modifications involved complete digestion instead of partial digestion, and separation of fragments with size ranging from 100 to 600 bp. These libraries were subjected to 454 sequencing and the sequences were used for SNP discovery leading to the identification of 126, 683 SNPs, mainly from genic regions (Gore et al. 2009).

2. *Complexity reduction of polymorphic sequences* (*CRoPS*). CRoPS is an approach, where complexity reduction is based on a method similar to the Amplified Fragment Length Polymorphism (AFLP) method, in which PCR primers for selective amplification are designed. After PCR amplification, the products from two or more samples are pooled and sequenced. The technology was used for mining 1,200 known SNPs between maize lines B73 and Mo17 (Mammadov et al. 2010; van Orsouw et al. 2007). CRoPS was also applied to tetraploid durum wheat (*Triticum durum*), where four cultivars were screened to identify 2,659 SNPs. After validation, a set of 275 robust SNPs were made available for wheat breeding programs (Trebbi et al. 2011).

Restriction Site-Associated DNA Sequencing (RAD-seq)

This approach involves sequencing of genomic regions flanking restriction sites. The following steps are involved: (1) genomic DNA is digested with a selected

restriction enzyme; (2) restriction fragments are ligated with bar-coded adaptors; (3) the adapter ligated fragments are pooled, randomly sheared and size selected to 300–700 bp; (4) Y-shaped adaptors with divergent ends are ligated to the fragments with and without the first adaptors; (5) the fragments are PCR amplified with primers that are specific to the two classes of adapters; the second adaptor is completed when fragments containing the first adaptor are bound by their primer and copied, and the second primer only binds to completed second adaptors. In this manner, only fragments with both the adaptors (the fragments containing restriction sites) are amplified. Illumina sequencing is performed on the products to identify SNPs (Miller et al. 2007).

Targeted Region-Capture or Enrichment for SNP Discovery and Genotyping

If the sequence of regions of interest are known, these can be used as baits in the form of oligonucleotides for the capturing and enrichment of the regions for sequencing. For example, regions associated with traits can be captured for targeted analysis. An extension of targeted genome capture involves exome capture where a portion or all of the predicted expressed sequences are captured for SNP discovery or genotyping. Hybridization methods for exome capture involve the use of long oligonucleotides either in a solid-phase in the form of microarrays, or in a liquid-phase as biotinylated baits.

The above approach, has been used in maize, leading to identification of 2,500 SNPs (Fu et al. 2010). Similarly in tetraploid wheat, liquid-phase exome capture was applied for the discovery of 4,386 SNPs (Saintenac et al. 2011), and solid phase exome capture was used in hexaploid wheat, leading to discovery of ~100,000 SNPs (Winfield et al. 2012). Large capture designs (~60 Mb) have been developed in barley and wheat as part of two international consortia in collaboration with Roche-Nimblegen, so further reports are expected in the future.

RNA Sequencing (RNA-seq) for SNP Discovery

In addition to gene expression analysis and genome annotation, RNA sequencing has also been applied for SNP discovery. As an example in wheat, >14,000 SNPs were discovered, when cDNA samples from several wheat varieties were sequenced using Illumina sequencing (Allen et al. 2011). In tetraploid wheat, RNA seq was used to identify SNPs between two near-isogenic lines differing across a ~30 cM interval including the *Gpc-B1* locus. Thirty nine new SNPs were identified across a 12.2 cM interval containing the *Gpc-B1* (Trick et al. 2012).

Low Coverage Genotyping by Sequencing (Skim Sequencing)

An alternative approach to reduced representation involves low coverage sequencing. This is suitable for genotyping populations where the parental genotypes are known. In this method, SNPs are usually predicted from the parents using medium coverage

sequencing, followed by low coverage skim sequencing to call the allelic variation in diverse individuals or segregating populations. Advantages of this approach include the relatively simple library preparation compared to complexity reduction and the flexibility of read depth, where following initial skim sequencing, libraries from selected individuals may be sequenced in greater depth for high resolution analysis or further SNP discovery.

3.2.4 Tools for NGS-Based SNP Discovery

There are many tools available for the discovery of SNPs from next generation sequence data, but few have been designed specifically for SNP discovery in cereal populations (Lee et al. 2012).

Software for SNP Discovery in Cereals

An important tool for SNP discovery that is based on autoSNP software (Barker et al. 2003; Batley et al. 2003a) uses redundancy and haplotype co-segregation for SNP discovery. Similarly, AutoSNPdb (Duran et al. 2009a) combines the SNP discovery pipeline of autoSNP with a relational database, hosting information on the polymorphisms, cultivars and gene annotations, to enable efficient mining and interrogation of the data. Users may search for SNPs within genes with specific annotation or for SNPs between defined cultivars. AutoSNPdb was originally developed for rice, barley and Brassica Sanger sequence data (Duran et al. 2009c), but has recently been applied to discover SNPs from wheat 454 data (Lai et al. 2012b) (http://autosnpdb.appliedbioinformatics.com.au/). More recently, second generation sequencing SGSautoSNP has been developed for SNP mining in complex genomes such as Brassica or hexaploid wheat using Illumina sequence data (Lorenc et al. 2012).

SNPs and Sequencing Errors

The challenge of *in silico* SNP discovery is not the identification of polymorphic nucleotide positions, but the differentiation of true inter-varietal polymorphisms from the abundant sequence errors. This is particularly true for NGS data, which generally has a higher error rate than traditional DNA sequencing. NGS remains prone to inaccuracies that are as frequent as one error every 20 bp. These errors impede the electronic mining of this data to identify biologically relevant polymorphisms. There are several different types of error which need to be taken into account when differentiating between sequence errors and true polymorphisms, and the approach is highly dependent on the NGS platform used to generate the data, as they each have distinct error profiles. A major source of sequence error comes from the fine balance between the desire to obtain the greatest sequence length, and the confidence that bases are called correctly. Because of this, sequence trimming, filtering and further processing

is often applied to reduce the abundance of sequence errors. A second cause of error which is particularly an issue with the short reads produced by NGS technology is the incorrect mapping of sequences to a reference. This can occur at any genomic region which has two or more similar copies in the genome, due to the presence of multigene families, genome duplications or polyploidy.

The identification of true polymorphisms in a background of sequence errors can be based on the following four criteria: (1) sequence quality values; (2) redundancy of the polymorphism in an alignment; (3) co-segregation of SNPs to define a haplotype, and (4) specificity of an allele call with a variety. The application of each of these methods also depends on the method of SNP discovery and sequencing technology applied, for example identifying SNPs from the assembly of Roche 454 transcriptome data or from the mapping of paired Illumina data to a reference. By using the various measures of SNP confidence assessment, true SNPs may be identified with reasonable confidence from NGS data. The above four criteria that are used for distinguishing between sequencing error and true SNPs are briefly discussed.

Sequence Quality

Sequence read quality is a basic factor in determining the quality of SNP calling. These quality scores can be applied in two ways. Firstly, low quality data may be removed by trimming and filtering before SNP discovery. This is particularly appropriate for very large data sets. Alternatively, low quality data may be included within the assembly or mapping, but not considered during the SNP calling process.

SNP Redundancy Score

The frequency of occurrence of a polymorphism at a particular locus provides one of the best measures of confidence in the SNP representing a true polymorphism, and is referred to as the SNP redundancy score (Barker et al. 2003). By examining SNPs that have a redundancy score equal to or greater than two (two or more of the aligned sequences represent the polymorphism), the vast majority of sequencing errors are removed. Although some true genetic variation is also ignored due to its presence only once within an alignment, the high degree of redundancy within NGS data permits the rapid identification of large numbers of true SNPs using this approach.

Co-Segregation of SNPs Within a Haplotype

While redundancy based methods for SNP discovery are highly efficient, the non-random nature of sequence error may lead to certain sequence errors being repeated between runs around locations of complex DNA structure. Errors at these loci would have a relatively high SNP redundancy score and appear as confident SNPs. In order to eliminate this source of error, an additional independent SNP confidence measure may be required. This can be determined by the co-segregation of SNPs to define a haplotype. True SNPs that represent divergence

between homologous genes co-segregate to define a conserved haplotype, whereas sequence errors do not co-segregate with a haplotype. Thus, a co-segregation score, based on whether a SNP position contributes to defining a haplotype is a further independent measure of SNP confidence. Determining haplotypes from short read data is challenging as sequence reads rarely traverse multiple SNPs. This is less of an issue for longer sequence reads from the Roche 454 system or in the application of paired reads from the Illumina or AB SOLiD platforms.

Genotype-Specific SNP Alleles

A final assessment of SNP confidence is the definition of a unique variety specific allele at the SNP position. SNP discovery in cereals often uses plant material that is predominantly homozygous across the genome. In these cases, there should be only one allele represented for each variety at any position. This is different from the approach used in human research and so requires the use of custom software. The use of homozygous plant material and the requirement for only a single variety specific allele at a SNP position allows discrimination of true SNPs from errors caused by mismapping of duplicate or similar genomic regions.

3.2.5 Some Examples of NGS-Based SNP Discovery in Cereals and Other Crops

In one of the first examples of cereal SNP discovery from NGS data, more than 7,000 candidate SNPs were identified between maize lines B73 and Mo17, with over 85 % validation rate (Barbazuk et al. 2007). This success is particularly impressive considering the complexity of the maize genome and the early version of Roche 454 sequencing applied, which produced an average read length of only 101 bp. More recent work has attempted to reduce SNP miscalling due to sequence errors (Brockman et al. 2008); this has the potential to improve cereal SNP prediction accuracy using Roche 454 data. The characterisation of linkage blocks assists the application of genomic selection methods for cereal crop improvement (Duran et al. 2010b). Some examples of cereals, where NGS-based SNP development has been reported are described below.

SNPs in Maize (Zea mays)

The large data volumes from Illumina sequencing platforms enable confident discovery of very large numbers of genome-wide SNPs (Imelfort et al. 2009b). Using this platform, more than ~1.2 million SNPs were identified among six elite inbred maize varieties selected for their commercial importance and genetic relationships, (Lai et al. 2010). This study also identified a large number of presence/absence variations (PAVs) which may be associated with heterosis in this species.

SNPs in Wheat (Triticum aestivum)

The large size of wheat genome has led to diverse approaches to reduce the cost of data production. These include the targeted re-sequencing of captured exome fragments (Winfield et al. 2012) and the establishment of consortia to share the cost of genome sequence data generation (Edwards et al. 2012c). In another study, 14,078 putative SNPs were identified across representative samples of UK wheat germplasm using Illumina sequencing of cDNA libraries, with a proportion of these SNPs validated using KASPar assays (Allen et al. 2011). In addition, several efforts for large–scale SNP development in bread wheat were undertaken in USA, U.K., France and Australia, leading to the development of millions of SNPs. These SNPs will be extensively used for molecular breeding in wheat (Lorenc et al. 2012).

SNPs in Rice (Oryza sativa)

In rice, around 3.6 million SNPs were identified by re-sequencing 517 rice landraces (Huang et al. 2010). These SNPs were used for a genome-wide association study (GWAS), which allowed association of SNPs with complex traits. More recently, the genomes of 40 cultivated accessions selected from the major groups of cultivated rice (*Oryza sativa*) and 10 accessions of their wild progenitors (*Oryza rufipogon* and *Oryza nivara*) were re-sequenced, identifying 6.5 million SNPs (Xu et al. 2012).

SNPs in Goat Grass(Aegilops tauschii)

An annotation-based, genome-wide SNP discovery pipeline (called AGSNP) was developed and used for SNP discovery in *Ae. tauschii*, the diploid progenitor of the D genome of hexaploid wheat. In this pipeline, one genotype (AL8/78) was used for generating long reads with a low coverage using Roche 454; these were annotated in order to distinguish single-copy sequences from repetitive sequences. Another genotype (AS75) was used to generate multiple genome equivalents of shotgun reads using SOLiD or Illumina sequencing. These short reads were then mapped to the annotated Roche 454 reads to identify candidate SNPs. A total of 497,118 SNPs were discovered, which included 195,631 SNPs in gene sequences, 155,580 SNPs in uncharacterized single-copy and 145,907 SNPs in repeat junctions. Only 81.3–88.0 % SNPs among three different classes could be validated.

SNPs in Rye (Secale cereale)

Rye genetic and genomic resources are limited relative to those for other members of Triticeae. In a recent study, the transcriptomes of five winter rye inbred lines were sequenced using Roche/454 (Haseneyer et al. 2011). More than 2.5 million reads were assembled into 115,400 contigs. These assemblies were used to identify 5,234 SNPs, which will prove useful for future research involving rye genetics, genomics and the breeding of improved rye cultivars.

3.3 Simple Sequence Repeats

SSRs are short stretches of DNA sequences occurring as tandem repeats of mono-, di-, tri-, tetra-, penta- and hexa-nucleotides. These short repeats have been found to be abundant and dispersed throughout the genomes of all prokaryotes and eukaryotes analysed (Katti et al. 2001; Toth et al. 2000). SSRs are highly polymorphic due to frequent variation in the number of repeat units. SSR markers are co-dominant and multi-allelic in nature and have been shown to be highly reproducible. The hypervariability of SSRs among related organisms makes them excellent markers for a wide range of applications, including genetic mapping, the molecular tagging of genes, genotype identification, the analysis of genetic diversity, phenotype mapping, marker-trait association and marker assisted selection (Powell et al. 1996; Tautz and Schlotterer 1994). In particular, SSRs demonstrate a high degree of transferability between species, as PCR primers designed for a particular SSR within one species would frequently amplify a corresponding locus in related species, enabling comparative genetic and genomic analysis (Gupta et al. 2003). Regions flanking SSRs are also known to be highly polymorphic and a valuable source of SNP polymorphisms (Batley et al. 2003b; Mogg et al. 2002).

Studies of the potential biological function and evolutionary relevance of SSRs is leading to a greater understanding of genomes (Subramanian et al. 2003). SSRs were initially considered to be evolutionally neutral (Awadalla and Ritland 1997), however, evidence suggests an important role in genome evolution (Moxon and Wills 1999). SSRs are also believed to be involved in gene expression and regulatory functions (Gupta et al. 1994; Kashi et al. 1997). There are numerous lines of evidence suggesting that SSRs in non-coding regions may also have functional significance (Mortimer et al. 2005).

3.3.1 SSRs Based on First Generation Sequence Data

Initially, the discovery of SSR loci required the construction of genomic DNA libraries enriched for SSR sequences, followed by DNA sequencing (Edwards et al. 1996). The increasing availability of sequence data made it more economical and efficient to use computational tools to identify SSR loci from the available genomic and transcriptomic sequences (Sharma et al. 2007). Several computational tools were developed for the identification of SSRs within sequence data as well as for the design of PCR primers for the amplification of these SSRs (Jewell et al. 2006; Robinson et al. 2004). Large sets of SSRs have been developed using EST or genomic databases (Batley and Edwards 2009b; Edwards et al. 2009; Gupta et al. 2003; Varshney et al. 2005).

3.3.2 SSRs Based on NGS Data

The identification of SSRs from NGS data has proven to be more challenging than with Sanger sequence data, due to the relatively short length of the reads, but

there are reported examples of successful use of NGS data for SSR development. For example, in wheat, following initial assembly of the data, large numbers of SSRs were identified (Nie et al. 2012). SSRs from several species were identified from Roche 454 data using two SSR enrichment methods (Santana et al. 2009). However, in cereals, NGS has not been applied extensively for the discovery of SSRs, as the emphasis has largely shifted to SNPs. The read length of some of the third generation sequencing technologies is likely to increase the utility of NGS data for SSR marker discovery. NGS can also be used to identify SSR markers in difficult material such as fossil species, as demonstrated by the generation of sequence data and identification of a single SSR marker from the fossil bone of an extinct New Zealand Moa (Allentoft et al. 2009). This opens opportunities to examine cereal genome evolution through genotyping ancient specimens which were around during early cereal domestication.

Due to the redundancy in NGS data, and with datasets often being derived from several distinct cultivars, it is possible to predict the polymorphism of SSRs computationally. Using an extended version of SSR Primer, polymorphic SSRs are distinguished from monomorphic SSRs by the representation of varying motif lengths within an alignment of sequence reads (unpublished results). The identification of SSRs that are predicted to be polymorphic between defined varieties greatly reduces the cost associated with the development of these markers.

3.4 Insertion Site-Based Polymorphisms (ISBP)

ISBP markers are based on polymorphisms surrounding transposable element insertion sites. Roche 454 sequencing allows the high-throughput discovery of ISBP markers (Paux et al. 2010). This group also developed the software 'IsbpFinder.pl' for the identification of ISBP markers. The high repeat content of some of the cereal genomes makes ISBPs an attractive alternative to SSRs and SNPs (Kumar and Hirochika 2001; Schulman et al. 2004). While reads from the current second generation sequencing technologies are too short for the high-throughput discovery of these markers, longer reads from third generation technologies may provide them in abundance. Assemblies from genome shotgun data may also provide a rich source of ISBP markers. Mining of the recent assembly of the wheat chromosome arm, 7BS (Berkman et al. 2011, 2012b) identified more than 10,000 ISBPs that can be used to genetically map the assembly fragments (unpublished results). It is predicted that there may be as many as four million potential ISBP markers in wheat (Paux et al. 2010).

3.5 Structural Variants: CNVs and PAVs

Structural variations (SVs) include indels (insertions/deletions), translocations, inversions, copy number variations (CNVs) and presence-absence variations (PAVs). SVs have been discovered in several plant systems including Arabidopsis,

maize and rice, and it is anticipated that further studies will be published in future. Genomic structural variants have been broadly classified into the following two groups: (1) unbalanced SVs, which involve a difference in the content of DNA sequence (CNVs due to deletions or duplications leading to loss or gain of DNA segments); and (2) balanced SVs, which do not involve any difference in the content, but differ due to the arrangement of DNA sequence, as is the case in inversions and translocations. Structural variations can be detected through any of the following methods: (1) fluorescence in situ hybridization (FISH), (2) array-based comparative genomic hybridization (aCGH), (3) *RT-qPCR*, (4) SNP/SFP genotyping arrays, (5) NGS technologies. The advent of NGS technologies promises to revolutionize structural variation studies. Different available approaches for the identification of SVs have recently been reviewed (Alkan et al. 2011).

3.6 Recombination Bins as Markers

Recombination bins can be determined using SNPs and NGS. A sliding window approach can call recombination breakpoints that define recombination bins (Huang et al. 2009). These recombination bins (each bin spanning the segment between two adjacent recombination break points) can then be used for the construction of a genetic maps and QTL analysis. It is possible that this method of genetic analysis involving the use of recombination bins may eventually replace conventional marker-based genetic analysis.

3.7 Summary and Outlook

During the last three decades, the development of new marker types has been a continuous process. However, a majority of these markers were low throughput (time consuming) and not cost-effective. Many of the early markers were anonymous and could not be related to physical genome segments. These markers are increasingly being replaced by sequence-based markers that can be directly located on reference genomes. Next generation sequencing technology and associated bioinformatics continue to drive discoveries in genetics and genomics and this is likely to continue with the rapid advances in these technologies. The expansion of genotyping by sequencing (GBS) allows the direct link between genetic markers and the physical genome. The drive for using NGS for marker development and genotyping is still continuing. This research trend in crop genetics and genomics, particularly in cereals, is likely to continue with the rapid advances in these technologies. The cost of DNA markers will continue to decline over the coming years, which combined with increasing numbers of reference genome sequences, will greatly expand the applications of genomics for cereal crop improvement and diversity analysis.

References

Seifollah K, Alina A, Akhunov E (2013) Application of next-generation sequencing technologies for genetic diver-sity analysis in cereals. In: Gupta PK, Varshney RK (eds) Cereal genomics II. Springer, Berlin

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

Allen AM, Barker GLA, Berry ST, Coghill JA, Gwilliam R, Kirby S, Robinson P, Brenchley RC, D'Amore R, McKenzie N, Waite D, Hall A, Bevan M, Hall N, Edwards KJ (2011) Transcript-specific, single-nucleotide polymorphism discovery and linkage analysis in hexaploid bread wheat (Triticum aestivum L.). Plant Biotechnol J 9(9):1086–1099

Allentoft ME, Schuster SC, Holdaway RN, Hale ML, McLay E, Oskam C, Gilbert MTP, Spencer P, Willerslev E, Bunce M (2009) Identification of microsatellites from an extinct moa species using high-throughput (454) sequence data. Biotechniques 46:195

Altshuler D, Pollara V, Cowles C, Van Etten W, Baldwin J, Linton L, Lander E (2000) An SNP map of the human genome generated by reduced representation shotgun sequencing. Nature 407:513–516

Ammiraju JSS, Luo M, Goicoechea JL, Wang W, Kudrna D, Mueller C, Talag J, Kim H, Sisneros NB, Blackmon B, Fang E, Tomkins JB, Brar D, MacKill D, McCouch S, Kurata N, Lambert G, Galbraith DW, Arumuganathan K, Rao K, Walling JG, Gill N, Yu Y, SanMiguel P, Soderlund C, Jackson S, Wing RA (2006) The Oryza bacterial artificial chromosome library resource: construction and analysis of 12 deep-coverage large-insert BAC libraries that represent the 10 genome types of the genus Oryza. Genome Res 16:140–147

Ammiraju JSS, Song X, Luo M, Sisneros N, Angelova A, Kudrna D, Kim H, Yu Y, Goicoechea JL, Lorieux M, Kurata N, Brar D, Ware D, Jackson S, Wing RA (2010) The *Oryza* BAC resource: a genus-wide and genome scale tool for exploring rice genome evolution and leveraging useful genetic diversity from wild relatives. Breed Sci 60:536–543

Appleby N, Edwards D, Batley J (2009) New technologies for ultra-high throughput genotyping in plants. In: Somers D, Langridge P, Gustafson J (eds) Plant genomics. Humana Press, New York, pp 19–40

Awadalla P, Ritland K (1997) Microsatellite variation and evolution in the Mimulus guttatus species complex with contrasting mating systems. Mol Biol Evol 14:1023–1034

Azam S, Thakur V, Ruperao P, Shah T, Balaji J, Amindala B, Farmer AD, Studholme DJ, May GD, Edwards D, Jones JDG, Varshney RK (2012) Coverage-based consensus calling (CbCC) of short sequence reads and comparison of CbCC results to identify SNPs in chickpea (Cicer arietinum; Fabaceae), a crop species without a reference genome. Am J Bot 99:186–192

Barbazuk WB, Emrich SJ, Chen HD, Li L, Schnable PS (2007) SNP discovery via 454 transcriptome sequencing. Plant J 51:910–918

Barker G, Batley J, O'Sullivan H, Edwards KJ, Edwards D (2003) Redundancy based detection of sequence polymorphisms in expressed sequence tag data using autoSNP. Bioinformatics 19:421–422

Batley J, Edwards D (2007) SNP applications in plants. In: Oraguzie N, Rikkerink E, Gardiner S, De Silva H (eds) Association mapping in plants. Springer, New York, pp 95–102

Batley J, Edwards D (2009a) Genome sequence data: management, storage, and visualization. Biotechniques 46:333–336

Batley J, Edwards D (2009b) Mining for single nucleotide polymorphism (SNP) and simple sequence repeat (SSR) molecular genetic markers. In: Posada D (ed) Bioinformatics for DNA sequence analysis. Humana Press, New York, pp 303–322

Batley J, Barker G, O'Sullivan H, Edwards KJ, Edwards D (2003a) Mining for single nucleotide polymorphisms and insertions/deletions in maize expressed sequence tag data. Plant Physiol 132:84–91

Batley J, Mogg R, Edwards D, O'Sullivan H, Edwards KJ (2003b) A high-throughput SNuPE assay for genotyping SNPs in the flanking regions of Zea mays sequence tagged simple sequence repeats. Mol Breeding 11:111–120

Batley J, Jewell E, Edwards D (2007) Automated discovery of single nucleotide polymorphism (SNP) and simple sequence repeat (SSR) molecular genetic markers. In: Edwards D (ed) Plant bioinformatics. Humana Press, New York, pp 473–494

Berkman PJ, Skarshewski A, Lorenc MT, Lai K, Duran C, Ling EYS, Stiller J, Smits L, Imelfort M, Manoli S, McKenzie M, Kubalakova M, Simkova H, Batley J, Fleury D, Dolezel J, Edwards D (2011) Sequencing and assembly of low copy and genic regions of isolated *Triticum aestivum* chromosome arm 7DS. Plant Biotechnol J 9:768–775

Berkman PJ, Lai K, Lorenc MT, Edwards D (2012a) Next generation sequencing applications for wheat crop improvement. Am J Bot 99:365–371

Berkman PJ, Skarshewski A, Manoli S, Lorenc MT, Stiller J, Smits L, Lai K, Campbell E, Kubalakova M, Simkova H, Batley J, Dolezel J, Hernandez P, Edwards D (2012b) Sequencing wheat chromosome arm 7BS delimits the 7BS/4AL translocation and reveals homoeologous gene conservation. Theor Appl Genet 124:423–432

Brockman W, Alvarez P, Young S, Garber M, Giannoukos G, Lee WL, Russ C, Lander ES, Nusbaum C, Jaffe DB (2008) Quality scores and SNP detection in sequencing-by-synthesis systems. Genome Res 18:763–770

Chagné D, Batley J, Edwards D, Forster JW (2007) Single nucleotide polymorphism genotyping in plants. In: Oraguzie N, Rikkerink E, Gardiner S, De Silva H (eds) Association mapping in plants. Springer, New York, pp 77–94

Ching A, Caldwell K, Jung M, Dolan M, Smith O, Tingey S, Morgante M, Rafalski A (2002) SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genet 3:19

Close T, Bhat P, Lonardi S, Wu Y, Rostoks N, Ramsay L, Druka A, Stein N, Svensson J, Wanamaker S, Bozdag S, Roose M, Moscou M, Chao S, Varshney R, Szucs P, Sato K, Hayes P, Matthews D, Kleinhofs A, Muehlbauer G, DeYoung J, Marshall D, Madishetty K, Fenton R, Condamine P, Graner A, Waugh R (2009) Development and implementation of high-throughput SNP genotyping in barley. BMC Genomics 10:582

Duran C, Appleby N, Clark T, Wood D, Imelfort M, Batley J, Edwards D (2009a) AutoSNPdb: an annotated single nucleotide polymorphism database for crop plants. Nucleic Acids Res 37:D951–D953

Duran C, Appleby N, Edwards D, Batley J (2009b) Molecular genetic markers: discovery, applications, data storage and visualisation. Curr Bioinform 4:16–27

Duran C, Appleby N, Vardy M, Imelfort M, Edwards D, Batley J (2009c) Single nucleotide polymorphism discovery in barley using autoSNPdb. Plant Biotechnol J 7:326–333

Duran C, Boskovic Z, Imelfort M, Batley J, Hamilton NA, Edwards D (2010a) CMap3D: a 3D visualisation tool for comparative genetic maps. Bioinformatics 26:273–274

Duran C, Eales D, Marshall D, Imelfort M, Stiller J, Berkman PJ, Clark T, McKenzie M, Appleby N, Batley J, Basford K, Edwards D (2010b) Future tools for association mapping in crop plants. Genome 53:1017–1023

Edwards D (2007) Bioinformatics and plant genomics for staple crops improvement. In: Kang MS, Priyadarshan PM (eds) Breeding major food staples. Blackwell, Oxford, pp 93–106

Edwards D (2011) Wheat bioinformatics. In: Bonjean A, Angus W, Van Ginkel M (eds) The world wheat book. Lavoisier, France, pp 851–875

Edwards D, Batley J (2004) Plant bioinformatics: from genome to phenome. Trends Biotechnol 22:232–237

Edwards D, Batley J (2008) Bioinformatics: fundamentals and applications in plant genetics, mapping and breeding. In: Kole C, Abbott AG (eds) Principles and practices of plant genomics. Science Publishers Inc., USA, pp 269–302

Edwards D, Batley J (2010) Plant genome sequencing: applications for crop improvement. Plant Biotechnol J 7:1–8

Edwards D, Wang X (2012) Genome Sequencing Initiatives. In: Edwards D, Parkin IAP, Batley J (eds) Genetics, genomics and breeding of Oilseed Brassicas. Science Publishers Inc., New Hampshire, pp 152–157

Edwards KJ, Barker JHA, Daly A, Jones C, Karp A (1996) Microsatellite libraries enriched for several microsatellite sequences in plants. Biotechniques 20:758

Edwards D, Forster JW, Chagné D, Batley J (2007a) What are SNPs? In: Oraguzie NC, Rikkerink EHA, Gardiner SE, De Silva HN (eds) Association mapping in plants. Springer, New York, pp 41–52

Edwards D, Forster JW, Cogan NOI, Batley J, Chagné D (2007b) Single nucleotide polymorphism discovery. In: Oraguzie N, Rikkerink E, Gardiner S, De Silva H (eds) Association mapping in plants. Springer, New York, pp 53–76

Edwards D, Hansen D, Stajich J (2009) DNA sequence databases. In: Edwards D HDaSJ (ed) Bioinformatics: tools and applications. Springer, Berlin, pp 1–11

Edwards D, Batley J, Snowdon R (2012a) Accessing complex crop genomes with next-generation sequencing. Theor Appl Genet 126:1–11

Edwards D, Henry RJ, Edwards KJ (2012b) Preface: advances in DNA sequencing accelerating plant biotechnology. Plant Biotechnol J 10:621–622

Edwards D, Wilcox S, Barrero RA, Fleury D, Cavanagh CR, Forrest KL, Hayden MJ, Moolhuijzen P, Gagnere GK, Bellgard MI, Lorenc MT, Shang CA, Baumann U, Taylor JM, Morell MK, Langridge P, Appels R, Fitzgerald A (2012c) Bread matters: a national initiative to profile the genetic diversity of Australian wheat. Plant Biotechnol J 10:703–708

Eisenstein M (2012) Oxford nanopore announcement sets sequencing sector abuzz. Nat Biotech 30:295–296

Emberton J, Ma J, Yuan Y, SanMiguel P, Bennetzen JL (2005) Gene enrichment in maize with hypomethylated partial restriction (HMPR) libraries. Genome Res 15:1441–1446

Feltus FA, Wan J, Schulze SR, Estill JC, Jiang N, Paterson AH (2004) An SNP resource for rice genetics and breeding based on subspecies Indica and Japonica genome alignments. Genome Res 14:1812–1819

Fu Y, Springer NM, Gerhardt DJ, Ying K, Yeh C-T, Wu W, Swanson-Wagner R, D'Ascenzo M, Millard T, Freeberg L, Aoyama N, Kitzman J, Burgess D, Richmond T, Albert TJ, Barbazuk WB, Jeddeloh JA, Schnable PS (2010) Repeat subtraction-mediated sequence capture from a complex genome. Plant J 62:898–909

Glenn TC (2011) Field guide to next-generation DNA sequencers. Mol Ecol Resour 11:759–769

Gore M, Chia J, Elshire R, Sun Q, Ersoz E, Hurwitz B, Peiffer J, McMullen M, Grills G, Ross-Ibarra J (2009) A first-generation haplotype map of maize. Science 326:1115–1117

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

Gupta M, Chyi YS, Romeroseverson J, Owen JL (1994) Amplification of DNA markers from evolutionarily diverse genomes using single primers of simple-sequence repeats. Theor Appl Genet 89:998–1006

Gupta PK, Roy JK, Prasad M (2001) Single nucleotide polymorphisms: a new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. Curr Sci 80:524–535

Gupta PK, Rustgi S, Sharma S, Singh R, Kumar N, Balyan HS (2003) Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. Mol Genet Genomics 270:315–323

Gupta PK, Rustgi S, Mir RR (2013) Array-based high-throughput DNA markers and genotyping platforms for cereal genetics and genomics. In: Gupta PK, Varshney RK (eds) Cereal genomics II. Springer, Berlin

Haseneyer G, Schmutzer T, Seidel M, Zhou R, Mascher M, Schon C-C, Taudien S, Scholz U, Stein N, Mayer K, Bauer E (2011) From RNA-seq to large-scale genotyping—genomics resources for rye (Secale cereale L.). BMC Plant Biol 11:131

Huang X, Feng Q, Qian Q, Zhao Q, Wang L, Wang A, Guan J, Fan D, Weng Q, Huang T, Dong G, Sang T, Han B (2009) High-throughput genotyping by whole-genome resequencing. Genome Res 19:1068–1076

Huang XH, Wei XH, Sang T, Zhao QA, Feng Q, Zhao Y, Li CY, Zhu CR, Lu TT, Zhang ZW, Li M, Fan DL, Guo YL, Wang A, Wang L, Deng LW, Li WJ, Lu YQ, Weng QJ, Liu KY, Huang T, Zhou TY, Jing YF, Li W, Lin Z, Buckler ES, Qian QA, Zhang QF, Li JY, Han B (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. Nat Genet 42:U961–U976

Hyten D, Cannon S, Song Q, Weeks N, Fickus E, Shoemaker R, Specht J, Farmer A, May G, Cregan P (2010a) High-throughput SNP discovery through deep resequencing of a reduced representation library to anchor and orient scaffolds in the soybean whole genome sequence. BMC Genomics 11:38

Hyten D, Song Q, Fickus E, Quigley C, Lim J, Choi I, Hwang E, Pastor-Corrales M, Cregan P (2010b) High-throughput SNP discovery and assay development in common bean. BMC Genomics 11:475

Imelfort M, Edwards D (2009) De novo sequencing of plant genomes using second-generation technologies. Briefings Bioinf 10:609–618

Imelfort M, Batley J, Grimmond S, Edwards D (2009a) Genome sequencing approaches and successes. In: Somers D, Langridge P, Gustafson J (eds) Plant genomics. Humana Press, New York, pp 345–358

Imelfort M, Duran C, Batley J, Edwards D (2009b) Discovering genetic polymorphisms in next-generation sequencing data. Plant Biotechnol J 7:312–317

Jewell E, Robinson A, Savage D, Erwin T, Love CG, Lim GAC, Li X, Batley J, Spangenberg GC, Edwards D (2006) SSR Primer and SSR Taxonomy Tree: biome SSR discovery. Nucleic Acids Res 34:W656–W659

Kashi Y, King D, Soller M (1997) Simple sequence repeats as a source of quantitative genetic variation. Trends Genet 13:74–78

Katti MV, Ranjekar PK, Gupta VS (2001) Differential distribution of simple sequence repeats in eukaryotic genome sequences. Mol Biol Evol 18:1161–1167

Kota R, Rudd S, Facius A, Kolesov G, Thiel T, Zhang H, Stein N, Mayer K, Graner A (2003) Snipping polymorphisms from large EST collections in barley (Hordeum vulgareL.). Mol Genet Genomics 270:24–33

Kumar A, Hirochika H (2001) Applications of retrotransposons as genetic tools in plant biology. Trends Plant Sci 6:127–134

Lai JS, Li RQ, Xu X, Jin WW, Xu ML, Zhao HN, Xiang ZK, Song WB, Ying K, Zhang M, Jiao YP, Ni PX, Zhang JG, Li D, Guo XS, Ye KX, Jian M, Wang B, Zheng HS, Liang HQ, Zhang XQ, Wang SC, Chen SJ, Li JS, Fu Y, Springer NM, Yang HM, Wang JA, Dai JR, Schnable PS, Wang J (2010) Genome-wide patterns of genetic variation among elite maize inbred lines. Nat Genet 42:U1027–U1158

Lai K, Berkman PJ, Lorenc MT, Duran C, Smits L, Manoli S, Stiller J, Edwards D (2012a) WheatGenome.info: an integrated database and portal for wheat genome information. Plant Cell Physiol 53:1–7

Lai K, Duran C, Berkman PJ, Lorenc MT, Stiller J, Manoli S, Hayden MJ, Forrest KL, Fleury D, Baumann U, Zander M, Mason AS, Batley J, Edwards D (2012b) Single nucleotide polymorphism discovery from wheat next-generation sequence data. Plant Biotechnol J 10:743–749

Lai K, Lorenc MT, Edwards D (2012c) Genomic databases for crop improvement. Agronomy 2:62–73

Landjeva S, Korzun V, Borner A (2007) Molecular markers: actual and potential contributions to wheat genome characterization and breeding. Euphytica 156:271–296

Lee H, Lai K, Lorenc MT, Imelfort M, Duran C, Edwards D (2012) Bioinformatics tools and databases for analysis of next generation sequence data. Briefings Funct Genom 2:12–24

Liu L, Li Y, Li S, Hu N, He Y, Pong R, Lin D, Lu L, Law M (2012) Comparison of next-generation sequencing systems. J Biomed Biotechnol 2012:11

Lorenc MT, Hayashi S, Stiller J, Lee H, Manoli S, Ruperao P, Visendi P, Berkman PJ, Lai K, Batley J, Edwards D (2012) Discovery of single nucleotide polymorphisms in complex genomes using SGSautoSNP. Biology 1:370–382

Love CG, Batley J, Edwards D (2003) Applied computational tools for crop genome research. J Plant Biotechnol 5:193–195

Mammadov J, Chen W, Ren R, Pai R, Marchione W, Yalçin F, Witsenboer H, Greene T, Thompson S, Kumpatla S (2010) Development of highly polymorphic SNP markers from the complexity reduced portion of maize [Zea mays L.] genome for use in marker-assisted breeding. Theor Appl Genet 121:577–588

Marshall D, Hayward A, Eales D, Imelfort M, Stiller J, Berkman P, Clark T, McKenzie M, Lai K, Duran C, Batley J, Edwards D (2010) Targeted identification of genomic regions using TAGdb. Plant Methods 6:19

McCouch SR, Zhao K, Wright M, Tung C-W, Ebana K, Thomson M, Reynolds A, Wang D, DeClerck G, Ali ML, McClung A, Eizenga G, Bustamante C (2010) Development of genome-wide SNP assays for rice. Breed Sci 60:524–535

McNally KL, Childs KL, Bohnert R, Davidson RM, Zhao K, Ulat VJ, Zeller G, Clark RM, Hoen DR, Bureau TE, Stokowski R, Ballinger DG, Frazer KA, Cox DR, Padhukasahasram B, Bustamante CD, Weigel D, Mackill DJ, Bruskiewich RM, Rätsch G, Buell CR, Leung H, Leach JE (2009) Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. Proc National Acad Sci 106(30):12273–12278

Metzker ML (2010) Applications of next-generation sequencing, Sequencing technologies—the next generation. Nat Rev Genet 11:31–46

Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA (2007) Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome Res 17:240–248

Mogg R, Batley J, Hanley S, Edwards D, O'Sullivan H, Edwards KJ (2002) Characterization of the flanking regions of Zea mays microsatellites reveals a large number of useful sequence polymorphisms. Theor Appl Genet 105:532–543

Mortimer J, Batley J, Love C, Logan E, Edwards D (2005) Simple Sequence Repeat (SSR) and GC distribution in the Arabidopsis thaliana genome. J Plant Biotechnol 7:17–25

Moxon ER, Wills C (1999) DNA microsatellites: agents of evolution? Sci Am 280:94–99

Nie X, Li B, Wang L, Liu P, Biradar SS, Li T, Dolezel J, Edwards D, Luo MC, Weining S (2012) Development of chromosome-arm-specific microsatellite markers in Triticum aestivum (Poaceae) using NGS technology. Am J Bot 99:e369–e371

Paux E, Faure S, Choulet F, Roger D, Gauthier V, Martinant J, Sourdille P, Balfourier F, Le Paslier M, Chauveau A (2010) Insertion site-based polymorphism markers open new perspectives for genome saturation and marker-assisted selection in wheat. Plant Biotechnol J 8:196–210

Pennisi E (2012) Search for pore-fection. Science 336:534–537

Poland JA, Brown PJ, Sorrells ME, Jannink J-L (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS ONE 7:e32253

Powell W, Machray GC, Provan J (1996) Polymorphism revealed by simple sequence repeats. Trends Plant Sci 1:215–222

Rafalski A (2002) Applications of single nucleotide polymorphisms in crop genetics. Curr Opin Plant Biol 5:94–100

Ravel C, Martre P, Romeuf I, Dardevet M, El-Malki R, Bordes J, Duchateau N, Brunel D, Balfourier F, Charmet G (2009) Nucleotide polymorphism in the wheat transcriptional activator spa influences its pattern of expression and has pleiotropic effects on grain protein composition, dough viscoelasticity, and grain hardness. Plant Physiol 151:2133–2144

Robinson AJ, Love CG, Batley J, Barker G, Edwards D (2004) Simple sequence repeat marker loci discovery using SSR primer. Bioinformatics 20:1475–1476

Rustgi S, Bandopadhyay R, Balyan HS, Gupta PK (2009) EST-SNPs in bread wheat: discovery, validation, genotyping and haplotype structure. Czech J Genet Plant Breed 45:106–116

Saintenac C, Jiang D, Akhunov E (2011) Targeted analysis of nucleotide and copy number variation by exon capture in allotetraploid wheat genome. Genome Biol 12:R88

Santana QC, Coetzee MPA, Steenkamp ET, Mlonyeni OX, Hammond GNA, Wingfield MJ, Wingfield BD (2009) Microsatellite discovery by deep sequencing of enriched genomic libraries. Biotechniques 46:217–223

Schulman AH, Flavell AJ, Ellis THN (2004) The application of LTR retrotransposons as molecular markers in plants. Mob Genet Elem: Protoc Genomic Appl 260:145–173

Sharma PC, Grover A, Kahl G (2007) Mining microsatellites in eukaryotic genomes. Trends Biotechnol 25:490–498

Shen Y-J, Jiang H, Jin J-P, Zhang Z-B, Xi B, He Y–Y, Wang G, Wang C, Qian L, Li X, Yu Q-B, Liu H-J, Chen D-H, Gao J-H, Huang H, Shi T-L, Yang Z-N (2004) Development of genome-wide DNA polymorphism database for map-based cloning of rice genes. Plant Physiol 135:1198–1205

Shendure J, Ji HL (2008) Next-generation DNA sequencing. Nat Biotechnol 26:1135–1145

Subramanian S, Mishra RK, Singh L (2003) Genome-wide analysis of microsatellite repeats in humans: their abundance and density in specific genomic regions. Genome Bio 4(2):R13

Tautz D, Schlotterer C (1994) Concerted evolution, molecular drive and natural-selection—reply. Current Biol 4:1165–1166

Toth G, Gaspari Z, Jurka J (2000) Microsatellites in different eukaryotic genomes: Survey and analysis. Genome Res 10:967–981

Trebbi D, Maccaferri M, de Heer P, Sørensen A, Giuliani S, Salvi S, Sanguineti M, Massi A, van der Vossen E, Tuberosa R (2011) High-throughput SNP discovery and genotyping in durum wheat (Triticum durum). Theor Appl Genet 123:555–569

Trick M, Adamski N, Mugford S, Jiang C-C, Febrer M, Uauy C (2012) Combining SNP discovery from next-generation sequencing data with bulked segregant analysis (BSA) to fine-map genes in polyploid wheat. BMC Plant Biol 12:14

van Orsouw NJ, Hogers RCJ, Janssen A, Yalcin F, Snoeijers S, Verstege E, Schneiders H, van der Poel H, van Oeveren J, Verstegen H, van Eijk MJT (2007) Complexity reduction of polymorphic sequences (crops™): a novel approach for large-scale polymorphism discovery in complex genomes. PLoS ONE 2:e1172

Varshney RK, Sigmund R, Börner A, Korzun V, Stein N, Sorrells ME, Langridge P, Graner A (2005) Interspecific transferability and comparative mapping of barley EST-SSR markers in wheat, rye and rice. Plant Sci 168:195–202

Winfield MO, Wilkinson PA, Allen AM, Barker GLA, Coghill JA, Burridge A, Hall A, Brenchley RC, D'Amore R, Hall N, Bevan MW, Richmond T, Gerhardt DJ, Jeddeloh JA, Edwards KJ (2012) Targeted re-sequencing of the allohexaploid wheat exome. Plant Biotechnol J 10:733–742

Wing R, Kim H, Foicoechea J, Yu Y, Kudrna D, Zuccolo A, Ammiraju J, Luo M, Nelson W, Ma J (2007) The oryza map alignment project (omap): a new re-source for comparative genome studies within oryza. In: Upadhyaya NM (ed) Rice functional genomics. Springer, New York, pp 395–409

Xu X, Liu X, Ge S, Jensen JD, Hu F, Li X, Dong Y, Gutenkunst RN, Fang L, Huang L, Li J, He W, Zhang G, Zheng X, Zhang F, Li Y, Yu C, Kristiansen K, Zhang X, Wang J, Wright M, McCouch S, Nielsen R, Wang J, Wang W (2012) Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. Nat Biotech 30:105–111

Yamamoto T, Nagasaki H, Yonemaru J-i, Ebana K, Nakajima M, Shibaya T, Yano M (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

You F, Huo N, Deal K, Gu Y, Luo M-C, McGuire P, Dvorak J, Anderson O (2011) Annotation-based genome-wide SNP discovery in the large and complex Aegilops tauschii genome using next-generation sequencing without a reference genome sequence. BMC Genomics 12:59

Chapter 4
Application of Next-Generation Sequencing Technologies for Genetic Diversity Analysis in Cereals

Seifollah Kiani, Alina Akhunova and Eduard Akhunov

4.1 Introduction

A genome-wide study of molecular variation is required for high-resolution genetic analysis of complex phenotypic traits (Tian et al. 2011; Huang et al. 2010; Yu and Buckler 2006). A plethora of experimental methods for detecting various types of molecular variation has been developed (Akbari et al. 2006; Akhunov et al. 2009; Chao et al. 2010; Rostoks et al. 2006; Hyten et al. 2008). While all these methods differ in their scalability, throughput and cost, the most comprehensive and reliable approach for variant analysis still relies on direct DNA sequencing, which until recently, due to the high cost of the classical Sanger sequencing method, could not be applied to large populations. The assessment of genetic diversity directly from DNA sequence data (1) overcomes problems associated with ascertainment bias resulting from non-random sampling of individuals for variant discovery (Clark et al. 2005) and (2) allows for simultaneous analysis of different types of molecular variation including single nucleotide polymorphism (SNP), copy number and presence/absence variation (CNV/PAV) and small and large scale insertions and deletions (indels). The recent advances in NGS technologies have made it feasible to scale up analyses of genetic variation to the whole genome level (Huang et al. 2009; Lai et al. 2010; Elshire et al. 2011; Chia et al. 2012; Hufford et al. 2012).

One of the major challenges for the analysis of genetic variation in cereal genomes is their complexity, which is especially prominent in the large genomes of maize, barley and wheat (Schnable et al. 2009; Akhunov et al. 2007; Mayer et al.

S. Kiani · E. Akhunov (✉)
Department of Plant Pathology, Kansas State University, Manhattan, KS 66502, USA
e-mail: eakhunov@ksu.edu

A. Akhunova
Integrated Genomics Facility, Kansas State University, Manhattan, KS 66502, USA

P. K. Gupta and R. K. Varshney (eds.), *Cereal Genomics II*,
DOI: 10.1007/978-94-007-6401-9_4,

2011; Wicker et al. 2011; Salse et al. 2008; Choulet et al. 2010). Analysis of genetic variation in these cereal genomes is complicated by high repetitive DNA content and a high proportion of duplicated genes resulting from ancient and recent segmental and/or whole-genome duplications (Salse et al. 2008; Paterson et al. 2010). However, the confounding effect of these factors on variant analysis can be reduced if complete genome sequence data is available. In recent years the complete genome sequencing of several important cereal crops such as maize (Schnable et al. 2009), rice (International Rice Genome Sequencing Project 2005), sorghum (Paterson et al. 2009) and barley (International Barley Genome Sequencing Consortium 2012) brought forth reference sequences that were shown to be critical for fast analysis and precise mapping of newly discovered polymorphisms (Huang et al. 2009; Lai et al. 2010; Elshire et al. 2011). In the near future, an effort led by the International Wheat Genome Sequencing Consortium www.wheatgenome.org will create a resource for genome-scale diversity analyses even in this complex genome.

4.2 Next-Generation Sequencing Technologies

A wide range of NGS platforms are available on the market each differing in the cost of sequencing, throughput, methods of template preparation, principles used for DNA sequencing, and the length and number of reads generated per instrument run (Metzker 2007). A unique combination of these features makes each NGS platform suitable for various types of applications ranging from de novo genome sequencing to metagenomics to whole genome or targeted re-sequencing of thousands individuals in a population. A comparison of some of the characteristics of major commercial NGS platforms is provided in Table 4.1.

According to the length of reads, NGS platforms can be roughly grouped into those that produce fewer long reads and those that produce more short reads. There are currently two major long read sequencing technologies (Roche GS FLX+ and PacBio) with Ion Torrent moving toward achieving the capacities of Roche GS FLX+. The Roche GS FLX+ sequencing system uses a pyrosequencing approach based on incorporation of dNTPs by DNA polymerase followed by detection of released inorganic pyrophosphate by converting it into light signals (Margulies et al. 2005). The method does not use the termination of DNA synthesis in sequencing and, therefore, the light intensity is directly proportional to the number of nucleotides incorporated into synthesized DNA. The major error type produced by this system are insertions that occur due to inability of the system to correctly relate the intensity of the light signal to the number of incorporated bases during sequencing of homopolymeric stretches of DNA. While the cost of pyrosequencing is significantly higher than the cost of other sequencing technologies, the long read length offered by GS FLX system was shown to be useful for assembling highly repetitive regions of genomes or for resolving haplotypes in metagenomics samples (Metzker 2007).

Another long read NGS technology that uses single-molecule sequencing for reading DNA is developed by Pacific Biosciences (www.pacificbiosciences.com).

Table 4.1 Comparison of next-generation sequencing platforms

Sequencing platform	Read length, bp	Amount of data/run[a]	Million reads/run	Run time	Types of libraries[b]	Error rate (%)[c]	Type of errors
Roche/GS FLX+	700	700 Mb	1	23 h	MP, SR	1	Indel
Illumina, HiSeq2000	2 × 100 bp	540–600 Gb	1,000	11 days	MP, SR	≥0.01	Subst.
Pacific Biosciences	1,000	50 Mb	0.05	0.5–2 h	SR	16	CG del.
Ion Torrent (318 chip)	400	1 Gb	8	2 h	MP, SR	1	Indel
SOLiD—5500xl (4hq)	75 + 35	155 Gb	1,410	8 days	MP, SR	>0.01	A-T bias

[a]Maximum amount of data generated by a platform; [b] *MP* mate-pair libraries, *SR* single read; [c] percentage of errors per single read

The technology is based on real-time detection of fluorescently labeled dNTPs as they get incorporated by DNA polymerase into the synthesized strand of DNA. The reactions are performed in small chambers called zero-mode waveguide (ZMW); each chamber contains a single molecule of DNA polymerase attached to the bottom glass surface. The incorporation of labeled nucleotides is monitored by using a laser that excites the fluorescent labels by penetrating up from the holes. The diameter of the holes is specifically selected to prevent the laser from illuminating over 30 nm from the bottom of the ZMW, thereby increasing the signal-to-noise ratio for reliable base detection. The technology is currently capable of generating reads with the average length of 1,000 bp (Schadt et al. 2010), with the possibility of generating reads in excess of 10,000 bp. As with any single-molecule sequencing technology, however, the error rate per read is very high (Table 4.1) but may be overcome by increasing the data coverage.

The recently developed Ion Torrent platform, the Ion PGM sequencer, that can also deliver long reads based on semiconductor technology. This instrument is capable of sequencing DNA by directly sensing ions produced by template-directed DNA synthesis (Rothberg et al. 2011). The system uses ion-sensitive chips to perform massively parallel sequencing in up to 11 million wells, generating reads up to 400 bp with total output of 1–2 Gb.

The two major short read sequencing platforms generating reads ranging from 35 to 150 bp are Illumina's GAII and Life Science's SOLiD. While these technologies are based on different principles, they are capable of producing large number of short reads that are optimal for re-sequencing projects and analysis of gene expression. The HiSeq2000 is one of the latest versions of Illumina NGS platforms capable of producing 1 billion reads per run generating up to 600 Gb of sequence data. Currently, this is the most broadly used NGS platform utilized in a wide range of applications including expression analysis, de novo genome assembly and genome-wide re-sequencing. The HiSeq2000 uses bridge PCR occurring on the surface of a glass slide to amplify individual DNA molecules captured by oligonucleotides attached to the surface of the slide, thus generating small clusters of identical molecules. The sequencing is performed using a method similar to the Sanger sequencing except that Illumina uses reversible terminators which can be used for further DNA synthesis after cleavage of the fluorescent dye. The SOLiD platform is based on ligation of short, fluorescently labeled oligonucleotides to determine the sequence of a DNA template. Preparation of sequencing libraries for SOLiD is similar to that used for GS FLX platform.

4.3 Detection of DNA Sequence Variation in Next-Generation Sequence Data

Variant detection in next-generation sequence data includes multiple steps and is usually performed by aligning NGS reads to reference sequences, which could be represented by complete genome sequences, shotgun genome assemblies or EST contigs. Next the alignment is scanned for variable sites. While using

NGS platforms for the analysis of genetic variation, several factors needs to be taken into consideration: sequencing errors, errors in the assembly and missing data. Some of these factors are platform specific or depend on sequence coverage (sequencing errors), while some are the result of assembly algorithms (errors in the assembly) or an inherent feature of next-generation shotgun sequence data (missing data). The latter factor plays an important role in experiments relying on low coverage sequencing (<5× per diploid genome) where there is a high probability of sampling only one of the alleles at a variable site. Although the accuracy of genotype calling may be improved by increasing the depth of sequence coverage, the demand for sequencing an even larger numbers of individuals suggests that low to medium depth of sequence coverage will be the most common type of data generated in future NGS experiments. This design is not only cost- effective, but has also increased power to detect low-frequency variants in sequenced populations of large size (Nielsen et al. 2011). Therefore, in designing the NGS experiments, one should consider the selection of a sequencing platform, the amount of NGS data generated per individual and appropriate bioinformatical and statistical approaches for variant discovery and genotype calling.

There are a number of excellent bioinformatical tools, both commercial and publicly available, for mapping NGS reads to the reference genome (Table 4.2). These programs differ in the algorithms used for read alignment, in their ability to process NGS data from different sequencing platforms and also in their ability to detect different types of genetic variation (SNPs, CNV, PAV, or indels). The most commonly used aligners use alignment algorithms based on 'hashing' or data compression referred to as 'Burrows-Wheeler transform' (BWT) (Burrows and Wheeler 1994). The BWT-algorithm is fast and computationally efficient. It is implemented in, for example, such popular programs as Bowtie (Langmead et al. 2009), SOAP (Li et al. 2009a, b) and BWA (Li and Durbin 2009). Even more sensitive, but more computationally intensive hash-based algorithms capable of generating alignments for the most accurate genotype calls, are implemented in MAQ

Table 4.2 List of non-commercial NGS alignment software

Alignment software	Alignment algorithm[a]	Long read mapping[b]	Use PE[c]	Use Q[d]	Gapped alignment
SOAP2	BW	Yes	Yes	Yes	Yes
BWA	BW	Yes	Yes	No	Yes
Bowtie	BW	No	Yes	Yes	No
MAQ	Hashing reads	No	Yes	Yes	Yes
Novoalign	Hashing reference	No	Yes	Yes	Yes
Mosaik	Hashing reference	Yes	Yes	No	Yes

[a]Burrows-Wheeler indexing (BW); more complete list of alignment software can be found at http://seqanswers.com/wiki/Software/list
[b]Sanger and 454 reads can be aligned
[c]Can map paired-end reads
[d]Uses base quality in alignment

(maq.sourceforge.net), Novoalign (www.novocraft.com) and Stampy (Lunter and Goodson 2011).

The genotype calling aimed at determining the genotypes of each individual in the sample starts with SNP calling step (or variant calling). During this step, variable sites in the alignment where aligned sequences are different from the reference are identified. The early methods of genotype and SNP calling were based on a fixed cutoff value for allele or genotype counts (Table 4.3). The examples of programs using this SNP and genotype calling approach are Roche's GS Mapper (Roche) or CLC Genomic Workbench. However, the allele counting approach is less suitable for low coverage sequence data since it can underestimate heterozygous genotypes and does not provide estimates of genotype quality (Nielsen et al. 2011). The disadvantages of these allele counting methods were corrected in recently developed probabilistic methods of genotype calling that incorporate various sources of errors (sequence error, alignment errors, etc.) into the quality score that reflects the uncertainty associated with each inferred genotype (Li et al. 2008, 2009). Some of these new algorithms also incorporate information about the allele frequency and the estimates of linkage disequilibrium (LD) into a probabilistic framework for improving genotype calling accuracy (Durbin et al. 2010).

Table 4.3 List of non-commercial genotype calling software for NGS data

Software[a]	Calling method	Input data	Comments
SOAP2 (SOAPsnp)	Single sample	Aligned reads	Package for NGS data analysis; including SOAPsnp module for SNP calling
SAMtools	Multiple samples	Aligned reads	Package for NGS data analysis; samtools and bcftools modules are used for computation of genotype likelihoods and SNP and genotype calling
GATK	Multiple samples	Aligned reads	Package for NGS data analysis; Unified Genotyper is used for SNP and genotype calling; Variant Filtration is used for SNP filtering; Variant Recalibrator is used for variant quality re-calibration
IMPUTE2	Multiple sample LD	Genotype likelihoods, fine-scale linkage map	Imputation and phasing software with option for genotype calling
MaCH	Multiple sample LD	Genotype likelihoods provided in map order	MACH is a Markov Chain based haplotyper that can resolve long haplotypes or infer missing genotypes in samples of unrelated individuals

[a]More complete list of alignment software can be found at http://seqanswers.com/wiki/Software/list

Even though the methods of genotype calling based on probabilistic approaches provide posterior probabilities in each site, if not all sources of error are taken into account, additional post-processing data filtering may be required to improve accuracy. Such filters are usually data-specific and depending on experimental design can be based on low-quality scores, abnormal LD, extreme read depth, strand bias, deviation of SNP calling data from previously performed SNP genotyping of the same individuals etc. (Nielsen et al. 2011; Durbin et al. 2010). The list of freely available programs for SNP and genotype calling can be found in Table 4.3.

4.4 Genome-Level Analysis of Genetic Variation

4.4.1 Transcriptome-Based Analysis

RNA-Seq is a powerful sequencing-based approach that enables researchers to survey the entire transcriptome in a high-throughput manner utilizing NGS technologies. By providing single-base resolution for annotation, and having the ability to generate an enormous number of reads in one sequencing run, the RNA-seq analysis can be both qualitative and quantitative. The RNA-seq approach has been broadly used for the analysis of genetic variation in the transcribed portion of cereal genomes (Barbazuk et al. 2007; Oliver et al. 2011; Allen et al. 2011). Analysis of genetic variation in RNA samples is simplified due to the reduction of sample complexity, while at the same time still allowing for variant discovery in functionally relevant parts of genome. Overall, transcriptome-based analysis of genetic variation studies demonstrated that, with the application of appropriate bioinformatical procedures, it is possible to discover SNPs even in complex polyploid genomes with a high level of confidence.

The successful SNP discovery efforts in cereal genomes were performed by sequencing cDNA libraries using 454 or Illumina sequencing platforms (Barbazuk et al. 2007; Oliver et al. 2011; Allen et al. 2011; Eveland et al. 2008). In these studies, of special interest are the computational approaches used to deal with the absence of a reference genome and/or polyploidy. The study performed by Barbazuk et al. (2007) used the early version of the 454 sequencing platform to generate about 0.5 million 454 reads with an average length of 100 bp. These reads were aligned using CROSS_MATCH to the maize assembled genomic islands (MAGIs) composed of gene-enriched fraction of the maize genome (Fu et al. 2005). The usage of genomic sequences for alignment helps to avoid problems associated with alternative splicing. The SNP discovery was performed using POLYBAYES (Marth et al. 1999) with paralog filter activated to reduce SNP calling errors due to the alignment of 454 reads to paralogs. Although POLYBAYES also identifies indel polymorphisms, to avoid variant calling errors due to high incidence of indel errors (Margulies et al. 2005; Table 4.1) in 454 sequencing data, only single-base substitutions were recorded. Application of a stringent filtering criterion ($\geq$2 reads/SNP allele) to this dataset revealed 7,016 SNPs with an overall 71 % validation rate.

In another study in maize, Eveland et al. (2008) presented a new 454 sequencing strategy for expression profiling to capture the information-rich 3′-UTR of mRNA. In this study, the analysis of approximately 229,000 3′-anchored sequences from maize ovaries identified 14,822 unique transcripts. Based on the analysis of SNPs identified within B73 MAGI genomic assemblies, they confirmed at least 89.9 % of polymorphisms independently by identical cDNA matches. These data were consistent with a study of Barbazuk et al. (2007) in which 88 % of SNPs sampled by two or more 454 reads were validated by Sanger sequencing. In addition, they showed that homopolymer base-calling errors (Margulies et al. 2005) have a minor impact on the ability to detect polymorphisms in maize cDNAs. A similar pyrosequencing-based SNP discovery approach was applied to the oat genome as well, where EST contigs were assembled using newly generated 454 sequencing data and used as a reference for mapping 454 reads. In spite of the low coverage data generated per genotype (about 250,000 reads) and the complication of polyploidy, the variant discovery performed in this dataset using the commercial gsMapper program (Oliver et al. 2011) resulted in 9,448 SNPs. Furthermore, a total of 48 SNPs of 96 (50 %) were positively validated by high-resolution melt analysis.

Alternative sequencing technology and bioinformatical approach were used for the discovery of SNPs in the wheat genome (Allen et al. 2011). The study was designed to specifically discover SNPs in the polyploid genome and reduce the effect of homoelogous mutations on false SNP calling rate. In this study, the authors used the Illumina GAII platform to generate paired-end 75 bp reads from the normalized transcriptomes of five wheat varieties. The reference sequences for read mapping were prepared by assembling publicly available EST data and Illumina reads. The mapping of NGS reads from five varieties was performed using the ELAND program (Illumina Inc.). Alignments were processed using custom Perl scripts designed to identify inter-varietal SNPs and remove homeologous SNPs usually present as intra-varietal polymorphisms. Only those variable sites that contained more than 2 reads per SNP allele with phred score $\geq$20 in a 3-base window were considered polymorphic, yielding an average of five varietal SNPs per kilobase across the five varieties sequenced. This estimate is consistent with the previous estimate of 4.3 varietal SNPs per kilobase in wheat (Barker and Edwards 2009). The genotyping experiment performed in the study by Allen et al. (2011) included 1,659 SNPs and demonstrated that 71 % of SNPs could be converted into working genotyping assays based on KASPar technology (KBioscience, UK). By screening monomorphic SNPs in the panel of 28 Chinese Spring nullisomic lines (Kimber and Sears 1979), it was demonstrated that the majority of them (93 %) represent homoelogous SNPs differentiating wheat genomes from each other. The KASPar genotyping assay conversion rate (71 %) was only slightly lower than that reported for a GoldenGate genotyping assay (76 %, Akhunov et al. 2009) developed using SNPs discovered in Sanger resequencing data (Akhunov et al. 2010).

In another study, single-base resolution analysis of the rice transcriptome was performed by mapping 40- and 76-bp paired-end RNA-seq reads generated with

Illumina GAII platform to the rice reference genome (Lu et al. 2010). RNA-seq comparison of the cultivated rice *Oryza sativa indica* and *japonica* was used to indentify transcriptionally active genomic regions and differentially expressed genes, characterize the patterns of alternative splicing and estimate the number of SNPs differentiating these two rice subspecies from each other. This study took advantage of available reference genome to map short reads to genomic sequence using SSAHA2 software (www.sanger.ac.uk/resources/software/ssaha2). For SNP detection, only those sites that showed the quality score ≥20 and read depth ≥5 were used resulting in the discovery of ~64,000–67,000 SNPs between *indica* and *japonica* subspecies. Of these SNPs only half were found within the annotated gene models, with 60.8 % of SNPs being located in the coding regions and 29.1 and 10.1 % of SNPs found in the 3′- and 5′-UTRs, respectively. The ratio of SNPs within nonsynonymous and synonymous substitution sites was nearly 1:1.06 (Lu et al. 2010).

4.4.2 Targeted Sequence Capture

The size of some of the cereal genomes (wheat, barley, maize, oat etc.) makes it impractical to analyze the large number of samples by direct whole-genome sequencing, even considering the throughput of modern NGS instruments. Therefore, the reduction of the input DNA sample complexity by enriching it with the genomic targets of interest is a cost-effective approach for studying genetic diversity in a large number of samples. The possible applications of this approach are genome-wide association mapping experiments, population genetics studies or re-sequencing of mutant populations for cataloging mutations in coding parts of genome.

One of the more recently developed methods of sequence capture use long oligonucleotide probes for enrichment of shotgun genomic libraries with the sequences of interest (Albert et al. 2007; Porreca et al. 2007; Gnirke et al. 2009; Okou et al. 2007). The targets of interest are captured using the pool of oligonucleotide probes, also referred to as "baits", that are complementary to the targets followed by elution of captured DNA. These types of captures can be performed using either solid- or liquid-phase hybridization assays (Fu et al. 2010; Saintenac et al. 2011) both of which showed similar levels of efficiency and specificity (Teer et al. 2010). However, the advantage of the liquid-phase assay is the high level of multiplexing that can be achieved using liquid-handling robotics. Further reduction in the cost of sequencing and increase in the throughput can be achieved by using multiplexing adaptor sequences which are added during library preparation (Bansal et al. 2011). Eluted target DNA is then sequenced using NGS platforms, followed by mapping reads to reference sequence. The sequence capture methodologies have shown high reproducibility and target specificity and have been effectively used for the analysis of genetic diversity in the human genome.

Recently, sequence capture methods have been successfully been applied to characterize genetic diversity in maize (Fu et al. 2010), soybean (Haun et al. 2011) and wheat (Saintenac et al. 2011). In maize, array-based sequence capture

was used for discovering 2,500 high-quality SNPs between the reference accessions B73 and Mo17 in a 2.2 Mb region (Fu et al. 2010). In this study, the authors developed two-step enrichment approach. The first step includes hybridization with the Nimblegen array containing maize repetitive elements; this step depletes the genomic library of repetitive elements. The second step enriches library for the targets of interest. This capture approach achieved 1,800–3,000-fold target enrichment with up to 98 % of targets being represented in the final capture library. The authors used Mosaik aligner for mapping reads to the maize B73 reference using parameters favoring only unique alignments. SNP discovery was performed making use of the GigaBayes package (http://bioinformatics.bc.edu/marthlab/Software_Release). The estimated false SNP discovery rate was below 3 % even in the presence of captured paralogous sequences.

Liquid-phase sequence capture assay has been successfully tested for the analysis of genetic variation in the large polyploid wheat genome (Saintenac et al. 2011). Sequence capture in polyploid wheat was performed using a liquid-phase hybridization assay. A total of 55,000 120-mer RNA baits were designed to target 3.5 Mb of genic sequence selected from a non-redundant set of 3,497 genome-wide distributed full-length cDNAs. Each homoeologous set of genes in the capture assay was represented by only one full-length cDNA sequence. The genomic DNA of tetraploid wild emmer *T. dicoccoides* (Td) and cultivated durum wheat *T. durum* cv. Langdon (Ld) were captured and mapped to the cDNA reference sequences. Since short read NGS platforms like Illumina GAII are less suitable for reconstructing haplotypes of individual wheat genomes, Illumina reads from homoeologous or paralogous copies of genes were mapped to the same region of the reference sequence. The primary challenge for variant discovery in these complex alignments is distinguishing allelic variation between lines (henceforth, SNPs) from sequence divergence between the wheat genomes (henceforth, genome-specific sites or GSS) (Fig. 4.1).

SNP calling in this study was performed using the SAMtools software (Li et al. 2009) followed by filtering detected variable sites to reduce false positives. Filtering parameters were specifically selected for identifying GSS and SNP sites in the polyploid genome and were adjusted experimentally using multilocus Sanger re-sequencing data. The variant filtering was based on the overall depth of

Fig. 4.1 Possible types of variable sites in a tetraploid wheat genome alignment. At genome-specific sites (*GSS*), both nucleotide variants represent fixed divergent mutations that differentiate the diploid ancestors of the wheat A and B genomes. SNP sites originate due to a mutation in one of the wheat genomes (in this example, in the A genome of Ld) and, therefore, are polymorphic in population

coverage, the minimum coverage at variable sites and the ratio of variant coverage (Saintenac et al. 2011). The level of target enrichment achieved in the study (2,900-fold) was comparable to that obtained for maize using the two-step enrichment procedure (Fu et al. 2010), demonstrating that a liquid-phase sequence capture can be the efficient tool for the targeted re-sequencing of polyploid wheat genome. It was also shown that multiple libraries can be labeled using barcoded adaptors and pooled together before sequence capture allowing for an even higher level of throughput for large scale re-sequencing projects.

The wheat exonic sequence alignments were analyzed to catalogue GSSs, SNPs sites, CNVs and PAVs between two accessions of cultivated and wild tetraploid wheat. Variant analysis detected 14,499 GSSs and 3,487 SNPs with 1 and 15 % false positive rate. The factor that had a major impact on SNP calling false positive rate in the polyploid wheat genome was a failure to the recover second variant at a SNP site due to high level of targeted sequence divergence, which either resulted in low efficiency of sequence capture or precluded reads from being aligned to the reference. Overall, the level of sequence divergence between reference cDNAs used for designing capture baits and targeted sequences biased the efficiency of capture toward capturing sequences more similar to the reference than to their homoeologous counterparts. This result suggests that both conditions of capture hybridization and the design of capture baits for polyploid genomes will need to be further optimized for better performance.

However, in spite of this capture bias the study successfully proved that sequence capture assays designed using EST/cDNA sequences can be effectively used for analyzing genetic variation in polyploid genomes. The obtained results suggest that baits designed using only one of the homoeologous copies of a gene are capable of capturing diverged gene copies from both genomes in tetraploid wheat. Therefore, it should be possible to capture homoeologous copies of genes in polyploids using the reduced set of baits designed to target only “diploid gene complement”. The exon capture assay designed using wheat genomic resources can also be effective for targeted enrichment of exons from the genomes of species closely related to wheat, many of which represent valuable sources of genes relevant in agriculture.

4.4.3 Reduced Representation Sequencing

Restriction site-associated DNA (RAD) markers, which are short DNA fragments adjacent to each instance of a particular restriction enzyme recognition site, are another efficient marker system for the study of genetic diversity (Miller et al. 2007). In first generation of RAD marker development based on microarray technology, hybridization of RAD tags to DNA microarrays allowed the parallel screening of thousands of polymorphic markers to map natural variation and induced mutations in diverse organisms (Miller et al. 2007). More recently, however, RAD genotyping through next-generation sequencing has been developed and successfully used to map an induced mutation in *Neurospora crassa*

(Baird et al. 2008). The RAD method uses restriction enzymes as a complexity reduction strategy to limit the sequenced portion of the genome to regions adjacent to restriction sites. It also facilitates the creation of highly multiplexed NGS libraries, thereby reducing library preparation costs (Baird et al. 2008; Ganal et al. 2009). It is noteworthy, however, that in restriction site-associated marker systems the selection of restriction enzymes that leave 2–3 bp overhangs and do not cut frequently in the major repetitive fraction of the genome is of critical importance.

In cereals, RAD sequencing by NGS has been successfully used to simultaneously discover and genotype SNP markers in sorghum (Nelson et al. 2011) and barley (Chutimanitsakun et al. 2011). It has been used to discover and genotype SNPs for linkage map construction in an Oregon Wolfe Barley (OWB) mapping population (Chutimanitsakun et al. 2011). The genomic DNA of both parental lines and 93 individuals from DH population (~300 ng) were digested with *Sbf*I restriction enzyme and modified Solexa adapters containing sample-specific barcodes were ligated to each DNA before pooling and random shearing. The constructed OWB libraries were run on an Illumina GAII instrument using single read (1 × 36 bp) sequencing chemistry. A total of 2,010,583 36-bp sequence reads were obtained for the parents of the OWB mapping population, and 27,704,592 sequence reads were obtained for the 93 DH mapping population. For map construction, raw 36-bp Illumina reads were assigned to each sample using barcoded adaptors. Data from each individual was then collapsed into RAD sequence clusters, excluding the sequences with <8× and >500× coverage levels. Homologous RAD clusters from parental lines were compared using a custom k-mer matching algorithm permitting exact sequence matches (monomorphic loci), single mismatch (one SNP per read) and two nucleotide mismatches (two SNPs per read) per 28 bp sequence. Of the 10,000 RAD clusters interrogated between parental lines, 530 (5.3 %) polymorphic co-dominant SNPs were identified among which, 436 were used for the final map construction after scoring RAD sequences in the DH individuals and removing RADs with >15 % missing data. The high quality of the RAD data was further confirmed by the comparable linkage map lengths for the RAD only, RAD+ prior marker, and DArT OWB maps.

In sorghum, three complexity-reduced libraries (one semi-random "SR" library made using the *Hpa*II enzyme and two RAD libraries prepared using the *Pst*I and *Bsr*FI enzymes ("Floragenex, Inc. Eugene, Oregon, USA") have been constructed and sequenced using Illumina instrument. The objective was to discover SNPs and assess their distribution in a sample of 8 accessions, including the reference accession of sorghum, BTx623 (Nelson et al. 2011). Two methods were used to identify candidate SNPs: (1) SOAP2 (Li et al. 2009b) was used to align the reads to the reference genome sequence, allowing unique alignment with a maximum of two mismatches. In the second method, Novoalign (http://www.novocraft.com) was used for alignment and SAMtools (http://samtools.sourceforge.net/) was used to identify candidate SNPs and indels with the minimum coverage depth of 3 reads, minimum SNP quality of 20 and minimum indel quality of 50, with default parameters for window size and nearby SNP and gaps. A total of 247 million reads were obtained, yielding 6 Gb (6 genome equivalents) of data that was aligned to

one-third of the sorghum reference sequence. Two methods of analysis produced slightly different results: while, 237,000 SNPs passed the Novoalign SNP filter, only 155,000 SNPs were called from SOAP2 alignments based on a simple filter requiring ≥6 alternative calls with average alternative-allele base-quality score ≥20. However, the high SNP validation rate obtained using SOAP2 (79 %) and Novoalign (82 %) suggested that both programs produce reliable SNP calls.

Another approach, referred to as "genotyping-by-sequencing" (GBS), has been introduced more recently. In this method the complexity of the DNA sample, like in RAD, is reduced by digesting with restriction enzymes. The advantage of the GBS method over a RAD approach is a simplified library construction procedure that does not require random shearing of libraries and second adaptor ligation step (Elshire et al. 2011). This strategy showed promising results for genetic map construction in barley and maize recombinant inbred lines (RILs). Libraries for NGS have been constructed using the *Ape*KI, a type II 5 bp sequence methylation-sensitive restriction enzyme. The throughput of GBS was scaled up by multiplexing up to 96 samples per lane of Illumina GAII instrument. The filtered sequence reads were aligned to the Maize reference genome (B73 RefGen v1) using the BWA program, allowing for a maximum of four mismatches and one gap of up to 3 bp. The GBS reads (tags) were scored as presence/absence in IBM mapping population that contained already 644 genetically mapped SNPs. The test for co-segregation of SNPs with GBS tags resulted in mapping a minimum of 200,000 (25,185 biallelic and 167,494 co-dominant) sequence tags in the IBM mapping population and roughly 25,000 tags in barley. The position of 90.8 % of mapped tags in IBM population agreed with the positions in reference genome suggesting that the GBS approach can be successfully used for genome-wide analysis of genetic variation in species lacking developed molecular tools or as a cost-effective marker system for the analysis of genetic variation in breeding populations.

The complexity reduction approach relying on methylation sensitive and insensitive restriction enzymes was also used to re-construct the first haplotype map of maize (Gore et al. 2009). The three complementary restriction enzyme-anchored genomic libraries were used to re-sequence the low-copy fraction of the maize genome in a diverse panel of 27 inbred lines. These lines were founders of the maize nested association mapping (NAM) panel selected to represent the diversity of maize breeding efforts and world-wide diversity (McMullen et al. 2009). More than 1 billion Illumina reads (>32 Gb of sequence data) provided low coverage of ~38 % of the maize genome. The authors found that 39 % of the sequenced low-copy DNA fraction was derived from intron and exons, thereby covering up to 32 % of the total genic regions in maize. A total of 3.3 million SNPs and indels were detected, 41 % of which were due to paralogous sequences in the inbred lines and a result of ancestral duplications in maize genome (Schnable et al. 2009). In total, 7.8 % of reads were unique or unalignable in these inbred lines, and the B73 genome could only capture 70 % of the alignable low-copy fraction represented by 27 inbred lines. Two estimates of recombination rate ("*R*", estimated from crossover frequency in the NAM population and "ρ", historical recombination rate estimated from the genetic diversity data) were highly correlated,

suggesting that the recombination rate was stable over time. Highly suppressed recombination in pericentromeric regions was thought to perhaps affect the efficiency of selection in maize and proposed to be one of the components of heterosis. The study uncovered multiple regions in the maize genome with low diversity and an excess of low-frequency variants that may have been loci subjected to selection and probably are involved in geographic adaptation of maize. Notably, in the high recombination fraction of genome, 148 regions were found with diversity less than that of the domestication gene *tb1* (Wang et al. 1999), of which at least one region has been shown to be involved in domestication on chromosome 10 (Tian et al. 2009).

4.4.4 Whole Genome Re-sequencing

Resequencing for genome-wide surveys of genetic diversity first became reasonable with the increased throughput of NGS platforms (Wheeler et al. 2008). Although this approach is bioinformatically challenging, it provides an unbiased estimation of genetic diversity across the genome in both coding and non-coding regions. In addition, it allows for the detection of various types of genetic variation; this detection capability includes not only SNPs and small indels, but also large mega-base scale indels, PAVs and CNVs. The full potential of whole genome re-sequencing is realized only in species for which complete genome sequence data is available. The bioinformatical analysis of genome-scale sequencing data has been significantly simplified by the development of fast and efficient alignment algorithms for mapping millions of NGS reads to the reference genome and performing variant discovery (Tables 4.2, 4.3).

The first whole-genome sequencing studies were performed on model plant species with small genomes such as *Arabidopsis* (Ossowski et al. 2008) and later, with the increased throughput and cost-reduction of NGS platforms, whole-genome sequencing was applied to rice landraces and maize elite inbred lines with the purpose of map construction and genetic diversity assessment (Huang et al. 2009; Lai et al. 2010; Mammadov et al. 2010; Arai-Kichise et al. 2011).

In cereals, whole-genome re-sequencing using Illumina GA platform was used to construct a genetic map from a cross between two accessions of rice *O. sativa ssp. indica* cv. 93–11 and *ssp. japonica* cv. Nipponbare (Huang et al. 2009). In this study, the authors investigated the accuracy of genotype calling in re-sequencing data obtained for the population of RILs and its relationship to sequence coverage, SNP density, genome size and error rate in the reference sequence. The shallow sequencing coverage (0.02 × rice genome equivalents) was achieved in the study for each RIL, which in addition to a large amount of missing data, makes an individual SNP genotype calling an error prone process. To deal with these issues, the authors developed an approach that considers a group of 15 SNPs within a sliding window for defining local genotypes. Two parental lines along with 150

RILs have been sequenced using a multiplexed bar-coded sequencing strategy by pooling indexed DNA libraries of 16 RILs in each lane of Illumina GA. Analysis of parental lines resulted in detection of 1,226,791 SNPs (3.2 SNPs/Kb) whereas the analysis of high quality reads obtained for the populations of RILs resulted in detection of 1,493,461 SNPs (25 SNPs/Mb). Re-sequencing of parental genotypes showed a SNP error rate of 4.12 and 0.71 % for *indica* and *japonica* parents, respectively. Based on these error estimates 3.41 % more *japonica* SNPs than *indica* SNPs in the heterozygous region of RILs were expected. These SNP error rates were taken into account during the calculation of the probability of occurrence of each genotype for a given SNP ratio in the sliding window. After all genotypes were called and corrected for SNP errors, a total of 5,074 breakpoints were found for 150 RILs (33.8 per RIL).

As the authors have pointed out, SNP error rate is an important issue in genotyping by whole-genome sequencing methods that needs to be taken into account. Three sources of errors contribute to the final SNP error rate: (1) errors in the 3-base barcode used for sample multiplexing (estimated to be 0.3 %); (2) errors in 33-mer Illumina reads (estimated to be 0.03 %); and (3) errors in the reference genomes of *indica* and *japonica* (estimated to be 3.9 and 0.39 %, respectively). Taken together, the SNP error rate for homozygous *indica* and *japonica* genotypes was estimated to be 4.23 and 0.72 %, respectively, which was similar to the experimentally estimated error rate of 4.12 and 0.71 % for re-sequenced parental lines. Finally, the authors reported a genetic map with a recombinant breakpoint of 40 kb apart in average and total length of 1539.5 cM, which not only provided much finer resolution compared to the existing genetic map but also was 20× faster in data collection and 35× more precise in recombination breakpoint detection. The utility of newly developed sequencing-based map for high-resolution gene mapping and subsequent cloning was confirmed by the detection of semi-dwarf gene *sd1*, responsible for the rice green revolution (Sasaki et al. 2002), localized to a 100-kb region.

Deep whole-genome sequencing (45× coverage) was also performed on rice landrace *O. sativa* L. cv. Omachi (used for *Japanese* rice wine production), in order to discover new SNPs and assess diversity between closely related cooking and non-cooking rice cultivars from *japonica* group (Arai-Kichise et al. 2011). Eight paired-end lanes of sequencing performed on Illumina GAII platform generated about 298 million 75 bp reads, 77 % of which were uniquely mapped onto the Nipponbare genome (IRGSP Build 4) using BWA software (Li and Durbin 2009). Both SNPs and insertions/deletion polymorphisms were identified using SAMtools software followed by applications of several filtering criteria to reduce the false positive SNPs and indels: target depth of ≥5, minimum mapping quality of 30, polymorphism call rate ≥90 % for SNPs and 30 % for indels. This re-sequencing study identified 132,462 SNPs and 35,766 indels (16,448 insertions and 19,318 deletions) between two *japonica* rice cultivars. The authors also showed that the number of detected SNPs and indels increase linearly with the increase of coverage up to 22.3×, and then slow down. Their observation suggests that 45× genome coverage is sufficient to detect most polymorphisms between these cultivars.

A combination whole-genome NGS shotgun sequencing with aligning whole-genome shotgun and cDNA sequence data for discovering SNPs was reported for the diploid ancestor of the polyploid wheat D genome *Aegilops tauschii* (You et al. 2011). The combination of Illumina, SOLiD and 454 sequencing technologies was used to generate sequence data which was mapped to the reference sequence of *Ae. tauschii* created by *de novo* assembling of whole-genome shotgun sequence data. Read mapping was performed using the BWA program with default settings followed by SNP discovery with SAMtools package. A set of post-processing filters was applied to variant calling data (Table 2 in You et al. 2011). A total of 497,118 putative SNPs were discovered with about ~80 % of them being positively validated by Sanger re-sequencing.

Whole-genome re-sequencing at a moderate level of coverage (5.4×) was used to assess the level of SNPs, indels and PAVs among 6 elite inbred lines of maize (Zheng58, 5003, 478, 178, Chang7-2 and Mo17) including the parents of the most productive hybrids in China (Lai et al. 2010). This data was used to assess the extent of genetic differentiation between maize inbred lines from different heterotic groups and to test the hypotheses of heterosis. The authors generated 1.26 billion 75-bp paired-end reads (83.7 Gb) to obtain 5.4× average depth of coverage for each inbred line. The reads were aligned to the maize reference genome using SOAP software (Li et al. 2009a). Uniquely mapped reads in non-repetitive regions were used for SNP calling using the probabilistic algorithm implemented in the SOAPsnp program. A set of filters including SNP coverage and maximum likelihood of SNP calls was applied to obtain high-quality genotype calls resulting in the discovery of 1,272,134 SNPs in non-repeat regions out of which 469,966 SNPs were found in the 32,540 maize genes and 130,053 SNPs were mapped to the coding regions. A total of 30,178 indel polymorphisms ranging from 1 to 6 bp were identified among 6 maize lines with only 571 indels present in the coding regions. The high efficiency of the approach for variant discovery applied by the authors was confirmed by sequencing 92 randomly selected PCR products which in turn validated 95 % of the predicted SNPs.

The authors of the above study identified SNPs and indels with large-effects on coding sequence among which 1,478 SNPs induced a stop codon, 97 altered initiation methionine residue, 828 disrupted splicing donor or acceptor sites and 322 indels caused frameshift. In addition, 1,087 SNPs removed stop codons resulting in longer coding sequences. Interestingly, of the large-effect SNPs, 101 were located in 45 genes that encode a disease-resistance protein with LRR domain, consistent with findings in *A. thaliana* and rice (Clark et al. 2007; McNally et al. 2009).

In addition to highly polymorphic genes, the authors found 393 genes that contained no SNPs among sequenced lines. To identify all chromosomal regions with low diversity, the number of segregating nucleotides per site within 1-Mb sliding window was calculated across the genome, and 101 genomic blocks ranging from 2.4 to 13 Mb in size with reduced diversity were identified. Intriguingly, the region with low diversity included the genes such as *bt2* and *su1* that are known to have been under selection during maize domestication (Whitt et al. 2002). This result

suggests that some of the identified low diversity genomic regions may contain genes subjected to selection.

One of the most interesting aspects of this study was the analysis of PAVs including full-length functional genes. This was performed by mapping Mo17 resequencing data to the B73 reference genome resulting in identification of 104 regions in the B73 genome in which at least 80 % of a 5-kb or longer genomic region and at least 90 % of its annotated transcriptional region did not have corresponding Mo17 reads. Moreover, using the same criteria, 296 high-confidence genes in B73 were missing from at least one of the six inbred lines. Comparison of contigs assembled from non-mapped reads of six inbred lines identified genomic regions present in the inbred lines but absent in B73. Annotation of these contigs resulted in 570 putative genes with an average length of 527 bp (considering only coding regions) that were absent in B73 reference genome. Out of 570 genes, 292 (55 %) showed homology with plant proteins and nearly half of them (267 out of 570) could be functionally classified. For example, seven were members of a LRR family and two belonged to the NB-ARC (nucleotide-binding adaptors shared by R proteins) family, suggesting that these PAV genes might be involved in strain-specific disease-resistance.

PAVs and the large-effect SNPs and indels, allowed for testing the hypotheses of heterosis. One of the hypotheses suggests that heterosis results from complementation of slightly deleterious recessive alleles (dominanace hypothesis) and the fixation of these alleles in inbred lines can result in inbreeding depression (Fu and Dooner 2002; Springer and Stupar 2007; Charlesworth and Willis 2009). The authors demonstrated that inbred lines from different heterotic groups contain different sets of deleted genes suggesting that these lines have large differences in gene content that could complement one another, thereby contributing to heterosis.

Low-pass whole-genome sequencing of pooled genomic DNA samples from wild and domesticated rice was used to identify genomic regions subjected to selection during domestication (He et al. 2011). Sixty-six rice accessions from three rice taxa (*O. sativa japonica*, *O. sativa indica* and *O. rufipogon*) were fully sequenced by both Illumina-Solexa-GA and ABI-SOLiD in order to (1) identify genomic regions exhibiting a genealogy distinct from the rest of the genome (2) explain how these regions reflect the process of domestication under artificial selection and (3) to identify additional domestication genes in these regions. In this study, sequencing was performed on pooled DNA samples of each subspecies (21–23 accessions per subspecies) rather than on individual accessions resulting in the coverage of about 30X for each pooled sample or 1.5X per accession. Even though low genome coverage has been achieved for individual accessions, the estimation of overall genetic diversity in domesticated rice and its wild ancestor could be efficiently performed in pooled samples (Lynch 2009). Watterson's θ, which corresponds to the total number of variable sites in the alignment, was used as a measure of genetic diversity. Only those polymorphic sites that were supported by data generated using both Illumina GA and ABI-SOLiD instruments were retained and used for diversity analysis. Interestingly, combined data from both sequencing technologies produced more reliable predictions of variable sites than that based on either SOLiD or Illumina GA data alone.

In order to detect genomic regions subjected to artificial selection during domestication, the authors of this study estimated the Watterson's θ in a sliding window of 100 kb (10 kb steps) across genome for all three rice samples. The regions showing the levels of diversity (LDRs) below the genome-wide average were selected. Then, based on a series of simulations, the authors concluded that the excess of LDRs in the domesticated cultivars can be partially attributed to the population-size reduction during domestication and the effect of selfing on recombination. To separate the effect of demography and selfing on selective sweeps due to domestication, the authors investigated the LDR between *indica* and *japonica* cultivars. Since both domesticated sub-species were selected for a similar suite of characteristics, it was reasonable to hypothesize that the same genes might be affected. The distribution of LDR regions across *indica* and *japonica* genomes suggested that they have been domesticated independently, with many overlapping domesticated regions originating only once and spreading across the entire domesticated population.

Recently, whole-genome sequencing of multiple accessions of maize wild relatives, landraces and cultivars was used to identify genomic regions subjected to selection during domestication and improvement (Chia et al. 2012; Hufford et al. 2012). By analyzing patterns of variation across maize genome, the evidence of stronger selection during maize domestication than improvement was found (Hufford et al. 2012). The selection scan based on genetic differentiation of linked markers identified new candidate genes showing the evidence of stronger selection than genes previously shown to underlie major morphological changes associated with domestication. In addition to tens of millions of SNPs, a large number of copy number variants were discovered in maize (Chia et al. 2012). It was shown that a significant fraction of associations in the GWAS results (15–27 %) is linked with structural variation suggesting its importance in phenotypic variation. These studies provided a first comprehensive view of genetic diversity in a crop genome during transition from wild forms to cultivated varieties.

4.5 Summary and Outlook

Large-scale NGS is gaining popularity as a tool for fast analysis of genome level sequence variation in a number of important crop species including maize, rice, wheat, barley and sorghum. The development of genome complexity reduction approaches based on targeted sequence capture or restriction enzyme digest combined with multiplexing capabilities of NGS approaches allows for cost-effective analysis of genetic variation in large populations for genome-wide association mapping, analysis of population structure or fine-scale linkage mapping. The improvement of computational algorithms specifically adapted to handle large volumes of NGS data significantly simplified sequence assembly, alignment and variant detection in NGS projects, leading to accelerated progress in the analysis of cereal genomes. The development of new NGS instrumentation holds a promise

to further reduce the cost of sequencing and increase the amount of data generated, making feasible in near future to sequence complete genomes of even very complex organisms. In spite of these advances, the analysis of NGS data and its application to cereal crop breeding and genetics is still in its infancy. Even though NGS data can now be easily obtained for any organism, the challenges associated with the experimental design, the error rate in NGS data and in assembled reference genomes, errors in read mapping, and the complexity of cereal genomes stemming from high repeat DNA content and polyploidy will still require active exploration in next few years. These factors should be taken into account while designing NGS experiments and analyzing data.

Acknowledgments This work was supported by the USDA National Institute of Food and Agriculture (2009-65300-05638), BARD (IS-4137-08), and Triticeae CAP (2011-68002-30029) grants. We would like to thank Miranda Gray for valuable comments on the earlier versions of this chapter.

References

Akbari M, Wenzl P, Caig V, Carling J, Xia L, Yang S, Uszynski G, Mohler V, Lehmensiek A, Kuchel H, Hayden MJ, Howes N, Sharp P, Vaughan P, Rathmell B, Huttner E, Kilian A (2006) Diversity arrays technology (DArT) for high-throughput profiling of the hexaploid wheat genome. Theor Appl Genet 113:1409–1420

Akhunov ED, Akhunova AR, Anderson OD, Anderson JA, Blake N, Clegg MT, Coleman-Derr D, Conley EJ, Crossman CC, Deal KR, Dubcovsky J, Gill BS, Gu YQ, Hadam J, Heo H, Huo N, Lazo GR, Luo MC, Ma YQ, Matthews DE, McGuire PE, Morrell PL, Qualset CO, Renfro J, Tabanao D, Talbert LE, Tian C, Toleno DM, Warburton ML, You FM, Zhang W, Dvorak J (2010) Nucleotide diversity maps reveal variation in diversity among wheat genomes and chromosomes. BMC Genom 11:702

Akhunov ED, Akhunova AR, Dvorak J (2007) Mechanisms and rates of birth and death of dispersed duplicated genes during the evolution of a multigene family in diploid and tetraploid wheats. Mol Biol Evol 24:539–550

Akhunov ED, Nicolet C, Dvorak J (2009) Single nucleotide polymorphism genotyping in polyploid wheat with the Illumina GoldenGate assay. Theor Appl Genet 119:507–517

Albert TJ, Molla MN, Muzny DM, Nazareth L, Wheeler D, Song X, Richmond TA, Middle CM, Rodesch MJ, Packard CJ, Weinstock GM, Gibbs RA (2007) Direct selection of human genomic loci by microarray hybridization. Nat Methods 4:903–905

Allen AM, Barker GL, Berry ST, Coghill JA, Gwilliam R, Kirby S, Robinson P, Brenchley RC, D'Amore R, McKenzie N, Waite D, Hall A, Bevan M, Hall N, Edwards KJ (2011) Transcript-specific, single-nucleotide polymorphism discovery and linkage analysis in hexaploid bread wheat (*Triticum aestivum* L.). Plant Biotechnol J. doi:10.1111/j.1467-7652.2011.00628.x

Arai-Kichise Y, Shiwa Y, Nagasaki H, Ebana K, Yoshikawa H, Yano M, Wakasa K (2011) Discovery of genome-wide DNA polymorphisms in a landrace cultivar of Japonica rice by whole-genome sequencing. Plant Cell Physiol 52:274–282

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE 3:e3376

Bansal V, Tewhey R, Leproust EM, Schork NJ (2011) Efficient and cost effective population resequencing by pooling and in-solution hybridization. PLoS ONE 6:e18353

Barbazuk WB, Emrich SJ, Chen HD, Li L, Schnable PS (2007) SNP discovery via 454 transcriptome sequencing. Plant J 51:910–918

Barker GL, Edwards KJ (2009) A genome-wide analysis of single nucleotide polymorphism diversity in the world's major cereal crops. Plant Biotechnol J 7:318–325

Burrows M, Wheeler D (1994) A block-sorting lossless data compression algorithm. HP Labs Technical Reports. http://www.hpl.hp.com/techreports/Compaq-DEC/SRC-RR-124.html

Chao S, Dubcovsky J, Dvorak J, Luo MC, Baenziger SP, Matnyazov R, Clark DR, Talbert LE, Anderson JA, Dreisigacker S, Glover K, Chen J, Campbell K, Bruckner PL, Rudd JC, Haley S, Carver BF, Perry S, Sorrells ME, Akhunov ED (2010) Population- and genome-specific patterns of linkage disequilibrium and SNP variation in spring and winter wheat (*Triticum aestivum* L.). BMC Genom 11:727

Charlesworth D, Willis JH (2009) The genetics of inbreeding depression. Nat Rev Genet 10:783–796

Chia JM, Song C, Bradbury PJ, Costich D, de Leon N, Doebley J, Elshire RJ, Gaut B, Geller L, Glaubitz JC, Gore M, Guill KE, Holland J, Hufford MB, Lai J, Li M, Liu X, Lu Y, McCombie R, Nelson R, Poland J, Prasanna BM, Pyhäjärvi T, Rong T, Sekhon RS, Sun Q, Tenaillon MI, Tian F, Wang J, Xu X, Zhang Z, Kaeppler SM, Ross-Ibarra J, McMullen MD, Buckler ES, Zhang G, Xu Y, Ware D (2012) Maize HapMap2 identifies extant variation from a genome in flux. Nat Genet 44:803–807

Choulet F, Wicker T, Rustenholz C, Paux E, Salse J, Leroy P, Schlub S, Le Paslier MC, Magdelenat G, Gonthier C, Couloux A, Budak H, Breen J, Pumphrey M, Liu S, Kong X, Jia J, Gut M, Brunel D, Anderson JA, Gill BS, Appels R, Keller B, Feuillet C (2010) Megabase level sequencing reveals contrasted organization and evolution patterns of the wheat gene and transposable element spaces. Plant Cell 22:1686–1701

Chutimanitsakun Y, Nipper RW, Cuesta-Marcos A, Cistué L, Corey A, Filichkina T, Johnson EA, Hayes PM (2011) Construction and application for QTL analysis of a restriction site associated DNA (RAD) linkage map in barley. BMC Genom 12:4

Clark AG, Hubisz MJ, Bustamante CD, Williamson SH, Nielsen R (2005) Ascertainment bias in studies of human genome-wide polymorphism. Genome Res 15:1496–1502

Clark RM, Schweikert G, Toomajian C, Ossowski S, Zeller G, Shinn P, Warthmann N, Hu TT, Fu G, Hinds DA, Chen H, Frazer KA, Huson DH, Schölkopf B, Nordborg M, Rätsch G, Ecker JR, Weigel D (2007) Common sequence polymorphisms shaping genetic diversity in Arabidopsis thaliana. Science 317:338–342

Durbin RM, Abecasis GR, Altshuler D, The 1000 Genomes Project Consortium et al (2010) A map of human genome variation from population-scale sequencing. Nature 467:1061–1073

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE 6:e19379

Eveland AL, McCarty DR, Koch KE (2008) Transcript profiling by 3′-untranslated region sequencing resolves expression of gene families. Plant Physiol 146:32–44

Fu H, Dooner HK (2002) Intraspecific violation of genetic colinearity and its implications in maize. Proc Natl Acad Sci USA 99:9573–9578

Fu Y, Emrich SJ, Guo L, Wen TJ, Ashlock DA, Aluru S, Schnable PS (2005) Quality assessment of maize assembled genomic islands (MAGIs) and large-scale experimental verification of predicted genes. Proc Natl Acad Sci USA 102:12282–12287

Fu Y, Springer NM, Gerhardt DJ, Ying K, Yeh CT, Wu W, Swanson-Wagner R, D'Ascenzo M, Millard T, Freeberg L, Aoyama N, Kitzman J, Burgess D, Richmond T, Albert TJ, Barbazuk WB, Jeddeloh JA, Schnable PS (2010) Repeat subtraction-mediated sequence capture from a complex genome. Plant J 62:898–909

Ganal MW, Altmann T, Roder MS (2009) SNP identification in crop plants. Curr Opin Plant Biol 12:211–217

Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W, Fennell T, Giannoukos G, Fisher S, Russ C, Gabriel S, Jaffe DB, Lander ES, Nusbaum C (2009) Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nat Biotechnol 27:182–189

Gore MA, Chia JM, Elshire RJ, Sun Q, Ersoz ES, Hurwitz BL, Peiffer JA, McMullen MD, Grills GS, Ross-Ibarra J, Ware DH, Buckler ES (2009) A first-generation haplotype map of maize. Science 326:1115–1117

Haun WJ, Hyten DL, Xu WW, Gerhardt DJ, Albert TJ, Richmond T, Jeddeloh JA, Jia G, Springer NM, Vance CP, Stupar RM (2011) The composition and origins of genomic variation among individuals of the soybean reference cultivar Williams 82. Plant Physiol 155:645–655

He Z, Zhai W, Wen H, Tang T, Wang Y, Lu X, Greenberg AJ, Hudson RR, Wu CI, Shi S (2011) Two evolutionary histories in the genome of rice: the roles of domestication genes. PLoS Genet 7:e1002100

Huang X, Feng Q, Qian Q, Zhao Q, Wang L, Wang A, Guan J, Fan D, Weng Q, Huang T, Dong G, Sang T, Han B (2009) High-throughput genotyping by whole-genome resequencing. Genome Res 19:1068–1076

Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z, Li M, Fan D, Guo Y, Wang A, Wang L, Deng L, Li W, Lu Y, Weng Q, Liu K, Huang T, Zhou T, Jing Y, Li W, Lin Z, Buckler ES, Qian Q, Zhang QF, Li J, Han B (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. Nat Genet 42:961–967

Hufford MB, Xu X, van Heerwaarden J, Pyhäjärvi T, Chia JM, Cartwright RA, Elshire RJ, Glaubitz JC, Guill KE, Kaeppler SM, Lai J, Morrell PL, Shannon LM, Song C, Springer NM, Swanson-Wagner RA, Tiffin P, Wang J, Zhang G, Doebley J, McMullen MD, Ware D, Buckler ES, Yang S, Ross-Ibarra J (2012) Comparative population genomics of maize domestication and improvement. Nat Genet 44:808–811

Hyten DL, Song Q, Choi IY, Yoon MS, Cregan PB (2008) High-throughput genotyping with the GoldenGate assay in the complex genome of soybean. Theor Appl Genet 116:945–952

International Barley Genome Sequencing Consortium (2012) A physical, genetic and functional sequence assembly of the barley genome. Nature 491:711–716

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

Kimber G, Sears ER (1979) Uses of wheat aneuploids. Basic Life Sci 13:427–443

Lai J, Li R, Xu X, Jin W, Xu M, Zhao H, Xiang Z, Song W, Ying K, Zhang M, Jiao Y, Ni P, Zhang J, Li D, Guo X, Ye K, Jian M, Wang B, Zheng H, Liang H, Zhang X, Wang S, Chen S, Li J, Fu Y, Springer NM, Yang H, Wang J, Dai J, Schnable PS, Wang J (2010) Genome-wide patterns of genetic variation among elite maize inbred lines. Nat Genet 42:1027–1030

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler Transform. Bioinformatics 25:1754–1760

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup (2009a) The Sequence alignment/map (SAM) format and SAMtools. Bioinformatics 25:2078–2079

Li H, Ruan J, Durbin RM (2008) Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res 18:1851–1858

Li R, Li Y, Fang X, Yang H, Wang J, Kristiansen K, Wang J (2009b) SNP detection for massively parallel whole-genome resequencing. Genome Res 19:1124–1132

Li R, Yu C, Li Y, Lam TW, Yiu SM, Kristiansen K, Wang J (2009c) SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics 25:1966–1967

Lu T, Lu G, Fan D, Zhu C, Li W, Zhao Q, Feng Q, Zhao Y, Guo Y, Li W, Huang X, Han B (2010) Function annotation of the rice transcriptome at single-nucleotide resolution by RNA-seq. Genome Res 20:1238–1249

Lunter G, Goodson M (2011) Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. Genome Res 21:936–939

Lynch M (2009) Estimation of allele frequencies from high-coverage genome sequencing projects. Genetics 182:295–301

Mammadov JA, Chen W, Ren R, Pai R, Marchione W, Yalçin F, Witsenboer H, Greene TW, Thompson SA, Kumpatla SP (2010) Development of highly polymorphic SNP markers from the complexity reduced portion of maize [*Zea mays* L.] genome for use in marker-assisted breeding. Theor Appl Genet 121:577–588

Margulies M, Egholm M, Altman WE et al (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437:376–380

Marth GT, Korf I, Yandell MD, Yeh RT, Gu Z, Zakeri H, Stitziel NO, Hillier L, Kwok PY, Gish WR (1999) A general approach to single-nucleotide polymorphism discovery. Nat Genet 23:452–456

Mayer KF, Martis M, Hedley PE, Simková H, Liu H, Morris JA, Steuernagel B, Taudien S, Roessner S, Gundlach H, Kubaláková M, Suchánková P, Murat F, Felder M, Nussbaumer T, Graner A, Salse J, Endo T, Sakai H, Tanaka T, Itoh T, Sato K, Platzer M, Matsumoto T, Scholz U, Dolezel J, Waugh R, Stein N (2011) Unlocking the barley genome by chromosomal and comparative genomics. Plant Cell 23:1249–1263

McMullen MD, Kresovich S, Villeda HS, Bradbury P, Li H, Sun Q, Flint-Garcia S, Thornsberry J, Acharya C, Bottoms C, Brown P, Browne C, Eller M, Guill K, Harjes C, Kroon D, Lepak N, Mitchell SE, Peterson B, Pressoir G, Romero S, Oropeza Rosas M, Salvo S, Yates H, Hanson M, Jones E, Smith S, Glaubitz JC, Goodman M, Ware D, Holland JB, Buckler ES (2009) Genetic properties of the maize nested association mapping population. Science 325:737–740

McNally KL, Childs KL, Bohnert R, Davidson RM, Zhao K, Ulat VJ, Zeller G, Clark RM, Hoen DR, Bureau TE, Stokowski R, Ballinger DG, Frazer KA, Cox DR, Padhukasahasram B, Bustamante CD, Weigel D, Mackill DJ, Bruskiewich RM, Rätsch G, Buell CR, Leung H, Leach JE (2009) Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. Proc Natl Acad Sci USA 106:12273–12278

Metzker ML (2007) Sequencing technologies—the next generation. Nat Rev Genet 11:31–46

Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA (2007) Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome Res 17:240–248

Nelson JC, Wang S, Wu Y, Li X, Antony G, White FF, Yu J (2011) Single-nucleotide polymorphism discovery by high-throughput sequencing in sorghum. BMC Genom 12:352

Nielsen R, Paul JS, Albrechtsen A, Song YS (2011) Genotype and SNP calling from next-generation sequencing data. Nat Rev Genet 12:443–451

Okou DT, Steinberg KM, Middle C, Cutler DJ, Albert TJ, Zwick ME (2007) Microarray-based genomic selection for high-throughput resequencing. Nat Methods 4:907–909

Oliver RE, Lazo GR, Lutz JD, Rubenfield MJ, Tinker NA, Anderson JM, Wisniewski Morehead NH, Adhikary D, Jellen EN, Maughan PJ, Brown Guedira GL, Chao S, Beattie AD, Carson ML, Rines HW, Obert DE, Bonman JM, Jackson EW (2011) Model SNP development for complex genomes based on hexaploid oat using high-throughput 454 sequencing technology. BMC Genom 12:77

Ossowski S, Schneeberger K, Clark RM, Lanz C, Warthmann N, Weigel D (2008) Sequencing of natural strains of Arabidopsis thaliana with short reads. Genome Res 18:2024–2033

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, Mehboob-ur-Rahman, Ware D, Westhoff P, Mayer KF, Messing J, Rokhsar DS (2009) The Sorghum bicolor genome and the diversification of grasses. Nature 457:551–556

Paterson AH, Freeling M, Tang H, Wang X (2010) Insights from the comparison of plant genome sequences. Annu Rev Plant Biol 61:349–372

Porreca GJ, Zhang K, Li JB, Xie B, Austin D, Vassallo SL, LeProust EM, Peck BJ, Emig CJ, Dahl F, Gao Y, Church GM, Shendure J (2007) Multiplex amplification of large sets of human exons. Nat Methods 4:931–936

Rostoks N, Ramsay L, MacKenzie K, Cardle L, Bhat PR, Roose ML, Svensson JT, Stein N, Varshney RK, Marshall DF, Graner A, Close TJ, Waugh R (2006) Recent history of artificial outcrossing facilitates whole-genome association mapping in elite inbred crop varieties. Proc Natl Acad Sci USA 103:18656–18661

Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, Leamon JH, Johnson K, Milgrew MJ, Edwards M, Hoon J, Simons JF, Marran D, Myers JW, Davidson JF, Branting A, Nobile JR, Puc BP, Light D, Clark TA, Huber M, Branciforte JT, Stoner IB, Cawley SE,

Lyons M, Fu Y, Homer N, Sedova M, Miao X, Reed B, Sabina J, Feierstein E, Schorn M, Alanjary M, Dimalanta E, Dressman D, Kasinskas R, Sokolsky T, Fidanza JA, Namsaraev E, McKernan KJ, Williams A, Roth GT, Bustillo J (2011) An integrated semiconductor device enabling non-optical genome sequencing. Nature 475:348–352

Saintenac C, Jiang D, Akhunov ED (2011) Targeted analysis of nucleotide and copy number variation by exon capture in allotetraploid wheat genome. Genome Biol 12:R88

Salse J, Bolot S, Throude M, Jouffe V, Piegu B, Quraishi UM, Calcagno T, Cooke R, Delseny M, Feuillet C (2008) Identification and characterization of shared duplications between rice and wheat provide new insight into grass genome evolution. Plant Cell 20:11–24

Sasaki A, Ashikari M, Ueguchi-Tanaka M, Itoh H, Nishimura A, Swapan D, Ishiyama K, Saito T, Kobayashi M, Khush GS, Kitano H, Matsuoka M (2002) Green revolution: a mutant gibberellin-synthesis gene in rice. Nature 416:701–702

Schadt EE, Turner S, Kasarskis A (2010) A window into third-generation sequencing. Hum Mol Genet 19:R227–R240

Schnable PS, Ware D, Fulton RS et al (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112–1115

Springer NM, Stupar RM (2007) Allelic variation and heterosis in maize: how do two halves make more than a whole? Genome Res 17:264–275

Teer JK, Bonnycastle LL, Chines PS, Hansen NF, Aoyama N, Swift AJ, Abaan HO, Albert TJ, NISC Comparative Sequencing Program, Margulies EH, Green ED, Collins FS, Mullikin JC, Biesecker LG (2010) Systematic comparison of three genomic enrichment methods for massively parallel DNA sequencing. Genome Res 20:1420–1431

Tian F, Bradbury PJ, Brown PJ, Hung H, Sun Q, Flint-Garcia S, Rocheford TR, McMullen MD, Holland JB, Buckler ES (2011) Genome-wide association study of leaf architecture in the maize nested association mapping population. Nat Genet 43:159–162

Tian F, Stevens NM, Buckler ES 4th (2009) Tracking footprints of maize domestication and evidence for a massive selective sweep on chromosome 10. Proc Natl Acad Sci USA 106(Suppl 1):9979–9986

Wang RL, Stec A, Hey J, Lukens L, Doebley J (1999) The limits of selection during maize domestication. Nature 398:236–239

Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, He W, Chen YJ, Makhijani V, Roth GT, Gomes X, Tartaro K, Niazi F, Turcotte CL, Irzyk GP, Lupski JR, Chinault C, Song XZ, Liu Y, Yuan Y, Nazareth L, Qin X, Muzny DM, Margulies M, Weinstock GM, Gibbs RA, Rothberg JM (2008) The complete genome of an individual by massively parallel DNA sequencing. Nature 452:872–876

Whitt SR, Wilson LM, Tenaillon MI, Gaut BS, Buckler ES (2002) Genetic diversity and selection in the maize starch pathway. Proc Natl Acad Sci USA 99:12959–12962

Wicker T, Mayer KF, Gundlach H, Martis M, Steuernagel B, Scholz U, Simková H, Kubaláková M, Choulet F, Taudien S, Platzer M, Feuillet C, Fahima T, Budak H, Dolezel J, Keller B, Stein N (2011) Frequent gene movement and pseudogene evolution is common to the large and complex genomes of wheat, barley, and their relatives. Plant Cell 23:1706–1718

You FM, Huo N, Deal KR, Gu YQ, Luo MC, McGuire PE, Dvorak J, Anderson OD (2011) Annotation-based genome-wide SNP discovery in the large and complex *Aegilops tauschii* genome using next-generation sequencing without a reference genome sequence. BMC Genom 12:59

Yu J, Buckler ES (2006) Genetic association mapping and genome organization of maize. Curr Opin Biotechnol 17:155–160

Chapter 5
Genome Sequencing and Comparative Genomics in Cereals

Xi-Yin Wang and Andrew H. Paterson

5.1 Introduction

Poaceae, formerly known as Gramineae, and commonly referred to as grasses, is a family of monocotyledonous flowering plants. With about 600 genera and ~10,000 species, grasses represent economically the most important family of flowering plants (Watson 1992), accounting for about 70 % of crops according to land use (Fig. 5.1). Grasses, particularly the cereals, are grown for their edible seeds, and are the primary source of human nutrition, providing more than half of all our calories and appreciable protein (Kellogg 2001). In particular, these crops, include rice as a staple food in southern and eastern Asia; maize in central and south America; wheat and barley in Europe, northern Asia and the Americas; and sorghum in some African countries. Sugarcane is the major source of sugar. The top four global agricultural commodities by quantity belong to crops from the grass family (sugarcane, maize, rice, wheat) (Global Perspective Studies Unit 2006). Many other grasses are grown for forage and fodder. Cow's milk, the sole animal product in the top 10 agricultural commodities by quantity, largely comes from grass-fed animals (Bevan et al. 2010). Other probable uses of grasses include building construction throughout east Asia (bamboo) and sub-Saharan Africa (sorghum), and paper-making (*Miscanthus*). They are also used in water treatment, wetland habitat preservation and land reclamation. Grasses with C4 photosynthesis, including *Miscanthus*, switchgrass, sugarcane, and sorghum, are attractive for

X.-Y. Wang (✉) · A. H. Paterson (✉)
Plant Genome Mapping Laboratory, University of Georgia, Athens, GA 30602, USA
e-mail: wang.xiyin@gmail.com

A. H. Paterson
e-mail: paterson@plantbio.uga.edu

X.-Y. Wang
Center for Genomics and Computational Biology,
School of Life Sciences and School of Sciences,
Hebei United University, Tangshan 063009 Hebei, China

P. K. Gupta and R. K. Varshney (eds.), *Cereal Genomics II*,
DOI: 10.1007/978-94-007-6401-9_5,

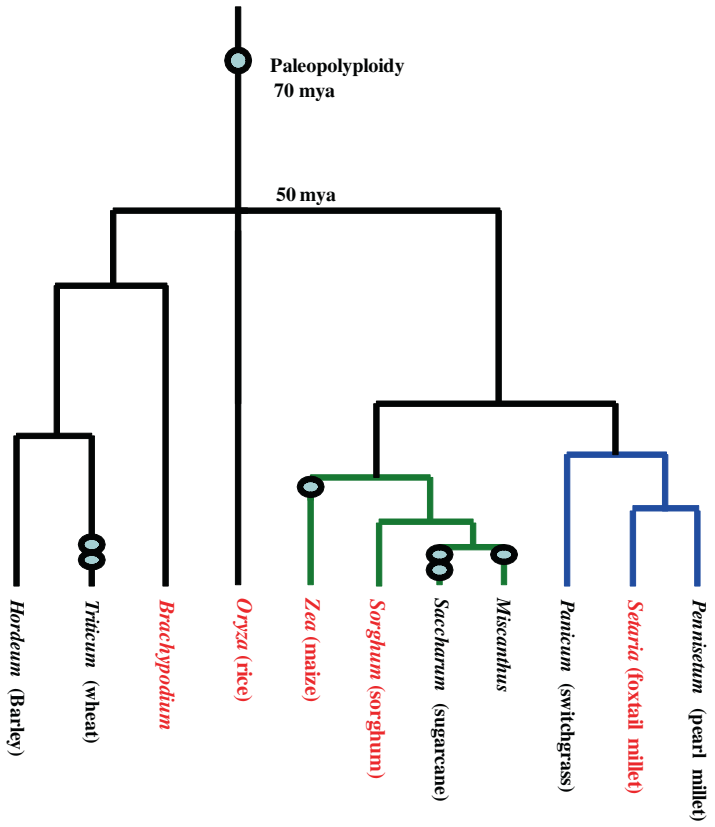

Fig. 5.1 Phylogeny of selected cereals. Polyploidization events are shown with *circles*. *Green* and *blue* branches show two C4 lineages, and the others are C3 plants. The names of five grasses having whole-genome sequenced are shown in *red*

biofuel production. A growing human population (9 billion by 2050) and expected increase in living standards will require the ongoing sustainable exploitation of grass resources.

In view of their importance to humanity, representatives of all three major grass clades have been sequenced. These include rice (International Rice Genome Sequencing Project 2005; Yu et al. 2005) from Ehrhartoideae, sorghum (Paterson et al. 2009a, b) and maize (Schnable et al. 2009) from Panicoideae, and *Brachypodium* (The International Brachypodium Initiative 2010) from pooideae (Fig. 5.1). Draft sequence of barley and wheat (Triticeae) group 1 chromosomes was also recently made available (Mayer et al. 2011; Wicker et al. 2011).

The importance of the rice genome is reflected in the fact that rice (*Oryza sativa*) was the first crop plant to have its genome sequenced. Actually, it was sequenced by four independent groups, including Beijing Institute of Genomics (BGI) (Yu et al. 2002, 2005), International Rice Genome Sequencing Project (IRGSP) (International

Rice Genome Sequencing Project 2005), Syngenta (Goff et al. 2002), and Monsanto. BGI analyzed the parental strains, 93–11 and PA64 s, of a popular super-hybrid rice, LYP9. Both rice indica and japonica subgenomes have been sequenced up to ~6x coverage. IRGSP presented a map-based, finished quality sequence that covers 95 % of the 389 Mb genome, virtually including all the euchromatic regions and even two complete centromeres (International Rice Genome Sequencing Project 2005). Both BGI and IRGSP reported ~38,000 rice genes. Syngenta also used a WGS method and released a 10x draft that incorporates the Syngenta data.

Sorghum (*Sorghum bicolor*) is a representative of the tribe Saccharinae, which includes some of the most efficient biomass accumulators to provide food and fuel, and also those having potential for use as cellulosic biofuel crops. This is the second grass, whose ~730 Mb genome sequence from *S. bicolor* (L.) Moench has been published (Paterson et al. 2009a). The WGS sequence has been carefully validated by genetic, physical and syntenic information. Despite a repeat content of 61 %, a high-quality genome sequence was assembled from homozygous sorghum genotype BTx623 by using WGS and incorporating the following: (1) 8.5 genome equivalents of paired-end reads from genomic libraries spanning a100-fold range of insert sizes, resolving many repetitive regions; and (2) high-quality read length averaging 723 bp, facilitating assembly. Comparison with 27 finished bacterial artificial chromosomes (BACs) showed the WGS assembly to be 98.46 % complete and accurate to ~1 error per 10 kbp.

Maize (*Zea mays*) is also an important model crop species, whose 2.3 Gb genome sequence was reported (Schnable et al. 2009). This whole genome sequence was assembled using genetic and physical maps. Over 32,000 genes were predicted, of which 99.8 % were placed on reference chromosomes. Nearly 85 % of the genome was shown to be composed of hundreds of families of transposable elements, dispersed non-uniformly across the genome.

Foxtail millet (*Setaria italica*), as an important grain crop in temperate, subtropical and tropical Asia, and in parts of southern Europe, and is also grown for forage in some other regions. It is a diploid grass with a relatively small genome (~515 Mb), which has been sequenced and published on the Joint Genome Institute website (Bennetzen et al. 2012; Zhang et al. 2012). The current annotation is version 2.1, and the sequence reached up to 8.3X coverage of the genome.

Brachypodium (*B distachyon*) is a wild grass and is a representative of the subfamily Pooideae, which contains bread wheat, whose genome size is large and complex. This wild grass was selected to have its genome sequenced, and the effort established a template for the analysis of economically important pooid grasses (The International Brachypodium Initiative 2010). The five compact Brachypodium pseudochromosomes contain 272 Mb, and the assembly was validated by cytogenetic analysis and alignment with two physical maps and sequenced BACs.

The genomes of the above five grasses share considerably gene collinearity; many genes were preserved at their ancestral location after millions years of divergence, which helps to perform profound comparative genomics analysis, as shown below. The sequencing efforts of these grasses will contribute to an understanding of the domestication and agricultural improvements of staple crops (Table 5.1).

Table 5.1 Grasses having whole-genome sequences so far

Species name	Common name	Release version
Brachypodium distachyon	Purple false brome	Phytozome v6.0
Oryza sativa ssp. japonica	Rice	RAP 2.0 (Nov 2007)
Oryza sativa ssp. indica	Rice	BGI v1
Sateria italica	Foxtail millet	JGI v2.1
Sorghum bicolor	Sorghum	Sbi 1.4 (Dec 2007)
Zea mays	Maize	Release 5a (Nov 2010)

Besides whole-genome sequences, transcriptome and other genome sequence datasets have been fast accumulating. In total, an astounding ~50 Poaceae species have transcriptome sequence databases (Buell 2009). A few prominent databases include Gramene [http://www.gramene.org/; (Youens-Clark et al. 2011)], the Rice Genome Annotation Resource (http://rice.tigr.org), the Rice Genome Automated Annotation System (RiceGAAS: hppt://ricegaas.dna.affrc.go.jp), the Rice Annotation Database (RAD: http://rad.dna.affrc.go.jp), the Integrated Rice Genome Explorer (INE: http://rgp.dna.affrc.go.jp/jiot/INE.html), and Oryzabase (http://www.shigen.nig.ac.jp/rice/oryzabase/top/top.jsp), Panzea (http://www.panzea.org), maizeGDB (http://www.maizedb.org), and the Comparative Saccharineae Genome Resource (http://csgr.agtec.uga.edu).

Although chloridoid and arundinoid grasses are explored only at the EST level to date, initial analyses of the available genome sequences have shown that most if not all grasses experienced at least two whole-genome duplications perhaps 20 million years or more prior to the divergence of these lineages. It has also been shown that the post-divergence evolution of gene contents and gene orders of the three grass branches with sequenced genomes are relatively conservative, except for additional lineage-specific polyploidizations. Taxa within each lineage have independently experienced polyploid formation and adaptation to the duplicated state (for example sugarcane, and durum or bread wheat). However, the model genomes were used as a good starting point for accelerating progress in the study and improvement of recursive polyploidy taxa. Indeed, the vast majority of monocots still lack sufficient genomic tools to investigate pertinent problems in agricultural productivity, conservation biology, ecology, invasion biology, population biology, and systematic biology (Paterson et al. 2009a, b). By developing and using comparative genomics tools, research starting from the sequenced genomes may shed light on these monocots (Van de Peer 2004; Wang et al. 2006; Lohithaswa et al. 2007; Tang et al. 2008b). The sequences of additional grasses and non-grass monocot genera such as *Elaeis*, *Musa*, and *Zostera* will clarify the functional innovation of their gene sets, further clarifying the structural and functional evolution of this important and interesting plant family. Re-sequencing of diverse germplasms of many species including wild and cultivated species promises to clarify the process of domestication of members of this important family.

5.2 Ancestral Polyploidy Preceded the Diversification of Grasses

5.2.1 Recursive Polyploidizations During Grass Evolution

Though neopolyploid taxa are not rare, it was surprising when the sequenced small genomes of Arabidopsis and rice each revealed ancestral polyploidy of both eudicots and monocots (The Arabidopsis Genome Initiative 2000; Bowers et al. 2003; Paterson et al. 2004), the two major angiosperm clades. These facts led to the thought that most if not all angiosperm lineages might have been shaped by a few common paleo-polyploidization events; some of these were further modified by additional recent polyploidization events (Paterson 2005; Soltis 2005). For instance, on the basis of RFLP maps, both sorghum and rice were initially shown to have large-scale duplications (Chittenden et al. 1994; Kishimoto et al. 1994; Nagamura et al. 1995). This inference later received support with the availability of the whole genome sequence of rice, *O. sativa* ssp. *Japonica* (Goff et al. 2002). A brief controversy about the scope of duplication (Vandepoele et al. 2003) was soon reconciled, by analyzing the genome sequence of another rice cultivar, *O. sativa* ssp. *Indica* (Yu et al. 2005), with the controversy attributed to differences in the approaches used to infer the duplicated blocks (Wang et al. 2005). This polyploidy event was dated to ~70 million years ago (mya) based on putatively neutral DNA substitution rates between duplicated genes, and suggested to be shared by all main lineages of grasses. This inference was later proved with genome sequences of sorghum, and Brachypodium (The International Branchypodium Initiative 2010). Maize (Schnable et al. 2009) had another polyploidy event (Gaut and Doebley 1997), which possibly contributed to its origin (Swigonova et al. 2004a, b). Multiple alignments of sequenced grass genomes shed light on a more ancient history of their common ancestor, and revealed one or more two earlier polyploidization(s) in the monocot lineage (Tang et al. 2010). This suggested that recursive polyploidy events may have shaped the evolution of grasses in a cyclic manner. Recursive polyploidizations have also been observed during the evolution of eudicots (Jaillon et al. 2007; Tang et al. 2008a).

5.2.2 Gene Collinearity Facilitated Paleogenomic Exploration

Characterization of gene synteny (gene content) and collinearity (gene order) among the major grass lineages reveals a "whole-genome duplication" pattern due to the shared polyploidy events (Fig. 5.2). The duplicated regions cover 68 % of the genome in rice (Paterson et al. 2004). Among the duplicated genes in these regions, 30–65 % have lost at least one duplicated copy, possibly soon after polyploidization and certainly prior to the radiation of the major grass lineages, so that the retention/loss patterns are largely orthologous between rice and sorghum (Paterson et al. 2009a, b). Gene losses often occurred in a complementary and segmental manner,

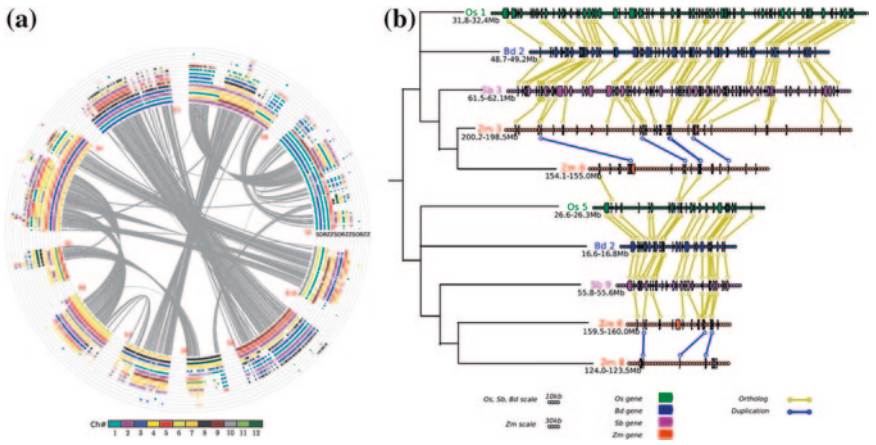

Fig. 5.2 Gene collinearity between chromosomes. **a** A *circular* display of aligned genomes of rice (O or Os), *Brachypodium* (B or Br), sorghum (S or Sb), and maize (Z or Zm), with rice as the reference. The colorful lines in *circles* denote genes along chromosomes and color scheme is according to the chromosome numbers as shown in the *lower part* of the figure. The *inner five circles* consist of orthologous chromosomes/blocks from the four grasses, and contain two *maize circles* for its specific polyploidization. The *outer five circles* are the duplicated blocks, produced in the ancestral polyploidization, respectively corresponding to the *inner circles*. The *curves* in the *innermost circles* show collinear genes produced by the pan-grass polyploidization. **b** A local segment of aligned chromosomes from the global alignment shown in (**a**). Chromosomes are shown in *parallel lines*, and genes are shown with *arrows*. Genes from a grass are in the same color. Orthologs and duplicated genes are linked with lines in different colors. A schematic tree is used to show the relationship of the chromosomal segments, and genes on them

resulting in a non-random patterns of retention/loss on corresponding duplicated DNA segments, in a process known as fractionation (Thomas et al. 2006). Genes may be removed by a short-DNA deletion mechanism (Woodhouse et al. 2010). In a pair of duplicates, gene loss may be universally biased to preserve the gene that is responsible for the majority of total expression (Schnable et al. 2011). Gene losses may have been accompanied by wide-spread genomic repatterning, due to the unstable nature of multi-valent meiosis, eventually restoring bivalent meiosis (Bowers et al. 2005). For more than 90 % of the preserved duplicated genes, the two copies have the same transcriptional orientations (Wang et al. 2005), and the exceptions may be a result of local DNA inversions or differential gains/losses of new tandemly duplicated genes in the paleo-duplicated regions.

Rice, sorghum and Brachypodium, which have not been affected by additional polyploidization after the split with one another, have preserved nearly perfect gene collinearity (Paterson et al. 2009a, b; The International Brachypodium Initiative 2010), making it possible to take them as a single genetic system to perform transitive genetics research across different grasses (Freeling 2001). Only a small fraction of genes show differential gene losses after the split of rice (1.8 %) and sorghum (3.1 %). A total of ~12,000 orthologous genes might

have represented the ancestral genome of these major cereals (Salse et al. 2009). Consistently, these grasses have similar gene content, having gene numbers ranging from 25,532 in Brachypodium to 32,540 in maize, despite nearly eightfold variation in genome size, ranging from 320 Mb in Brachypodium to 2.5 Gb in maize. These findings suggest that, after 70 mya polyploidization, the genome of the last universal common ancestor of grasses had already experienced most gene loss and reached a relatively stable state prior to the divergence of the major grass lineages about 50 mya (Paterson 2008; The International Brachypodium Initiative 2010).

Genes duplicated due to whole-genome duplication were differentially preserved in grasses, but show retention/loss patterns that are related to those observed in other taxa (Paterson et al. 2006). A comparison of duplicated genes produced in independent duplication events during the evolution of grasses, Arabidopsis, yeast and fish, indicated that retention or loss of protein functional domain-containing genes has been convergent. Preferential gene losses or retentions were also revealed in maize genome (Schnable et al. 2009).

5.2.3 *Rules of Large-Scale Genomic Repatterning After Polyploidization*

As noted above, a polyploidy event may result in genomic instability, consequently incurring a process of diploidization, characterized by wide-spread DNA rearrangements often accompanied by large-scale gene losses (The Arabidopsis Genome Initiative 2000; Paterson et al. 2004; Van de Peer 2004; Wang et al. 2006). These DNA rearrangements may result in chromosome number variations. Grasses range from 2 to 18 in their basic chromosome sets (Soderstrom et al. 1987; Hilu 2004). In the sequenced genomes, rice, sorghum, and Brachypodium have $n = 12$, 10, and 5 chromosomes, respectively. Though experiencing a whole-genome duplication since their divergence, modern day maize retains the same chromosome number (10) as sorghum. Comparison of grass genomes has shed light on the rules of chromosome number evolution and ancestral grass karyotypes (Salse et al. 2009; Murat et al. 2010). An ancestral karyotype of $n = 5$ chromosomes was inferred (Salse et al. 2009; Murat et al. 2010) before the grass-common polyploidization, with $n = 2$ $x = 10$ chromosomes after the duplication, then two chromosome fissions to result in $n = 2$ $x = 12$ chromosomes in the common ancestor of major cereals. However, the authors noted that an ancestral karyotype of $n = 6$–7 was also possible. They inferred that chromosome number variation/reduction from the common ancestor may be attributed to nonrandom centric double-strand break repair events. It was suggested that centromeric/telomeric illegitimate recombination between nonhomologous chromosomes led to nested chromosome fusions and synteny break points, and concluded that these breakpoints were meiotic recombination hotspots that corresponded to high sequence turnover loci through repeat invasion. These rules seem to explain most of the alterations in chromosome number in the grass genomes sequenced so far, especially the previously observed nested chromosome fusions in

Brachypodium (The International Brachypodium Initiative 2010). However, many details related to dynamics of centromeres and telomeres during the rearrangements remain unclear, and the wide range in possible ancestral karyotypes (from 12 to 24) suggests that further revision of thinking on this subject is likely.

5.2.4 Genome Size Variation is Mainly Attributable to Differential Accumulation of Repetitive DNA Sequence

Though gene contents (synteny) and even gene orders (collinearity) have been well preserved among grasses, there is a stark difference in their genome sizes, spanning at least a 50-fold range from ~300 million base pairs of Brachypodium to ~16,000 million base pairs of hexaploid wheat (Arumuganathan and Earle 1991). Genes often reside in the euchromatic regions, the sizes of which remain similar among different grasses (Feuillet and Keller 1999). Euchromatic regions also represent regions where gene collinearity has been well preserved and homologous recombination occurs at the highest frequency in each studied genome (Bowers et al. 2005). In contrast, heterochromatin, which contains the centromeres, often shows remarkable differences in size, arrangement, and content among taxa. For example, 65 % of sorghum DNA is in heterochromatin versus only about 22 % in rice, with the differing quantity of heterochromatin explaining 75 % of their 300 Mb genome size difference (Paterson et al. 2009a, b). The expansion of heterochromatin is also largely responsible for more than threefold size difference between maize (2.3 Gb) and sorghum (730 Mb) (Schnable et al. 2009). The expansion of heterochromatin is prominently due to the accumulation of long-terminal repeat retroelement-like (LTR) sequences, but not other repeats such as DNA transposons. However, different LTRs show different degrees of accumulation and are distributed somewhat differently across the genomic landscape. For example, most medium- and high-copy-number LTRs, such as *gypsy* LTRs, preferentially accumulated in gene-poor regions like pericentromeric heterochromatin, whereas a few high-copy-number LTRs, such as *copia* LTRs, exhibited the opposite bias and preferentially accumulate in gene-rich regions like euchromatin (Baucom et al. 2009; Devos 2010).

5.3 Illegitimate Recombination Between Duplicated DNA Segments may Contribute to Genetic Novelties and Species Diversification

5.3.1 Illegitimate Recombination in Grasses

Genetic recombination is central to molecular biology for its central role in DNA repair and reshuffling of mutations during crossovers between homologous

sequences. It is the major source of new combinations of alleles, which facilitate adaptation to environmental change, and consequently constitutes a driving force of biological evolution (Puchta et al. 1996). During meiosis, homologous chromosomes may recombine reciprocally, while during mitosis in somatic cells, recombination can be induced by DNA damage. Recombination between paralogous or homoeologous DNA regions is often termed "illegitimate", occurring between non-homologous DNA sequences. Recombination can be reciprocal involving symmetrical exchange of genetic information between paralogous loci, resulting in crossing-over; or non-reciprocal involving unidirectional transfer of information from one locus to its paralogous counterparts, resulting in gene conversion (Datta et al. 1997). Recombination, especially illegitimate recombination between paralogous loci, may produce severe chromosomal lesions, which are often deleterious but may in rare cases contribute to elimination of deleterious mutations (Khakhlova and Bock 2006). Notably, in plants, both meiotic and mitotic recombination outcomes can be transferred to the offspring, due to the lack of a predetermined germline.

Since polyploidization has affected nearly all plants during their evolution and produces massive duplicated regions in plant genomes, it is natural to question whether the appreciable sequence similarity of duplicated chromosomal regions may have permitted illegitimate recombination. Comparative and phylogenetic analysis of the sequenced genomes revealed extensive and long-lasting recombination between duplicated regions in rice and sorghum since their divergence, which was ~20 million years after the polyploidization (Wang et al. 2009b). It is estimated that at least 14 % and 12 % of rice and sorghum duplicated genes have been affected, resulting in gene conversion, possibly accompanied by crossing-over (Fig. 5.3).

Homoeologous recombination has occurred at very different rates among different duplicated regions of grass genomes, probably being restricted in most regions due to DNA rearrangement (Wang et al. 2009b). Inversion of DNA

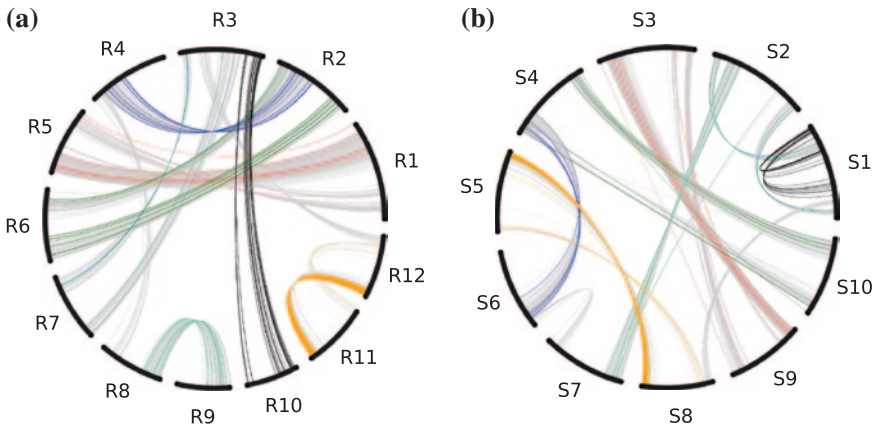

Fig. 5.3 Genome duplication and conversion pattern, in rice (**a**) and sorghum (**b**). Chromosomes from a grass form a *circle*. Duplicated (*grey lines*) and converted (*lines in other colors*) genes are linked with *curvy lines*, and the lines in the same color indicate that converted genes are from orthologous duplicated regions in rice and sorghum

segments in one region may hinder its recombination with the other duplicated copy. For example, inversion in the ancestor of rice chromosomes 1 or 5 may be responsible for lowering the gene conversion rate between them and between their sorghum orthologs. Clearer evidence was found in sorghum chromosomes 5 and 8, where two inversions have contributed to recombination suppression. Recombination might have been restricted in a non-synchronized manner, in that divergent gene conversion rates were found among duplicated regions. However, homoeologous recombination appears to be on-going on the initial 3 Mb of short arms' termini of rice chromosomes 11 and 12, where homoeologous sequences are very similar (Wang et al. 2007, 2009b). The authors also found that recombination occurs in higher frequency toward terminal regions of chromosomes. In rice, >50 % of fully converted genes are in the initial 2 Mb regions on the chromosomal termini, in which ~40 % of the duplicated genes have been converted. In sorghum, 48.6 % of wholly converted genes are in the initial 2 Mb regions on the chromosomal termini, in which ~34.5 % of the duplicated genes have been converted. Another interesting finding is that the rice and sorghum orthologous chromosomes/chromosomal segments have similar patterns of illegitimate recombination.

A comprehensive model has been proposed, to describe the pan genome dynamics after polyploidizationin grasses (Wang et al. 2009b). Soon after polyploidization, multiple homologous chromosomes or chromosomal segments may compete to pair and recombine with one another, forming multivalent structures during meiosis. This may confer genome instability as often inferred in paleogenomic duplications or observed in artificial polyploids. DNA rearrangement may inhibit the chance of pairing between affected chromosomes or chromosomal segments. Gradually, structural and sequence divergence may establish neo-homologous chromosome pairs, eventually re-establishing bivalent pattern during meiosis. The chromosomes or chromosomal segments sharing ancestry, but with lower level of similarity in structure and sequence, are then referred to as homoeologous chromosomes. Most DNA rearrangements, a major factor in restricting recombination, occurred before the divergence of the major grass lineages. Accordingly, different divergence levels between the ancestral homoeologous chromosomes in the cereal common ancestor may have led to variation in conversion rates among duplicated regions within a genome, but similar conversion rates between orthologous regions from different species. The size of duplicated regions is positively correlated to recombination rate (Wang et al. 2009b), with larger sizes thought to provide higher DNA similarity, facilitating recombination. When small duplicated blocks are buried in chromosomes that otherwise share little or no homoeology, they may have little chance to recombine. This may be particularly true when other regions of the chromosome do have homoeology with large segments of other chromosomes, leaving the small duplicated regions at a disadvantage in forming homoeologous duplexes. Higher homoeologous recombination near chromosome termini, just like higher homologous recombination therein, may be explained by higher gene density and less accumulation of repetitive sequences in comparison to the pericentromeric regions, which provide higher DNA-level similarity needed for chromosomes to pair and recombine.

5.3.2 *Homoeologous Recombination Accelerates Gene Evolution*

Gene conversion as a result of recombination homogenizes homoeologous gene sequences, and would make the affected paralogs appear more similar to one another than would be predicted based on their true age. Pa and Ps values (non-synonymous and synonymous nucleotide substitution percentages) between paralogs affected by homoeologous recombination are often smaller than between those not affected (Wang et al. 2009b). Converted rice homoeologs have an average Ps 0.15 and Pa 0.06, significantly smaller than those of not converted (0.49 and 0.20). In sorghum genome also, converted sorghum homoeologs have an averaged Ps 0.24 and Pa 0.08, that are also significantly smaller than those of not converted (0.50 and 0.19). An intriguing question is: do the converted genes evolve slowly? One could not find the answer based on the paralogs themselves, the pairwise distance between which could have been distorted by gene conversion. Comparative analysis of the corresponding orthologs can provide some insight. A comparison of the nucleotide differences of orthologs, whose paralogs were affected by whole gene conversion in rice and sorghum, to those whose paralogs were not affected in either species, indicated that the conversion-related group has a little larger Pa and Ps, showing that converted paralogs evolve faster than those not converted. Though converted paralogs are more similar to one another than those, not converted, this intriguing finding means that homoeologous recombination acts as a diversifying rather than a conservative element in evolution.

Illegitimate recombination may influence natural selection in converted genes. Converted paralogs in both rice and sorghum have an average Pa/Ps ratios that are significantly lower (0.34 and 0.31) than those which are not converted (0.44 and 0.42) (Wang et al. 2009b), suggesting that within a genome, converted paralogs tend to be more subject to purifying selection. Likewise, converted genes tend to have more similar expression patterns than non-converted duplicates. By checking the Pfam domains in the converted and non-converted duplicated genes, the authors found weak evidence for preferential conversion of genes with specific functions.

5.4 Interrelated Evolution of Two Grass Chromosomes is an Exception to Other Plant Chromosomes

Rice chromosomes 11 and 12 (R11 and R12) are a striking exception, when compared to all other chromosomes that had been so clearly affected by the 70 mya polyploidization (Wang et al. 2009b). R11 and R12, at first seemingly having avoided duplication during the pan-grass polyploidization, share a ~3 Mb duplicated DNA segment at the termini of their short arms, the formation of which had been dated on the basis of synonymous substitutions to ~5–7 mya (The Rice Chromosomes 11 and 12 Sequencing Consortia 2005; Wang et al. 2005; Yu et al.

2005). Remarkably, the corresponding region(s) of their orthologous chromosomes in sorghum genome (S5 and S8, respectively) also contained such an apparently recent duplication despite having diverged from rice about 50 mya (Paterson et al. 2009a, b). Based on physical and genetic maps, different groups reported shared terminal segments of the corresponding chromosomes in wheat (4, 5), foxtail millet (7, 8) and pearl millet (linkage groups 1, 4) (Devos et al. 2000; Singh et al. 2007). It would be exceedingly unlikely that segmental duplications in each of these crops occur independently at such closely corresponding locations in reproductively isolated lineages. A much more parsimonious hypothesis is that the rice R11/12 and *sorghum* S5/8 regions each resulted from the pan-grass duplication 70 mya but have an unusual evolutionary history (Paterson et al. 2009a, b).

5.4.1 Long-Lasting Illegitimate Recombination and Stepwise Restriction Along Homoeologous Chromosomes

Detailed analysis of the above two exceptional chromosomes and their homoeologs from Brachypodium (The International Brachypodium Initiative 2010) and maize (Schnable et al. 2009) suggested that illegitimate recombination has continued for millions of years after the divergence of these homoeologs, and remains on-going in rice and perhaps other grasses (Wang et al. 2007). Gradual and step-by-step restrictions on recombination, starting from the pericentromeric regions around the time of polyploidization 70 mya, have resulted in chromosome structural stratification (Wang et al. 2011), having produced old chromosomal strata (CSA, CSB, and CSC, short forms for "Common Strata A, B, and C", respectively) in their common ancestor, and produced some relatively younger strata in each species after rice-sorghum split, namely, RSA, RSB, and RSC in rice ("RS" is short form for "Rice Strata"), and SSA and SSB in sorghum ("SS" is short form for "Sorghum Strata") (Fig. 5.4).

Sequence similarity between homoeologs in the strata reflects the time(s) of recombination suppression rather than the times of their origin. Strata RSA-RSC were estimated to have formed <0.5, 9.4, and 39.1 mya, SSA and SSB were dated to 13.4 and 59.7 mya, and strata CSA–CSC, common to two species, were dated to 65.1, 80.1 and 53.5 mya (Wang et al. 2011). The corresponding regions in maize and Brachypodium also show prominent homoeologous recombination. However, wide-spread chromosomal rearrangement, especially in maize after its lineage-specific polyploidization, makes the stratification patterns more difficult to compare than in rice and sorghum.

The initial time of the formation of underlying homoeologous chromosome pair may be best indicated by the oldest stratum, CSB, estimated to have formed 72.3–91.6 mya (with 95 % confidence), perhaps even predating the 70 mya polyploidization (67.7–70 mya). In the modern rice and sorghum genomes, CSB regions of R12/S8 have much less DNA and fewer genes than their homoeologs R11/S5. This accounts for 2/3 of gene content differences between R11/12 and S5/8, which suggests fractionation (and degeneration) of the ancestral chromosome of R12/S8. Since R12 and S8 closely resemble one another and each differ from their homoeologs R11/

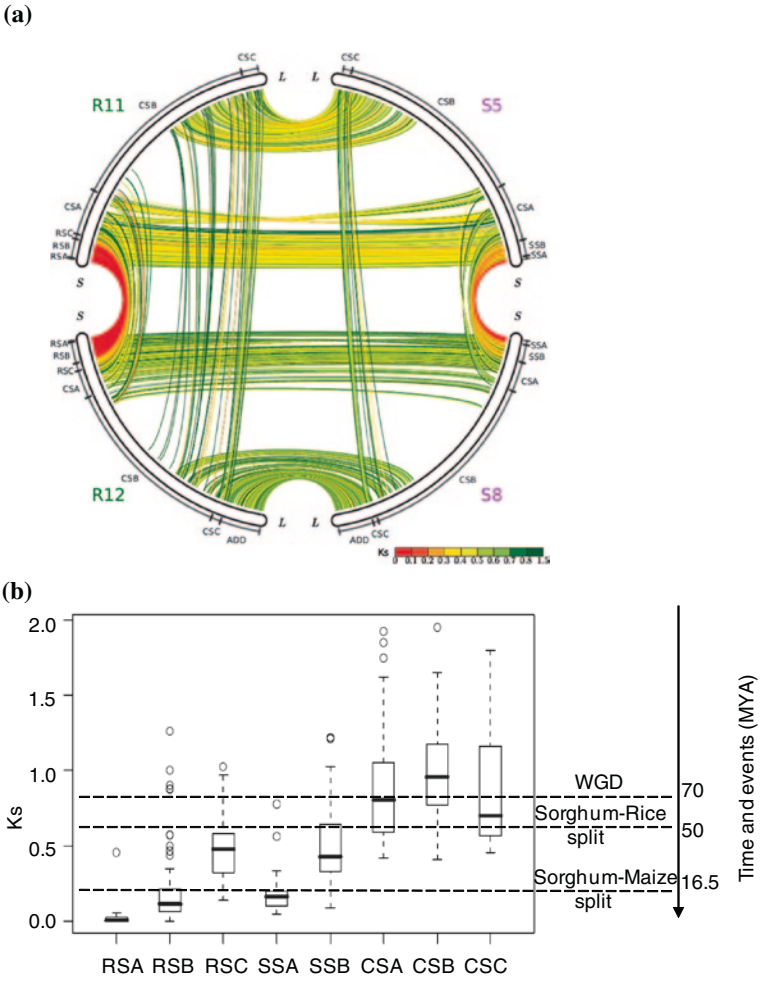

Fig. 5.4 Homology pattern and evolutionary model of chromosomes *R11*, *R12*, *S5* and *S8* and their common ancestor. **a** Strata (Chromosomal segments: *RSA–RSC*, *SSA–SSB*, and *CSA–CSC*) are displayed in subfigures. "*RS*" and "*SS*" mean strata formed after rice-sorghum divergence; *CS* means common strata formed before the divergence of two species. Additional segments on *R12* and *S8* are noted with "*ADD*". "*S*" and "*L*" indicate *short* and *long arms*. Lines between chromosomes connect syntenic genes, and colors correspond to *Ks* values. **b** The strata in (**a**) were dated according to synonymous nucleotide substitution rates (*Ks*)

S5, this fractionation presumably preceded the rice-sorghum divergence, possibly occurring near the polyploidization. If produced during the polyploidization, the two duplicated chromosomes may have initially recombined with one another and competed with their respective homologs. Their competence to recombine may have been reduced by stepwise homoeologous recombination suppression due to DNA inversions, accumulation of repetitive sequences and new members of multigene families

(such as resistance genes), with initial recombination suppression occurring in the ancestral CSB region. Each of these factors is supported by evidence. Two inversions in S8 obviously suppressed recombination between S8 and S5, contributing to the formation of new strata. The inversion breakpoints were enriched with repetitive sequences, possibly leading to DNA inversions. Notably, R11 and S5 are the chromosomes most enriched for NBS-LRR resistance genes in their respective species, accounting for up to one-fourth of the total in each species (Fig. 5.4), respectively. The parallel nature of this enrichment suggests that it had occurred before the rice-sorghum split, possibly resulting from or even contributing to recombination suppression. Recombination suppression may have led to unbalanced DNA deletion, with the ancestral chromosome of R12/S8 more affected than its homoeolog.

5.4.2 Weird Co-existence of high Intragenomic Similarity and High Intergenomic Divergence

While large-scale recombination in which other regions may have been mostly restricted, homoeologous recombination in RSA has been on-going on in the terminal regions on rice R11 and R12, during the past 400,000 years since the divergence of rice subspecies japonica and indica (Wang et al. 2007). Both intriguing and perplexing is a distal chromosomal region with the greatest DNA similarity between surviving duplicated genes, which represents the highest concentration of lineage-specific gene pairs found anywhere in these genomes with a significantly elevated gene evolutionary rate (Wang et al. 2011). Of 33 and 23 rice- and sorghum-specific gene pairs on these chromosomes, a respective 100 % and 90 % of them are in the young strata. Both members of a remarkable 50 % of the 16 duplicated RSA gene pairs are absent from sorghum, and 15 (38 %) of 39 SSA pairs are absent from rice. RSB gene pairs also have a high frequency of taxon-specific duplicated genes. Gene losses on any one of a pair of homoeologs experiencing concerted evolution may be commuted to the other, perhaps explaining the more than tenfold higher rate of gene loss in the RSA and SSA regions than the genome-wide averages of 1.8 % in rice and 3.1 % in sorghum since their divergence about 50 mya.

5.4.3 Factors Contributing to the Preservation of Homoeologous Recombination

Why has this pair of homoeologous chromosomes (R11 and R12) been so exceptional? It was proposed that the mechanics of chromosome pairing may have contributed to the patterns of intragenomic variation along the homoeologous chromosome pair. Homologous chromosome pairing in early meiotic prophase is accompanied by dynamic repositioning of chromosomes in the nucleus and formation of a

cytological structure called the telomere bouquet, i.e., chromosomes that are bundled at the telomere to form a bouquet-like arrangement (Ding et al. 2004; Bozza and Pawlowski 2008). This implies that if duplicated chromosomes preserve the telomeres and the proximal chromosomal regions, they may preserve the ability to pair and recombine with one another like homologous chromosomes. However, such illegitimate pairing near the termini may increase genomic instability due to multivalent formation and segregation disorder. Therefore, it is not unusual to find that at least one member of duplicated chromosome pairs have terminal breakages/inversions (Paterson et al. 2009a, b; The International Branchypodium Initiative 2010), which may reduce homoeologous chromosome pairing and contribute to genomic stability.

The singular evolutionary history of the above pair of grass chromosome needs further exploration. Elevated gene loss rates and elevated evolutionary rates of the preserved genes in young strata may facilitate speciation, since the loss of alternative copies of duplicated genes leads to reproductive isolation (Werth and Windham 1991; Lynch and Force 2000). The recently suggested inter-relationship between reproductive isolation and autoimmune responses (Bomblies et al. 2007; Yin et al. 2008) draws attention to the finding that orthologs R11 and S5 each contain ~25% of the NBS-LRR resistance genes (Zhou et al. 2004; Paterson et al. 2009a, b). Also, a high level of concerted evolution, associated stratification of chromosomal segments, and extensive homoeologous gene loss are each characteristics of sex chromosomes in organisms from divergent branches of the tree of life, including humans (Lahn and Page 1999), chickens (Lawson Handley et al. 2006), fungi (Charlesworth 2002), and plants (Ming and Moore Ming 2007). Moreover, unexpectedly close proximity between, and co-expansion of, NBS-LRR and several sex-determining gene analogs is found, particularly on S5 (*ts2* and two an1; three *d3* and *sk1*; another three *d3* and three *an1*; and *ts2* and twelve *d9*) (Fig. 5.5). *Ts2* (Calderon-Urrea and Dellaporta 1999), implicated in two of these four expansions, is required for a carefully regulated cell death process that forms the unisexual maize flower from an initially bisexual meristem, and is also expressed in rice floral organs that do not abort (DeLong et al. 1993). Primary maize ear pistils survive *ts2*-mediated cell death through the action of the silkless 1 (*sk1*) gene (Kim et al. 2007). *An1* (anther ear), also implicated in two NBS-LRR expansions, has mutants that permit development of perfect flowers on otherwise pistillate ears (Bensen et al. 1995). Transcription of related *An2* is stimulated by Fusarium attack although it is not known to be directly involved in defense-related metabolism, perhaps suggesting some interaction with NBS-LRR genes. The genotype with *d9* has an andromonecious (anthers in ears) dwarf phenotype (Winkler and Freeling 1994). Another hypothesis for further study is whether NBS-LRR enrichment on R11 and S5 reflects a mechanism by which genes in the candidate sex-determining region, and associated networks respectively influencing reproduction and speciation, could have some 'functional coherence' resembling that of the human Y chromosome (Lahn and Page 1999). If it proves true that CSB actually formed shortly before the 70 mya polyploidization, as albeit limited molecular clock type data suggest, then its formation may have involved divergence among homologs rather than homoeologs, yet another parallel with the evolution of heteromorphic sex chromosomes.

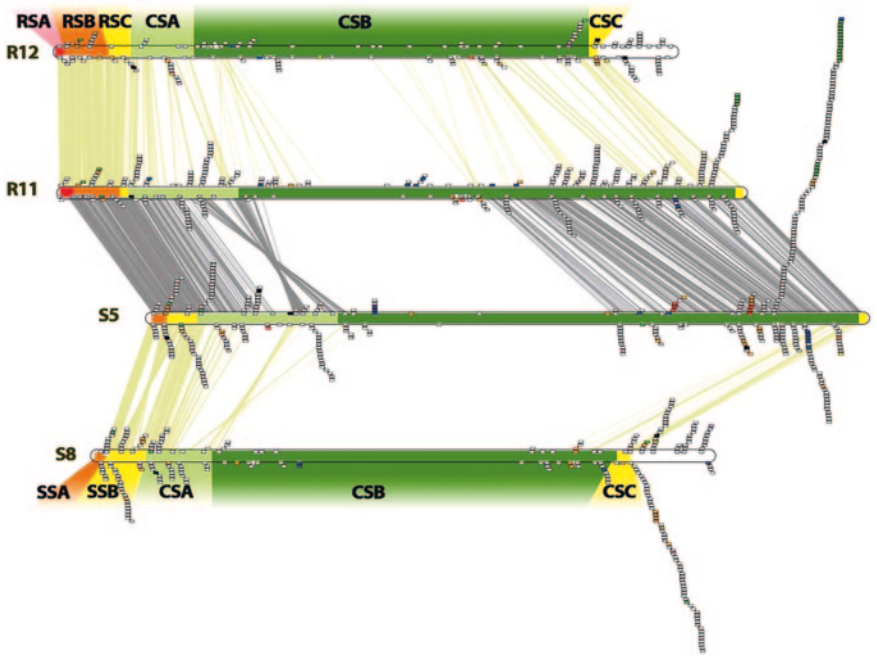

Fig. 5.5 Distribution of resistance genes on rice chromosomes 11 and 12, and their respective sorghum orthologous chromosomes 5 and 8. Chains of *boxes* show proximal locations of genes with respective transcriptional orientations *above* or *below* the chromosomes; Lines between chromosomes show gene synteny; Strata are shown in different colors, similar to Fig. 5.4 indicating their relative age. Strata produced by stepwise restriction of homoeologous recombination are shown in *different colors*

5.5 Comparative Genomics Research in Biological Pathways in Grasses

Duplicated genes have long been suggested to contribute to the evolution of new biological functions, in that redundant genes could be altered to produce new genes without disadvantage to the organism (Haldane 1932). A whole-genome duplication is a good source of duplicated genes, providing enormous opportunities for biological and genetic innovations. It was reported that duplicated genes produced by polyploidization may be adopted to form divergent pathways (Blanc and Wolfe 2004), possibly through gene functional changes soon after gene duplication (Lin et al. 2006).

While genome duplication has profoundly shaped grass genomes, processes not related to genome duplication have also been essential to grass diversity, an excellent example being the 'C4' photosynthetic pathway that is particularly abundant in grasses (Kellogg 1999). This led to the development of most productive biomass-producing crops. C4 plants are characterized by high rates of photosynthesis and efficient use of water and nitrogen, through morphological and biochemical innovations (Hatch and Slack 1966). These features are related to confer adaptation to hot, dry

environments or CO_2 deficiency (Ehleringer and Bjorkman 1978; Hattersley 1983; Seemann et al. 1987; Cerling et al. 1997). The C4 photosynthetic pathway is also an intriguing example of convergent evolution of a unique mode of CO_2 assimilation, featuring strict compartmentalization of photosynthetic enzymes into two distinct cell types, mesophyll and bundle-sheath. The C4 pathway is thought to have independently appeared at least 50 times during angiosperm evolution, including 17 times in grasses (Sage 2004; Mulhaidat et al. 2007). It has been inferred that C4 photosynthesis arose in grasses during the Oligocene epoch (24–35 mya) (Christin et al. 2008, 2011; Vicentini et al. 2008). The C4 grasses, including maize, sorghum, millets, and sugarcane; dominate most subtropical and tropical habitats (Christin et al. 2009). Multiple origins of the C4 pathway (Giussani et al. 2001; Pyankov et al. 2001) imply that its evolution may not be so complex, suggesting that there may have been some genetic pre-deposition in some C3 plants to C4 evolution (Sage 2004; Brown et al. 2011). In fact, genes encoding C4 enzyme are usually from families having multiple copies, implying that gene duplication may have potentially contributed to the establishment of the C4 pathway (Monson 2003). An ability to create and maintain large numbers of duplicated genes has been suggested to be one precondition for certain taxa to develop C4 photosynthesis (Monson 2003; Sage 2004).

Ironically, while whole-genome duplication should in principle have provided all the genes required for C4 photosynthesis to evolve, some duplicated genes were lost before becoming important for C4 photosynthesis. In these cases, independent re-duplications of these genes distributed over many millions of years were necessary for development of C4 pathway (Wang et al. 2009a). Some C4 genes (PEPC, PPCK, and NADP-ME) were recruited from duplicates of C3 photosynthetic genes produced by the polyploidization, while others (NADP-MDH, NADP-ME and PPDK-RP) were derived from tandem duplicates, which have whole-genome duplicates. C4 genes show ability for divergent duplications, with some genes seemingly duplication-philic (PEPC, NADP-ME, PPCK, and CA), and others duplication-phobic (NADP-MDH and PPDK). Many C4 genes show evidence of adaptive evolution, achieved though rapid mutations in DNA sequences, aggregated amino acid substitutions, and/or considerable increases of expression levels in specific cells. The evolution of CA genes is particularly interesting, featuring recursive tandem duplication and neighboring gene fusion, resulting in distinct isoforms of 1–3 functional units. The elongation of CA genes by recruiting extra domains may have directly contributed to the formation of more complex protein structures, as often observed in plants. In summary, some of the most important innovations in gene function may come from single-gene rather than whole-genome duplications, and may involve a long transition process.

5.6 Two Distinct Groups of Genes Exist in Grasses

5.6.1 Prominently Elevated GC Content in Grass Genes

Guanine and cytosine (GC) content of grass gene coding sequences (CDSs) shows a bi-modal distribution (Carels and Bernardi 2000), which is mainly attributed to

the third base in a codon (Fig. 5.6). GC contents at the first and second base of a codon follow a uni-modal distribution, and exhibit significant positive correlation (Shi et al. 2007). The peak values of third codon site GC content (GC3) are ~0.45 and 0.9, based on which genes can be classified as GC-poor or GC-rich. GC contents in intron (GC_{intron}) and intergenic (GC_{inter}) regions, each follow a uni-modal distribution. GC_{intron} is positively correlated with GC content in exons, but GC_{inter} is not. These observations imply co-variation of GC contents in the gene body, but not in gene regulatory regions. GC content in grasses is generally much higher than in eudicot plants, which have uni-modal GC content distribution (Carels and Bernardi 2000). A comparative analysis shows that GC contents in several monocots, including Musa, ginger, and grasses, each follow bi-modal distribution with long tails at the upper side, and grasses have higher GC content than other studied monocots (Lescot et al. 2008). These findings imply that prominent GC content elevation has occurred in grasses, and several other monocots.

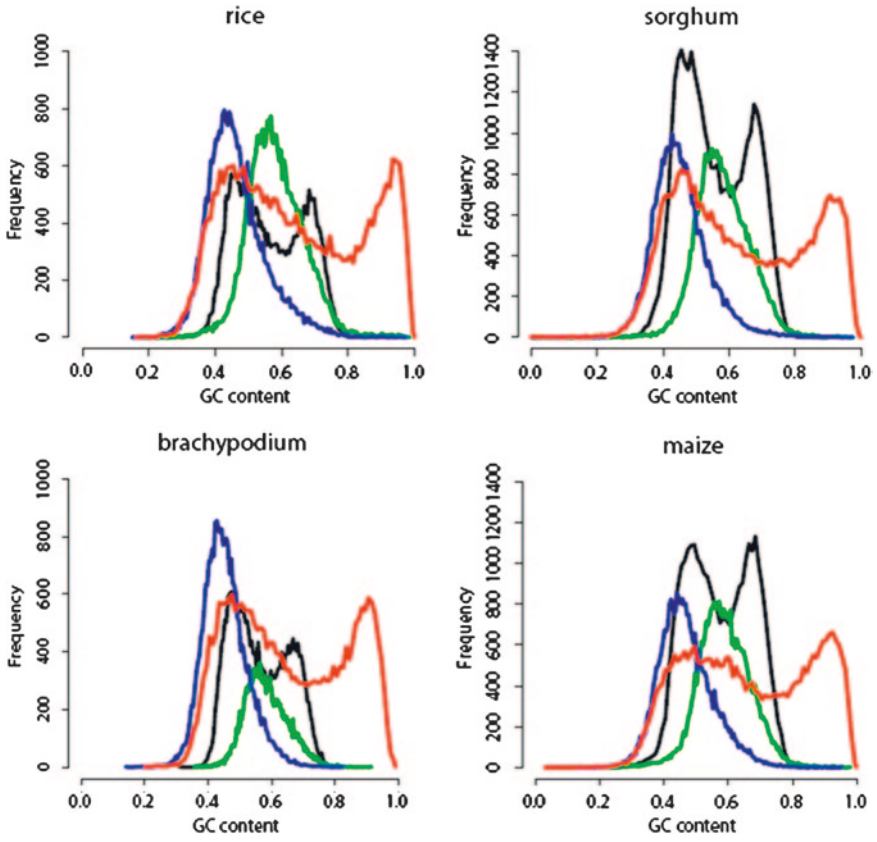

Fig. 5.6 Guanine and cytosine (*GC*) content in grass genes. *Black curves* show the *GC content* on all codon sites in genes, while *blue*, *green*, and *red curves* show *GC contents* at the first, second, and third codon sites, respectively

5.6.2 Biased Base Substitutions

By comparing duplicated genes produced by polyploidization and through inferred ancestral states of these bases in each, an excess of AT → GC substitution was observed, especially in GC-rich genes (Shi et al. 2007). In GC3-poor genes, transition (between purine bases or pyrimidine bases) frequencies were about two to threefold higher than transversion (between a purine and a pyrimidine) frequencies. Higher frequencies of transitions than transversions has been widely noted in other organisms (Brown et al. 1982; Gojobori et al. 1982). However, a different pattern of synonymous nucleotide substitution frequency bias was found in the GC3-rich genes. For substitutions at the third site of codons ending with G and C, the transversions C → G and G → C have been widely preferred, occurring about two to fivefold more frequently than the corresponding transition of C → T and G → A. For instance, the substitution frequency of transversion C → G (0.650) was more than twice the frequency of transition C → T (0.226).

5.6.3 Exploring the Cause of Elevation in GC Content

The two classes of grass genes that differ in GC content are inferred to be functionally different (Carels and Bernardi 2000), implying a possible effect of natural selection. Indeed, a distinct difference in GC content means different amino acid composition. GC-rich genes more frequently encode Glycine, Alanine, Arginine and Proline, whereas GC-poor genes more frequently encode Phenylalanine, Tyrosine, Methionine, Isoleucine, Asparagine, and Lysine (Wang et al. 2004). Instead of dividing cereal genes into two classes, it was suggested that the elevation in GC content had a negative gradient along the direction of transcription (Wong et al. 2002). They further attributed the GC elevation to transcription-related mutation bias or translation-related selection. Wang et al. (2004) analyzed rice genes and compared them with Arabidopsis genes, and proposed that mutation bias rather than natural selection was the primary cause of two classes of genes in cereals (Wang et al. 2004). Shi et al. (2007) considered many correlated factors, including the GC contents and lengths of different DNA components, exon numbers, and their correlation, and proposed that natural selection may be the cause for GC content changes. Gene conversion has also been linked by some to the elevation of GC content (Duret and Galtier 2009). It has been hypothesized that conversion may be accompanied by DNA repair of nucleotide mismatches. If the repair process were biased towards G and C (referred to biased gene conversion), an elevation in GC content would result (Galtier et al. 2001). The double-strand breaks are preferentially repaired by sister chromatids, not the homologs (Kadyk and Hartwell 1992). Theoretically, the homeologs may have even smaller chance to be taken as repair substrates due to sharing less sequence similarity. There is indirect evidence that non-allelic gene conversion can be related to GC changes in vertebrates (Galtier 2003; Kudla et al. 2004; Backstrom et al. 2005).

As to the possible contribution from homeologous gene conversion, no correlation was found between gene conversion and GC content in rice and sorghum (Wang et al. 2009b), with converted and non-converted paralogs usually having similar GC content. In spite of all these explorations, the cause for prominent elevation in grass gene GC content and the production of two distinct gene classes remains unknown.

5.7 Bioinformatics Resources and Tools for Comparative Genomics Analysis

Comparative genomics tools, not limited to the cereals, have already been developed to help exploit the valuable genomic resources accumulating at a rapid pace. To characterize gene synteny/collinearity, which is crucial to perform genome-level comparison as shown above, various software packages are available, which include the following: (1) ColinearScan [http://colinear.cbi.pku.edu.cn/; (Wang et al. 2006)], (2) i-ADHoRe [http://bioinformatics.psb.ugent.be/software/details/i–ADHoRe; (Vandepoele et al. 2002)], (3) DiagHunter [http://www.tc.umn.edu/~cann0010/Software.html; (Cannon et al. 2003)], (4) DAGchainer (Haas et al. 2003). MCSCAN (http://chibba.agtec.uga.edu/duplication/mcscan/) aims at multiple chromosome alignment to find gene collinearity in sophisticated plant genomes due to recursive polyploidizations (Tang et al. 2008b). Plant Genome Duplication Database (PGDD, http://chibba.agtec.uga.edu/duplication/) is a public database to identify and catalog plant genes in terms of intragenome or cross-genome collinearity relationships and focuses on plants with available whole genome sequences. CoGe (http://synteny.cnr.berkeley.edu/CoGe/) is a comparative genomics platform for genomes across all domains of life, helping find and compare homologous sequences (Lyons et al. 2008). Gobe (http://github.com/brentp/gobe) is a web-based tool for viewing comparative genomic data and supports viewing multiple genomic regions simultaneously, facilitating the comparison of gene content, conservation and loss in homologous regions within the same plant, or between different plants (Pedersen et al. 2011).

Finding orthologous genes is particularly important to phylogenetic research. Strategies to distinguish possible orthologs from paralogs can basically be classified into two groups: (1) phylogeny-based approaches, including Resampled Inference of Orthology [RIO: http://genome.crg.es/~talioto/OrthologySearch/; (Zmasek and Eddy 2002)] and Orthostrapper/Hierarchical grouping of Orthologous and Paralogous Sequences [HOPS: ftp://.cgb.ki.se/pub/data/HOPS/; (Storm and Sonnhammer 2003)], and (2) BLAST-based approaches, including OrthoMCL [http://www.orthomcl.org/cgi-bin/OrthoMclWeb.cgi; (Li, Stoeckert et al. 2003)], Cluster of Orthologous Groups (COG: http://www.ncbi.nlm.nih.gov/COG/new/), and In paranoid [http://inparanoid.cgb.ki.se; (O'Brien et al. 2005)]. Phylogeny-based methods typically exhibit high false negative rates, while BLAST-based methods exhibit high false positive rates (Chen et al. 2007). Gene collinearity across genomes is valuable information in finding orthologs. Based on MCSCAN outputs, a list of orthologous groups has been available on PGDD.

5.8 Summary and Outlook

The continuously declining cost and increasing speed to sequence a genome are revolutionizing many aspects of biological research, and especially inspiring vigorous explorations in comparative genomics. As one hotspot of sequencing efforts, many more genomes of cereals and their relatives will be available in the near future, which include important crops like barley and wheat. Moreover, there are a growing slate of resequencing efforts to catalog intraspecific variation of sequenced cereals (Schnable and Freeling 2011; Schnable et al. 2011). These efforts will add to our understanding of the biology of important cereals with regard to genome structural changes, gene functional innovations, speciation, and domestication, perhaps revealing underlying mechanisms and rules governing cereal diversity at the DNA sequence level. Meanwhile, more user-friendly bioinformatics tools and databases will be developed and strengthened to support many interesting comparisons.

Acknowledgments We are grateful to members in Paterson lab for useful discussion and collaboration in publishing many high-impact papers in comparative genomics. We appreciate financial support from the US National Science Foundation (MCB-1021718) and the J. S. Guggenheim Foundation to AHP, and from the China National Science Foundation (30971611, 31170212), and Hebei Natural Science Foundation distinguished young scholorship project China-Hebei New Century 100 Creative Talents Project to XW.

References

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

Backstrom N, Ceplitis H et al (2005) Gene conversion drives the evolution of HINTW, an ampliconic gene on the female-specific avian W chromosome. Mol Biol Evol 22(10):1992–1999

Baucom RS, Estill JC et al (2009) Exceptional diversity, non-random distribution, and rapid evolution of retro-elements in the B73 maize genome. PLoS Genet 5(11):e1000732

Bennetzen JL, Schmutz J et al (2012) Reference genome sequence of the model plant Setaria. Nat Biotechnol 30(6):555–561

Bensen RJ, Johal GS et al (1995) Cloning and characterization of the maize An1 gene. Plant Cell 7(1):75–84

Bevan MW, Garvin DF et al (2010) Brachypodium distachyon genomics for sustainable food and fuel production. Curr Opin Biotechnol 21(2):211–217

Blanc G, Wolfe KH (2004) Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. Plant Cell 16(7):1679–1691

Bomblies K, Lempe J et al (2007) Autoimmune response as a mechanism for a Dobzhansky-Muller-type incompatibility syndrome in plants. PLoS Biol 5(9):e236

Bowers JE, Chapman BA et al (2003) Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature 422(6930):433–438

Bowers JE, Arias MA et al (2005) Comparative physical mapping links conservation of microsynteny to chromosome structure and recombination in grasses. Proc Natl Acad Sci USA A102(37):13206–13211

Bozza CG, Pawlowski WP (2008) The cytogenetics of homologous chromosome pairing in meiosis in plants. Cytogenet Genome Res 120(3–4):313–319

Brown J, Ersland D et al (1982) Molecular aspects of storage protein synthesis during seed development. In: Khan A (ed) The physiology and biochemistry of seed development, dormancy, and germination. Elsevier Biomedical Press, Amsterdam, pp 3–42

Brown NJ, Newell CA et al (2011) Independent and parallel recruitment of preexisting mechanisms underlying C photosynthesis. Science 331(6023):1436–1439

Buell CR (2009) Poaceae genomes: going from unattainable to becoming a model clade for comparative plant genomics. Plant Physiol 149(1):111–116

Calderon-Urrea A, Dellaporta SL (1999) Cell death and cell protection genes determine the fate of pistils in maize. Development 126(3):435–441

Cannon SB, Kozik A, Chan B, Michelmore R, Young ND (2003) DiagHunter and GenoPix2D: programs for genomic comparisons, large-scale homology discovery and visualization. Genome Biology 4:R68

Carels N, Bernardi G (2000) Two classes of genes in plants. Genetics 154(4):1819–1825

Cerling TE, Harris JM et al (1997) Global vegetation change through the Miocene/Pliocene boundary. Nature 389:153–158

Charlesworth B (2002) The evolution of chromosomal sex determination. Novartis Found Symp 244:207–219 (discussion 220–204, 253–207)

Chen F, Mackey AJ et al (2007) Assessing performance of orthology detection strategies applied to eukaryotic genomes. PLoS One 2(4):e383

Chittenden LM, Schertz KF et al (1994) A detailed RFLP map of Sorghum bicolor and S. propinquum suitable for high-density mapping suggests ancestral duplication of Sorghum chromosomes or chromosomal segments. Theor Appl Genet 87:925–933

Christin PA, Osborne CP et al (2011) C4 eudicots are not younger than C4 monocots. J Exp Bot 62(9) : 3171–3181

Christin PA, Besnard G et al (2008) Oligocene CO2 decline promoted C4 photosynthesis in grasses. Curr Biol 18(1):37–43

Christin PA, Salamin N et al (2009) Integrating phylogeny into studies of C4 variation in the grasses". Plant Physiol 149(1):82–87

Datta A, Hendrix M et al (1997) Dual roles for DNA sequence identity and the mismatch repair system in the regulation of mitotic crossing-over in yeast. Proc Natl Acad Sci USA A94(18):9757–9762

DeLong A, Calderon-Urrea A et al (1993) Sex determination gene TASSELSEED2 of maize encodes a short-chain alcohol dehydrogenase required for stage-specific floral organ abortion. Cell 74(4):757–768

Devos KM (2010) Grass genome organization and evolution. Curr Opin Plant Biol 13(2):139–145

Devos KM, Pittaway TS et al (2000) Comparative mapping reveals a complex relationship between the pearl millet genome and those of foxtail millet and rice. Theoret Appl Genet 100(2):190–198

Ding DQ, Yamamoto A et al (2004) Dynamics of homologous chromosome pairing during meiotic prophase in fission yeast. Dev Cell 6(3):329–341

Duret L, Galtier N (2009) Biased gene conversion and the evolution of mammalian genomic landscapes. Annu Rev Genomics Hum Genet 10:285–311

Ehleringer JR, Bjorkman O (1978) A Comparison of Photosynthetic Characteristics of Encelia Species Possessing Glabrous and Pubescent Leaves. Plant Physiol 62(2):185–190

Feuillet C, Keller B (1999) High gene density is conserved at syntenic loci of small and large grass genomes. Proc Natl Acad Sci USA 96(14):8265–8270

Freeling M (2001) Grasses as a single genetic system: reassessment 2001. Plant Physiol 125(3):1191–1197

Galtier N (2003) Gene conversion drives GC content evolution in mammalian histones. Trends Genet 19(2):65–68

Galtier N, Piganeau G et al (2001) GC-content evolution in mammalian genomes: the biased gene conversion hypothesis. Genetics 159(2):907–911

Gaut BS, Doebley JF (1997) DNA sequence evidence for the segmental allotetraploid origin of maize. Proc Natl Acad Sci USA 94(13):6809–6814

Giussani LM, Cota-Sanchez JH et al (2001) A molecular phylogeny of the grass subfamily Panicoideae (Poaceae) shows multiple origins of C4 photosynthesis. Am J Bot 88(11):1993–2012

Global Perspective Studies Unit, F a A O o t U N (2006) FAQ: World Agriculture: towards 2030/2050. Interim Report, Rome, Italy

Goff SA, Ricke D et al (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science 296(5565):92–100

Gojobori T, Li WH et al (1982) Patterns of nucleotide substitution in pseudogenes and functional genes. J Mol Evol 18(5):360–369

Haas BJ, Delcher AL et al (2003) Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res 31(19):5654–5666

Haldane JBS (1932) The causes of evolution. Cornell University Press, Ithaca

Hatch MD, Slack CR (1966) Photosynthesis by sugar-cane leaves. A new carboxylation reaction and the pathway of sugar formation. Biochem J 101(1):103–111

Hattersley PG (1983) The distribution of C3 and C4 grasses in Australia in relation to climate. Oecologia 57:113–128

Hilu KW (2004) Phylogenetics and chromosomal evolution in the Poaceae (grasses). Aust J Bot 52:10

The International Brachypodium Initiative (2010) Genome sequencing and analysis of the model grass Brachypodium distachyon. Nature 463(7282):763–768

International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. Nature 436(7052):793–800

Jaillon O, Aury JM et al (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449(7161):463–467

Kadyk LC, Hartwell LH (1992) Sister chromatids are preferred over homologs as substrates for recombinational repair in Saccharomyces cerevisiae. Genetics 132(2):387–402

Kellogg EA (1999) Phylogenetic aspects of the evolution of C4 photosynthesis. In: Sage RF, Monson RK (eds) C4 plant biology. Academic Press, San Diego, CA, pp 411–444

Kellogg EA (2001) Evolutionary history of the grasses. Plant Physiol 125(3):1198–1205

Khakhlova O, Bock R (2006) Elimination of deleterious mutations in plastid genomes by gene conversion. Plant J 46(1):85–94

Kim JC, Laparra H et al (2007) Cell cycle arrest of stamen initials in maize sex determination. Genetics 177(4):2547–2551

Kishimoto NHH, Abe K, Arai S, Saito A, Higo K (1994) Identification of the duplicated segments in rice chromosomes 1 and 5 by linkage analysis of cDNA markers of known functions. Theor Appl Genet 88:722–726

Kudla G, Helwak A et al (2004) Gene conversion and GC-content evolution in mammalian Hsp70. Mol Biol Evol 21(7):1438–1444

Lahn BT, Page DC (1999) Four evolutionary strata on the human X chromosome. Science 286(5441):964–967

Lawson Handley LJ, Hammond RL et al (2006) Low Y chromosome variation in Saudi-Arabian hamadryas baboons (Papio hamadryas hamadryas). Heredity 96(4):298–303

Lescot M, Piffanelli P et al (2008) Insights into the Musa genome: syntenic relationships to rice and between Musa species. BMC Genomics 9:58

Li L, Stoeckert CJ et al (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res 13:2178–2189

Lin Y, Byrnes JK et al (2006) Codon usage bias versus gene conversion in the evolution of yeast duplicate genes. Proc Natl Acad Sci USA 103:14412–14416

Lohithaswa HC, Feltus FA et al (2007) Leveraging the rice genome sequence for comparative genomics in monocots. Theor Appl Genet 115:237–243

Lynch M, Force AG (2000) The origin of interspecific genomic incompatibility via gene duplication. Am Nat 156(6):590–605

Lyons E, Pedersen B et al (2008) Finding and comparing syntenic regions among arabidopsis and the outgroups papaya, poplar, and grape: CoGe with rosids. Plant Physiol 148(4):1772–1781

Mayer KF, Martis M et al (2011) Unlocking the barley genome by chromosomal and comparative genomics. Plant Cell 23(4):1249–1263
Ming R, Moore PH (2007) Genomics of sex chromosomes. Curr Opin Plant Biol 10(2):123–130
Monson RK (2003) Gene duplication, neofunctionalization, and the evolution of C4 photosynthesis. Int J Plant Sci 164(6920):S43–S54
Mulhaidat R, Sage RF et al (2007) Diversity of kranz anatomy and biochemistry in C4 eudicots. Am J Bot 94(3):20
Murat F, Xu JH et al (2010) Ancestral grass karyotype reconstruction unravels new mechanisms of genome shuffling as a source of plant evolution. Genome Res 20(11):1545–1557
Nagamura YIT, Antonio B, Shimano T, Kajiya H, Shomura A, Lin S, Kuboki Y, Kurata N et al (1995) Conservation of duplicated segments between rice chromosomes 11 and 12. Breed Sci 45:373–376
O'Brien KP, Remm M et al (2005) In paranoid: a comprehensive database of eukaryotic orthologs. Nucleic Acids Res 33:D476–D480 (database issue)
Paterson AH (2005) Polyploidy, evolutionary opportunity and crop adaptation. Genetica 123(1–2):191–196
Paterson AH (2008) Paleopolyploidy and its Impact on the Structure and Function of Modern Plant Genomes. Genome Dyn 4:1–12
Paterson AH, Bowers JE et al (2004) Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. Proc Natl Acad Sci USA 101(26):9903–9908
Paterson AH, Chapman BA et al (2006) Convergent retention or loss of gene/domain families following independent whole-genome duplication events in Arabidopsis, Oryza, Saccharomyces, and Tetraodon. Trends Genet 22:597–602
Paterson AH, Bowers JE et al (2009a) The Sorghum bicolor genome and the diversification of grasses. Nature 457(7229):551–556
Paterson AH, Bowers JE et al (2009b) Comparative genomics of grasses promises a bountiful harvest. Plant Physiol 149(1):125–131
Pedersen BS, Tang H et al (2011) Gobe: an interactive, web-based tool for comparative genomic visualization. Bioinformatics 27(7):1015–1016
Puchta H, Dujon B et al (1996) Two different but related mechanisms are used in plants for the repair of genomic double-strand breaks by homologous recombination. Proc Natl Acad Sci USA 93(10):5055–5060
Pyankov VI, Artyusheva EG et al (2001) Phylogenetic analysis of tribe Salsoleae (Chenopodiaceae) based on ribosomal ITS sequences: implications for the evolution of photosynthesis types. Am J Bot 88(7):1189–1198
Sage RF (2004) The evolution of C4 photosynthesis. New Phytol 161:341–370
Salse J, Abrouk M et al (2009) Reconstruction of monocotelydoneous proto-chromosomes reveals faster evolution in plants than in animals. Proc Natl Acad Sci USA 106(35):14908–14913
Schnable JC, Freeling M (2011) Genes identified by visible mutant phenotypes show increased bias toward one of two subgenomes of maize. PLoS One 6(3):e17855
Schnable PS, Ware D et al (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326(5956):1112–1115
Schnable JC, Springer NM et al (2011) Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. Proc Natl Acad Sci USA 108(10):4069–4074
Seemann JR, Sharkey TD et al (1987) Environmental effects on photosynthesis, nitrogen-use efficiency, and metabolite pools in leaves of sun and shade plants. Plant Physiol 84(3):796–802
Shi X, Wang X et al (2007) Evidence that natural selection is the primary cause of the GC content variation in rice genes. J Integr Plant Biol 49(9):1393–1399
Singh NK, Dalal V et al (2007) Single-copy genes define a conserved order between rice and wheat for understanding differences caused by duplication, deletion, and transposition of genes. Funct Integr Genomics 7(1):17–35

Soderstrom TR, Hilu KW, Campbell CS, Barkworth MA (1987) Grass systematics and evolution. Smithsonian Institution Press, Washington, DC

Soltis PS (2005) Ancient and recent polyploidy in angiosperms. New Phytol 166(1):5–8

Storm CE, Sonnhammer EL (2003) Comprehensive analysis of orthologous protein domains using the HOPS database. Genome Res 13(10):2353–2362

Swigonova ZJ, Lai S et al (2004b) Close split of sorghum and maize genome progenitors. Genome Res 14(10A):1916–1923

Swigonova Z, Lai JS et al (2004a) On the tetraploid origin of the maize genome. Compa Funct Genomics 5(3):281–284

Tang HB, Wang XY et al (2008b) Unraveling ancient hexaploidy through multiply aligned angiosperm gene maps. Genome Res 18(12):1944–1954

Tang H, Bowers JE et al (2008a) Synteny and colinearity in plant genomes. Science 320:486–488

Tang H, Bowers JE et al (2010) Angiosperm genome comparisons reveal early polyploidy in the monocot lineage. Proc Natl Acad Sci USA 107(1):472–477

The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408(6814):796–815

The Rice Chromosomes 11 and 12 Sequencing Consortia (2005) The sequence of rice chromosomes 11 and 12, rice in disease resistance genes and recent gene duplications. BMC Biol 3:20

Thomas BC, Pedersen B et al (2006) Following tetraploidy in an Arabidopsis ancestor, genes were removed preferentially from one homology leaving clusters enriched in dose-sensitive genes. Genome Res 16(7):934–946

Van de Peer Y (2004) Computational approaches to unveiling ancient genome duplications. Nat Rev Genet 5(10):752–763

Vandepoele K, Saeys Y et al (2002) The automatic detection of homologous regions (ADHoRe) and its application to microcolinearity between *Arabidopsis* and rice. Genome Res 12(11):1792–1801

Vandepoele K, Simillion C et al (2003) Evidence that rice and other cereals are ancient aneuploids. Plant Cell 15(9):2192–2202

Vicentini A, Barber JC et al (2008) The age of the grasses and clusters of origins of C4 photosynthesis. Glob Change Biol 14:15

Wang HC, Singer GA et al (2004) Mutational bias affects protein evolution in flowering plants. Mol Biol Evol 21(1):90–96

Wang X, Shi X et al (2005) Duplication and DNA segmental loss in the rice genome: implications for diploidization. New Phytol 165(3):937–946

Wang X, Shi X et al (2006) Statistical inference of chromosomal homology based on gene colinearity and applications to Arabidopsis and rice. BMC Bioinf 7(1):447

Wang X, Tang H et al (2007) Extensive concerted evolution of rice paralogs and the road to regaining independence. Genetics 177(3):1753–1763

Wang X, Gowik U et al (2009a) Comparative genomic analysis of C4 photosynthetic pathway evolution in grasses. Genome Biol 10(6):R68

Wang X, Tang H et al (2009b) Comparative inference of illegitimate recombination between rice and sorghum duplicated genes produced by polyploidization. Genome Res 19(6):1026–1032

Wang X, Tang H et al (2011) Seventy million years of concerted evolution of a homoeologous chromosome pair, in parallel, in major Poaceae lineages. Plant Cell 23(1):27–37

Watson L, Dallwitz MJ (1992) The grass genera of the world. CAB International, Wallingford

Werth CR, Windham MD (1991) A model for divergent, allopatric speciation of polyploid pteridophytes resulting from silencing of duplicate-gene expression. Am Nat 137(4):515–526

Wicker T, Mayer KF et al (2011) frequent gene movement and pseudogene evolution is common to the large and complex genomes of wheat, barley, and their relatives. Plant Cell 23(5): 1706–1718

Winkler RG, Freeling M (1994) Physiological genetics of the dominant gibberellin-nonresponsive maize dwarfs, dwarf-8 and dwarf-9. Planta 193(3):341–348

Wong GK, Wang J et al (2002) Compositional gradients in Gramineae genes. Genome Res 12(6):851–856

Woodhouse MR, Schnable JC et al (2010) Following tetraploidy in maize, a short deletion mechanism removed genes preferentially from one of the two homologs. PLoS Biol 8(6):e1000409

Yin T, Difazio SP et al (2008) Genome structure and emerging evidence of an incipient sex chromosome in Populus. Genome Res 18(3):422–430

Youens-Clark K, Buckler E et al (2011) Gramene database in 2010: updates and extensions. Nucleic Acids Res 39:D1085–D1094 (database issue)

Yu J, Hu S et al (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science 296(5565):79–92

Yu J, Wang J et al (2005) The genomes of Oryza sativa: a history of duplications. Plos Biol 3(2):266–281

Zhang G, Liu X et al (2012) Genome sequence of foxtail millet (Setaria italica) provides insights into grass evolution and biofuel potential. Nat Biotechnol 30(6):549–554

Zhou T, Wang Y et al (2004) Genome-wide identification of NBS genes in japonica rice reveals significant expansion of divergent non-TIR NBS-LRR genes. Mol Genet Genomics 271(4):402–415

Zmasek CM, Eddy SR (2002) RIO: analyzing proteomes by automated phylogenomics using resampled inference of orthologs. BMC Bioinf 3:14

Chapter 6
Transposons in Cereals: Shaping Genomes and Driving Their Evolution

Jan P. Buchmann, Beat Keller and Thomas Wicker

6.1 Introduction

A molecular understanding of genome evolution depends on the availability of the complete DNA sequence of an organism, ideally in the form of a complete sequence or at least as good datasets of partial sequences. Thus, evolutionary genomics and specifically our understanding of the role of transposons in the evolution of (cereal) genomes have been intimately linked to the progress in whole genome analysis. The first completely sequenced plant genome was from the model plant *Arabidopsis thaliana* with a size of ~120 Mbp (AGI 2000). Two years later, the complete sequences of the first grass genomes were published: *Oryza sativa* L ssp. *japonica* and *Oryza sativa* L ssp. *indica* (Goff et al. 2002; Yu et al. 2002). These first genomes were sequenced by the "BAC-by-BAC" approach: constructing a BAC library of the genomes, fingerprinting the BAC clones and assembling them into a minimum tiling path which was then sequenced by shotgun-sequencing. This approach creates a high quality sequence which is ordered along the chromosomes, but is very laborious and expensive.

The rapid development in the field of DNA sequencing technology has resulted in faster and cheaper methods, allowing to sequence and assemble *de novo* entire genomes using a whole-genome shotgun (WGS) approach. This approach was used to sequence the genomes of the two grass species Sorghum and *Brachypodium*, as well as soybean (Paterson et al. 2009; IBI 2010; Schmutz et al. 2010). However, WGS sequencing has limits in cases where size and complexity of the analyzed genomes are large, e.g. the sequencing of the large and repetitive genomes from barley or wheat will critically depend on anchoring

J. P. Buchmann · B. Keller (✉) · T. Wicker
Institute of Plant Biology, University of Zurich, Zollikerstrasse 107, 8008 Zürich, Switzerland
e-mail: bkeller@botinst.uzh.ch

P. K. Gupta and R. K. Varshney (eds.), *Cereal Genomics II*,
DOI: 10.1007/978-94-007-6401-9_6,

shotgun sequences or individual BACs to genetic maps. The largest plant genome sequenced so far is the 2,300 Mbp genome of maize (Schnable et al. 2009).

The goal of each sequencing project is to retrieve the so called "pseudomolecules" which represent the chromosomes of the sequenced organism. However, the obtained sequence will not be one complete molecule per chromosome, whatever sequencing strategy and method is used. Problematic sequences which introduce gaps in genomic sequences are usually highly repetitive regions, e.g. centromeric regions or ribosomal DNA clusters. Therefore, genome sequences are gradually and continuously improved after the first release and even high quality genomes like those from rice and *Arabidopsis*, which are now in the sixth and ninth release, still contain gaps (Table 6.1). Not all plant genomes have yet reached such high quality standards. For example, the genomes of *Physcomitrella patens* (Rensing et al. 2008), a moss plant species, poplar (Tuskan et al. 2006) or grapevine (Jaillon et al. 2007) went through only one initial round of shotgun sequencing and the resulting assemblies consist of "supercontigs" or "scaffolds", which in many cases have not yet been assigned to specific chromosomes and contain thousands of sequence gaps (Table 6.1). Newer sequencing techniques, e.g. 454 and Illumina, create an enormous amount of sequence reads in every run and those assemblies are still a great challenge for existing software. In addition, these low-cost and faster sequencing techniques led to a rapid growth in the number of available genomes which show different levels of completeness. At the end of 2011, there were 25 genome projects listed on www.phytozome.org, of which five are from the grass family (Mayer et al. 2011; Berkman et al. 2011).

Chain et al. (2009) proposed a classification system, where the genomic sequences are classified into five categories depending upon technology used and the quality of the assembly. The categories range from the lowest *Standard Draft* (category 1), which represents a basic automated assembly of raw sequences to the highest *Finished* (category 5) with no gaps and less than 1 sequence error in 100 kb. Some microbial genomes have reached the *Finished* status, while plant genomes are found between the categories 1 through 4.

Table 6.1 Numbers of scaffolds and gaps in a selection of publicly available plant genomes

Organism	Size (Mbp)	Version	Scaffolds[a]	Gaps[b]	Gaps/Mbp	References
Arabidopsis	119	9	5	96	0.8	AGI (2000)
Brachypodium	271	1	5	1,625	5.9	IBI (2010)
Rice	372	6	12	203	0.5	IRGSP (2005)
Physcomitrella	462	1.6	506	14,910	32.2	Rensing et al. (2008)
Poplar	405	2	236	13,341	32.9	Tuskan et al. (2006)
Sorghum	659	1	10	6,907	10.4	Paterson et al. (2009)
Grapevine	342	1	32	165,717	25.7	Jaillon et al. (2007)
Maize	2,066	5b.60	11	125,338	60.8	Schnable et al. (2009)
Wheat 3B	14	1	10	282	19.8	Choulet et al. (2010)

[a]Scaffolds or supercontigs larger than 100 kb
[b]Total number of gaps longer than 5 bp

The availability of complete sequences from several grass genomes has allowed genome-wide studies on genome structure. Gene content can be compared and analyzed and the repetitive part of the genomes can be analyzed in great detail. This has allowed us to get a deep insight into the structure of grass genomes. In addition, this knowledge about genomes has allowed us to develop approaches for comparative and evolutionary genomics. This has resulted in new insight into the evolution of plant genomes, both concerning genes and well as repetitive elements. For example, the comparison of the *Brachypodium* and rice genomes revealed a model for chromosome fusion which could explain variation in chromosome numbers in the family of grasses (Paterson et al. 2009; IBI 2010). Thus, comparative and evolutionary analysis of genomes can result in new and exciting insights into genome structure and its evolution. In this chapter, we will focus on the role of transposons or repetitive elements on genome structure and evolution, a particularly active research field which is profiting from genome-wide analysis.

6.2 Comparative and Evolutionary Genomics in Grasses: The Early Studies

In the mid 1980s, restriction fragment length polymorphism (RFLP) markers were developed for applications in plant breeding and genetic research (Gale and Devos 1998). This resulted in the first genetic maps of cereal crop species. The potential of RFLP probes to hybridize to highly similar, but not perfectly identical sequences and lack of abundance of available markers at that time, stimulated the use of probes from one species for genetic studies in related species. Colinearity across genomes was first reported in the late 1980s between tomato and potato (Bonierbale et al. 1988) and between the three diploid genomes of hexaploid wheat (Chao et al. 1988). Soon after, RFLP-based genetic maps were developed for homoeologous chromosomes of group 7 of bread wheat (*Triticum aestivum*), revealing a high colinearity of marker order between them (Chao et al. 1989). This early work was followed by a number of studies using RFLP markers to establish complete maps of the wheat genome. The first consensus map in the grasses, known today as the 'crop circle', was published in 1995 by Moore et al. (1995a, b), providing the foundation of much of the later research, elaborating and refining the concept and establishing the grasses as a single genetic system. These early studies also revealed some rearrangements between similar genomes, starting the highly productive field of evolutionary genomics. In one such study, cross–hybridization of RFLP markers derived from bread wheat with rye (*Secale cereale*) and barley revealed evidence for a few translocations of chromosome arms in the rye genome if compared to the wheat genomes, while most probes showed that the order of the loci was conserved between those three species (Devos et al. 1993; Moore et al. 1995b).

Investigating the genomic relationships of wheat in maize and rice, Moore et al. (1995a) showed that, despite the divergence of those species ~60–70 million years

ago (MYA) and their massive differences in genome size, the gene order was still conserved along large stretches of the chromosomes. Assuming that the colinearity between rice and wheat is preserved, the genetic map of rice, the smallest grass genome known at that time, was divided into linkage groups and aligned against the genetic maps of wheat and maize. Indeed, it was possible to reconstruct the wheat and maize genome with the rice linkage groups (Moore et al. 1995a). This approach was extended to sugarcane and foxtail millet and led to the first version of the crop circle mentioned above, which has been updated and expanded later (Devos 2005; Salse and Feuillet 2011). The crop circle indicates that the grasses diverged from a common ancestor and that the gene order seems to be well conserved during evolution even after millions of years, despite chromosomal reorganization and remarkable changes in genome sizes.

However, due to the use of only relatively few DNA probes, the genetic resolution of the original crop circle was quite low and did not necessarily reflect the situation at the genomic level. The advances in sequence technology and the subsequent drop in costs created a vast amount of sequence information which offered a unique opportunity to investigate colinearity at the molecular level. In fact, already the first studies of genomic colinearity at the sequence level revealed various exceptions, demonstrating that genes were not always found at the expected position and, therefore, the hypothesis of gene movement was formulated (Gallego et al. 1998; Guyot et al. 2004).

The further comparative analysis of grass genomes revealed many surprising insights into genome evolution. For example, it was found the intergenic regions diverge completely within a few million years. Only in case of very recent evolutionary divergence, both genes and intergenic regions are still conserved. The finding that the intergenic space is changing at a faster pace than the genic space can easily be explained by the lower evolutionary pressure for conservation compared to the genes (Petrov 2001). Therefore, insertions by transposable elements (TE) or deletions caused by illegitimate recombination or unequal crossing over drive the fast turnover of intergenic sequences (Devos et al. 2002; Wicker et al. 2003). These first discoveries laid the foundation for much of the later work in genome-wide analysis described below.

6.3 Discovery of Transposons in Plants: Selfish DNA and Beyond

The research on transposable elements (TEs) in plant genomes has two different historical origins, each with independent lines of research. Their findings converged only relatively recently. First, in her pioneering work Barbara McClintock discovered the existence of jumping genes based on the careful analysis of several biological phenomena observed in maize genetic studies. B. McClintock suggested that genetic factors can move in the genome, thereby modifying gene expression and contributing to genome evolution (McClintock 1950 and summarized

in McClintock 1984). At that time, the molecular basis of the proposed mobile genetic elements remained unknown. It was later found that DNA transposons caused the observed effects (e.g. the Activator (*Ac*)/Dissociator (*Ds*) elements) (Fedoroff et al. 1983). Second, in an independent line of research, the analysis of complete genomes resulted in the discovery that some plant genomes have very high contents of repetitive DNA. Mostly, such studies were done in the 1970s and 1980s by DNA reassociation experiments (Cot analysis) which are based on DNA hybridization (Britten et al. 1974). The observation of a rapidly annealing fraction of genomic DNA suggested that many plant genomes are highly repetitive (Flavell et al. 1974). Only later it was found that this highly repetitive part of the genome consists mostly of transposon and retrotransposon DNA. Based on the history of this discovery, transposons are frequently also called repetitive elements.

Transposons were identified in many different organisms and in cases such as the *P* and *I* elements in Drosophila, they were found to have dramatic negative effects for the survival of the organism under certain conditions. In general, transposons did not have obvious adaptive value and were described as "selfish" DNA based on their ability to multiply (Doolittle and Sapienza 1980). Indeed, TEs are small genetic units, actual "minimal genomes", which contain exactly enough information to be able to replicate, move around in the genome, or both. They use the DNA replication and translation machinery of their "host" and thrive within the environment of the genome. In their paper, Doolittle and Sapienza (1980) made an argument for the hypothesis that the only function of transposable elements is the survival in the genome. However, they explicitly included the possibility that this "raw material" can have some adaptive value later in evolution. The concept of "selfish" DNA was rapidly adapted by the community. In addition, the term "junk DNA" was introduced, reflecting the idea that what is present in such enormous amounts in the genome without obvious consequences on the phenotype must be useless junk. Of course, this contrasts with the original findings of biological function of transposons and the hypothesis of B. McClintock that transposable element contribute to evolution by stimulating chromosomal and genomic rearrangements, resulting in new configurations of genes and changes in gene expression. Interestingly, and a main topic in this chapter below, it was recently found that transposons are major factors in moving around genes in genomes (Wicker et al. 2010). This is a highly relevant finding for understanding the evolution of plant genomes and fits perfectly with the earlier arguments of McClintock. Thus, the jury is out on the final decision on the role of transposable elements in evolution, i.e. selfish DNA verses adaptive value, and there are some arguments for both hypotheses.

6.4 Genome Size of Plants: Genes and Repetitive Elements

In the 1970s, it was found that eukaryotic genomes show an extreme variation in size (Bennett and Smith 1976). Some studies reported an over 200,000-fold variation in genome size, namely between the Amoeba *Amoeba dubia* which was found to have a genome size of 670,000 Mbp (Gregory 2001) and the 2.9 Mbp genome

of the microsporidium *Encephalitozoon cuniculi* (Biderre et al. 1995; Katinka et al. 2001). Plant genomes in particular show a vast variation in genome sizes, even between very closely related species. Most interestingly, there is almost no correlation between genome size and phylogenetic distance in plants (Fig. 6.1). Among the dicotyledonous plants, A*rabidopsis* has one of the smallest genomes known with only about 120 Mbp (AGI 2000). In contrast, the closely related *Brassica* species which diverged from *Arabidopsis* only 15–20 MYA(Yang et al. 1999) have 5–10 times larger genomes. In monocotyledonous plants, variation is even more extreme: The grasses *Brachypodium dystachion*, rice and sorghum have genome sizes of 273, 389 and 690 Mbp, respectively, considerably larger than the *Arabidopsis* genome but roughly an order of magnitude smaller than the genomes of some agriculturally important grass species such as diploid wheat or maize with haploid genome sizes of 5,700 and 2,500 Mbp, respectively. And even they are still dwarfed by the genomes of some lilies, among them *Fritillaria uva-vulpis* which has a genome size of more than 87,000 Mbp, over 700 times the size of the *Arabidopsis* genome (Leitch et al. 2007). Also among *Dicotyledons*, closely related species often differ dramatically in their genome sizes. Maize and sorghum, for example, diverged only about 12 MYA (Swigonová et al. 2004), but the maize genome is more than 4 times the size of the sorghum genome (Tables 6.1 and 6.2).

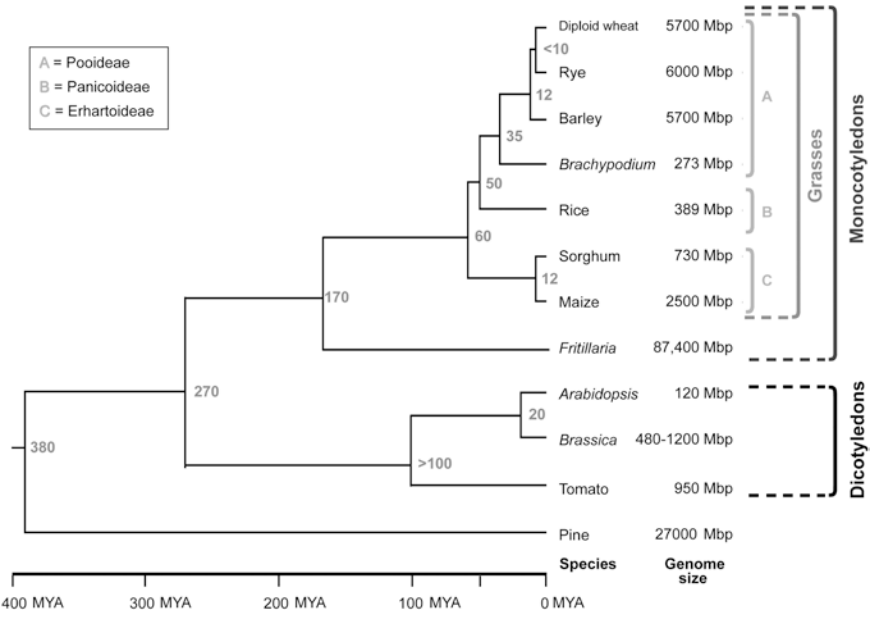

Fig. 6.1 Phylogenetic relationships and genome sizes in selected plant species. Divergence times of specific clades are indicated in *grey* numbers next to the corresponding branching. These numbers are averages of the published values provided in Table 6.1. The scale at the *bottom* indicates divergence times in million years ago (*MYA*). Major taxonomic groups that are discussed in the text are indicated at the *left*

Table 6.2 Plant genome sizes and gene numbers in a selection of publicly available genomes

Plant genomes	Size (Mbp)	Genes	Reference
Arabidopsis thaliana	120	26,200	AGI (2000)
Brachypodium distachyon	273	25,500	IBI (2010)
Fritillaria uva-vulpis	87,400	?	Leitch et al. (2007)
Hordeum vulgare	5,700	32,000	Mayer et al. (2011)
Oryza sativa	372	40,600	IRGSP (2005)
Physcomitrella patens	462	35,900	Rensing et al. (2008)
Populus trichocarpa	410	45,500	Tuskan et al. (2006)
Sorghum bicolor	659	34,500	Paterson et al. (2009)
Triticum aestivum	16,000	50,000	Choulet et al. (2010)
Vitis vinifera	342	30,400	Jaillon et al. (2007)
Zea mays	2,061	30,000	Schnable et al. (2009)

Abbreviations in references: *AGI Arabidopsis* Genome Initiative, *CSC C. elegans* Sequencing Consortium, *IBI* International *Brachypodium* Initiative, *ICGSC* International Chicken Genome Sequencing Consortium, *IHGSC* International Human Genome Sequencing Consortium, *IRGSP* International Rice Genome Sequencing Consortium, *MGSC* Mouse Genome Sequencing Consortium

Despite the vast differences in genomes sizes among plants, the number of genes is almost similar in all species investigated so far. In fact, in recent years, a consensus began to transpire that probably all angiosperm plants contain between 25,000 and 30,000 genes per haploid genome equivalent. This includes only protein-coding genes and excludes other components of gene space such as the highly repetitive ribosomal DNA clusters, tRNAs and small nucleolar and small interfering RNAs as well as conserved non-coding sequences (Freeling and Subramaniam 2009). However, the discussion about the actual gene number of plant genomes is far from over because of the technical difficulties of reliably predicting genes and the mere challenge of defining what a gene actually is.

6.5 Transposable Elements Determine Genome Size

The differences in genome sizes are caused by variation in the number and size of TEs. Especially in large genomes like barley, wheat or maize, TE contribute at least 80 % to the total genomic DNA (Schnable et al. 2009; Wicker et al. 2009b). Already early on, it became clear that there must be hundreds or even thousands of different TE families populating these large genomes (SanMiguel et al. 1998; Wicker et al. 2001). Thus, it has become an important research area to categorise and characterise at least the most abundant TE families in the different plant species. This is necessary for two practical reasons: first, TEs display such an enormous variety that some more exotic ones are often mistaken for genes and annotated as such and, second, transposable elements can cause problems during

sequencing, especially in large genomes. Good knowledge of TEs can therefore help order sequence fragments and close sequence gaps.

Although the necessity and practical value of databases of repetitive elements are apparent to researchers, in recent years, classification and characterisation of these repeats was done very much on a species-by-species level and no common guidelines and classification systems were ever consistently applied. In 2002, Jorge Dubcovsky (UC Davis, CA, USA), David Matthews (Cornell University, NY, USA) and Thomas Wicker (University of Zurich, Switzerland) initiated the first database for TE sequences from *Triticeae* (TREP, *Triticeae* repeat database). TREP originally included only sequences from wheat and barley (Wicker et al. 2002), but has been expanded to include other species since then. The 11th release of TREP contained over 1,500 DNA sequences of plant TEs plus 291 predicted TE protein sequences. Databases for TEs from *A. thaliana* (tigr.org) and rice (retroryza.org) have also become publicly available later.

In 2007, a group of TE experts met at the Plant and Animal Genome Conference in San Diego (California, USA) with the goal to define a broad consensus for the classification of all eukaryotic transposable elements. This included the definition of consistent criteria in the characterisation of the main superfamilies and families and a proposal for a naming system (Wicker et al. 2007a). The proposed system is a consensus of a previous TE classification system that groups all TEs into two major classes, 9 orders and 29 superfamilies (Fig. 6.2). Class 1 contains all TEs which replicate via an mRNA intermediate in a "copy-and-paste" process, while in Class 2 elements, the DNA itself is moved analogous to a "cut-and-paste" process. One novel aspect of the classification system is that the TE family name should be preceded by a three-letter code for class, order and superfamily (Fig. 6.2). This allows to immediately recognise the classification when seeing the name of TE. The proposed classification system is open to expansion as new types of TEs might still be identified in the future.

6.6 TE-Driven Genome Expansion

The most abundant TE class in plant genomes are long terminal repeat (LTR) retrotransposons. They replicate via an mRNA intermediate which is reverse transcribed and integrated elsewhere in the genome. Thus, each replication cycle creates a new copy of the element. Most of the probably hundreds of LTR retrotransposon families in a genome are present in low or moderate copy numbers. However, especially the large plant genomes contain retrotransposon families which are extremely successful colonisers. For example, *BARE1* elements contribute more than 10 % to the barley genome (Vicient et al. 1999; Kalendar et al. 2000; Soleimani et al. 2006). A whole genome survey in barley showed that 50 % of its genome is made up by only 14 TE families and 12 of them are LTR retrotransposons (Wicker et al. 2009b). It is not known what makes certain LTR retrotransposon families particularly successful. Some retrotransposons have been shown to be activated by stress conditions

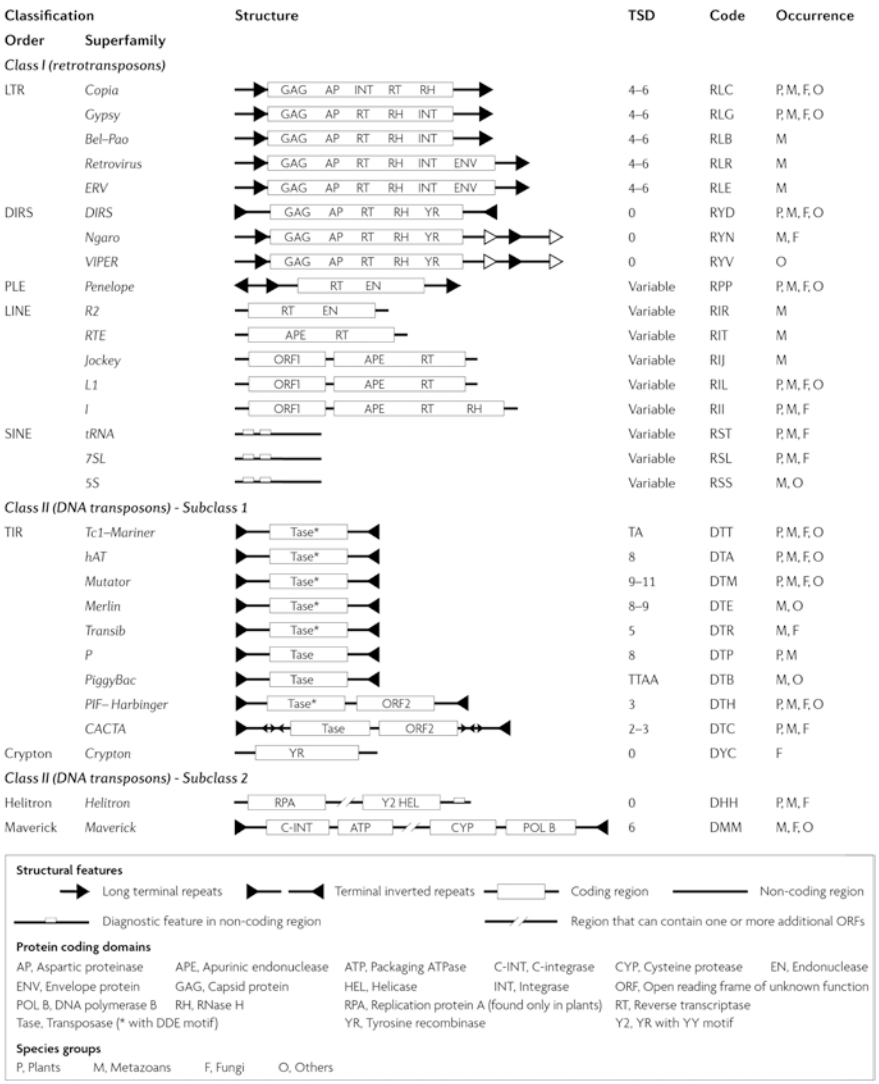

Fig. 6.2 Classification system for transposable elements (Wicker et al. 2007a). The classification is hierarchical and divides TEs into two main classes on the basis of the presence or absence of RNA as a transposition intermediate. They are further subdivided into subclasses, orders and superfamilies. The size of the target site duplication (*TSD*), which is characteristic for most superfamilies, can be used as a diagnostic feature. To facilitate identification, we propose a three-letter code that describes all major groups and that is added to the family name of each TE. *DIRS* Dictyostelium intermediate repeat sequence, *LINE* long interspersed nuclear element, *LTR* long terminal repeat, *PLE* Penelope-like elements, *SINE* short interspersed nuclear element, *TIR* terminal inverted repeat

such as drought (Kalendar et al. 2000). Additionally, analysis of *Copia* elements in rice and wheat showed that different families are active at different times in waves lasting for several hundreds of thousands of years (Wicker and Keller 2007).

In any case, the activity of LTR retrotransposons causes an increase in genome size. Indeed, it was shown that genome size in plants is largely determined by the amount of LTR retrotransposons, while all other TE superfamilies contribute only few percent to the total genomic DNA (Paterson et al. 2009; Schnable et al. 2009; Wicker et al. 2009b; IBI 2010). In large genomes, TEs often insert into one another, leading to complex nesting patterns with large regions that consist exclusively of TE sequences. This is illustrated in Fig. 6.3 which shows how the *rym4* locus in the barley variety *Morex* expanded to more than 65 kb by a series of TE insertions compared with the same locus in the variety *Haruna nijo*. The strong differences between the two varieties indicate that the two loci represent two ancient haplotypes which diverged approximately 930,000 years ago (Wicker et al. 2009a). These data illustrate that TE insertions can greatly expand intergenic regions within relatively short evolutionary time periods. Extensive regions consisting of nested TEs are a typical characteristic of large plant genomes, and they define the image of small gene islands being lost in an ocean of repetitive DNA.

6.7 Genome Contraction Through Deletion of Repetitive DNA

In the early 2000s, the discovery of genome expansion through TE replication led to the perception that plant genomes have "a one-way ticket to genomic obesity" (Bennetzen and Kellogg 1997). Indeed, the existence of plant genomes of several hundred times the size of the *Arabidopsis* genome suggested a one-way process. However, the model of ever-expanding genomes through TE activity could not explain why the genomes of some plant species would suddenly start to grow while others stayed small and compact. Neither could it explain the outright contradictions between taxonomy and genomes sizes (Fig. 6.1). For example, *Brachypodium* with its small genome lies in between two taxonomic groups with significantly larger genomes (*Triticeae* and *Panicoideae*, Fig. 6.1). The model of a one way-process could only explain this pattern if genome expansion had started in *Triticeae* and *Panicoideae* independently only after the three taxa had diverged.

Furthermore, comparative analysis of orthologous regions from barley and wheat revealed virtually no conservation of intergenic sequences (SanMiguel et al. 2002). Genes were found in the same linear order, while no TE was found to be conserved in both species in orthologous positions, i.e. TEs that have inserted in the common ancestor of wheat and barley (Fig. 6.4). Considering that wheat and barely have very similar genome sizes and diverged only about 12 MYA (Chalupska et al. 2008), this finding was surprising and could be best explained if there were mechanisms by which DNA could be removed from the genome.

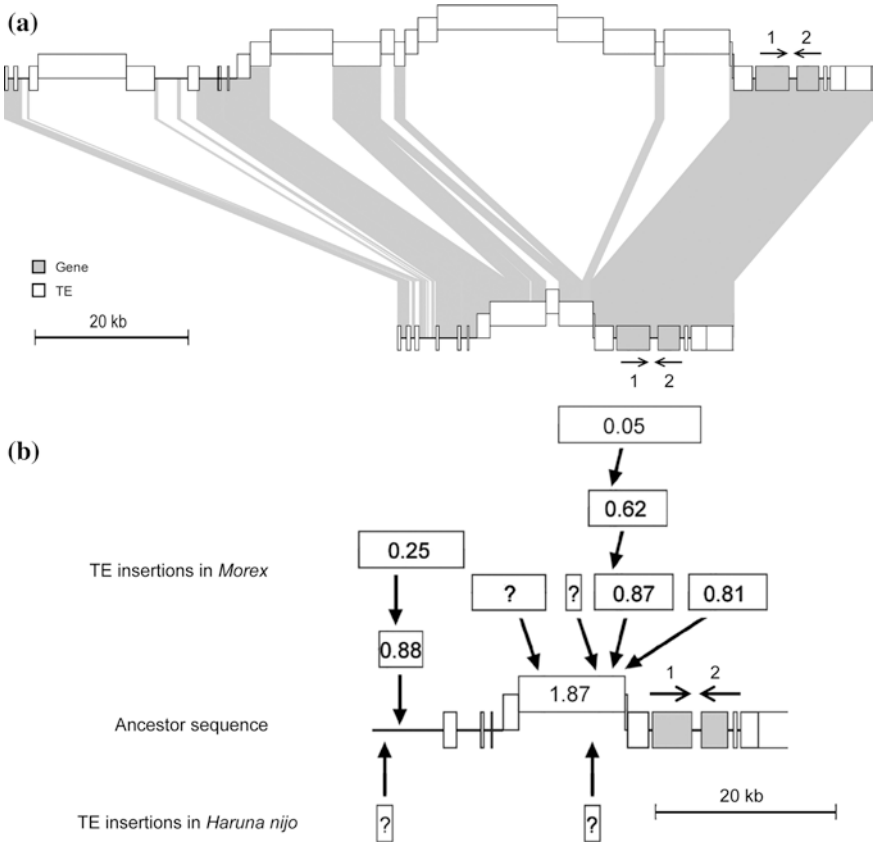

Fig. 6.3 Genome expansion through TE insertions. **a** Comparison of the *rym4* locus from the barley varieties *Morex* (*top*) and *Haruna nijo* (*bottom*). Two genes (#1 and #2) are conserved while intergenic regions differ strongly. The *Morex* locus is greatly expanded due to several TE insertions. Nested insertions of TEs are depicted as follows: TEs that have inserted into others are raised above the ones into which they have inserted. Regions that are conserved between the haplotypes of the two varieties are indicated with grey areas connecting the two maps. **b** Model for the evolution of the *rym4* locus in barley. The map depicts the sequence organization of the hypothetical ancestor sequence. Transposable elements that have subsequently inserted in *Morex* (*top*) and *Haruna nijo* (*bottom*) are indicated as colored boxes, with arrows pointing to their insertion sites. Estimated times of insertions in millions of years ago (*MYA*) are indicated inside the elements. Adapted from Wicker et al. (2009a)

One mechanism how TE sequences can be deleted from the genome is through unequal homologous recombination between the LTRs of retrotransposons. This leads to the generation of a solo-LTR while the internal domain and one equivalent of an LTR is eliminated from the genome (Fig. 6.5a). This phenomenon was long known in animals (who have only relatively few LTR retrotransposons) and was first described as a possible mechanism of genome size reduction in plants (Shirasu et al. 2000). This mechanism also provides an elegant explanation how large parts of retrotransposons

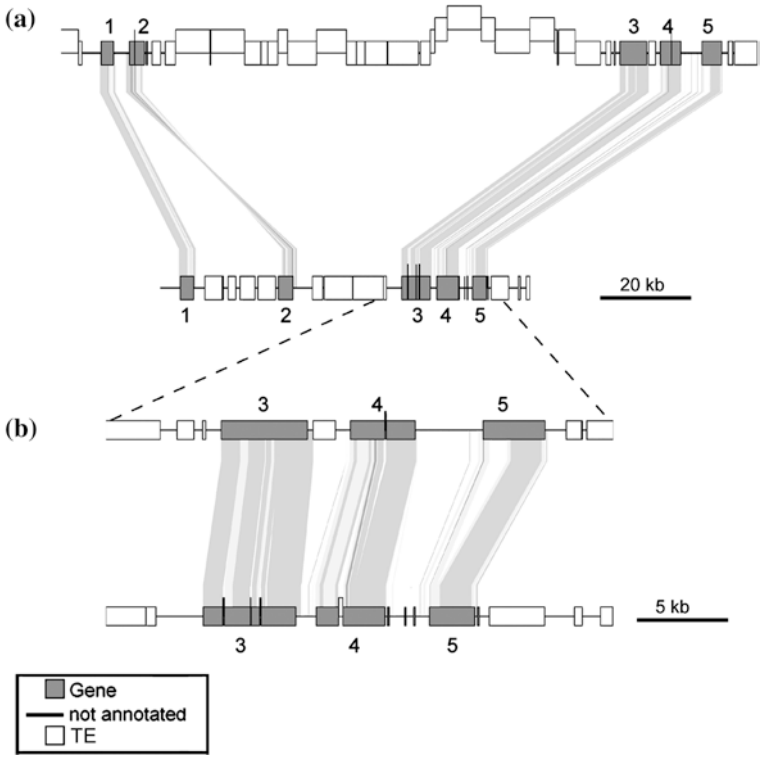

Fig. 6.4 Comparison of orthologous loci from diploid wheat *Triticum monococcum* and barley. **a** Comparison of complete BAC sequences with *T. monococcum* at the *top* and barley at the *bottom*. Orthologous regions conserved in both species are connected by shaded areas. Note that genes are conserved while intergenic regions are completely different due to genomic turnover caused by TE insertions and deletions of DNA. The only difference in the genes is an inversion of the gene #2. Nested insertions of TEs are depicted as described in Fig. 6.3. **b** Detail view of the gene island containing genes 3, 4 and 5. Almost exclusively coding sequences of genes are conserved while promoters and downstream regions have diverged to a degree that they can hardly be aligned. Based on SanMiguel et al. (2002)

can be eliminated from the genome. However, the resulting solo-LTRs still mean a net increase in genome size. The formation of solo-LTRs can therefore not explain the complete absence of colinearity in intergenic regions.

The discovery of apparently "random deletions" in the large intergenic sequences suggested a new mechanism by which repetitive DNA is eliminated independent of its sequence (Wicker et al. 2001). The mechanism that causes this kind of deletions was later described as "illegitimate recombination" (Devos et al. 2002). The term illegitimate recombination includes multiple molecular mechanisms such as replication slippage (reviewed by Lovett (2004)) or double strand break (DSB) repair through non-homologous end-joining (reviewed by van Rijk and Bloemendal (2003)) or single-strand annealing (SSA, reviewed by Hartlerode

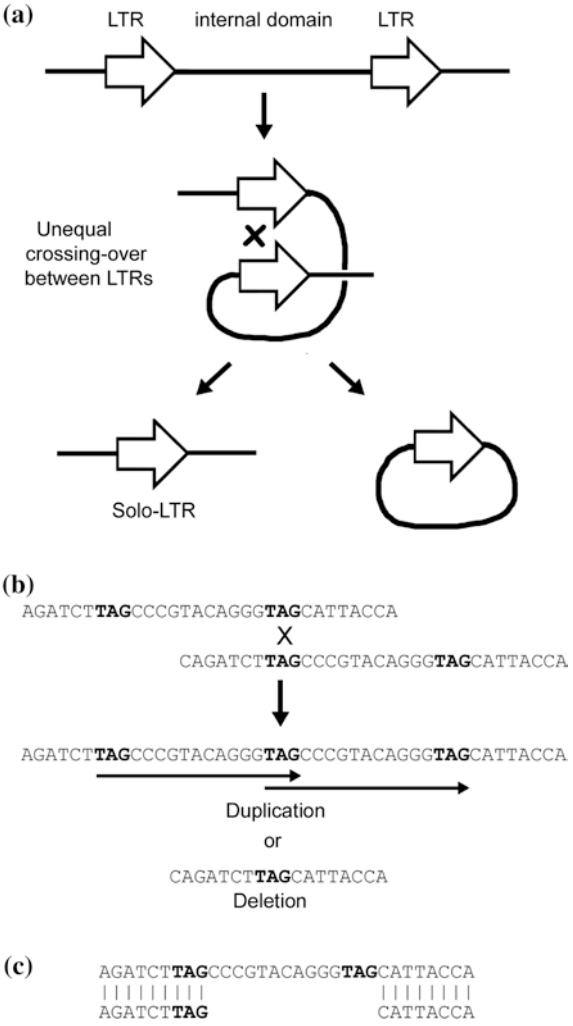

Fig. 6.5 Mechanisms for reduction of genome size. **a** Unequal homologous recombination can occur between the two LTRs of a retrotransposon. It results in a solo-LTR and a circular molecule which is then degraded. **b** Schematic depiction of illegitimate recombination. A short motif of only 3 bp that occurs twice by chance in a short stretch serves as a template for the recombination event. The two products of the recombination are a duplication and a deletion. In the case of the duplication, both units are flanked by the 3 bp motif that served as template. In that case, the 3 bp motif is referred to as the illegitimate recombination signature. **c** A deletion can only be detected if a sequence from an orthologous or paralogous locus is available. The typical illegitimate recombination signature (i.e. the sequences that served as templates for the recombination event) is printed in bold

and Scully (2009)). Whatever the precise molecular mechanisms are, they all result in recombination between very short stretches of homology (e.g. a single or a few bp), thus leading to the apparent random nature and distribution of illegitimate recombination events (Fig. 6.5b).

Studies in *Arabidopsis* (Devos et al. 2002), wheat (Wicker et al. 2003) and rice (Bennetzen et al. 2004) showed that illegitimate recombination is a major mechanism for genome contraction and might have a larger effect on genome size than the generation of solo-LTRs. As shown in Fig. 6.5b, illegitimate recombination leads either to a deletion or a duplication. However, our recent data indicate that deletions caused by DSB repair through SSA probably strongly outnumber duplications (Buchmann et al. 2012). Nevertheless, such duplications can sometimes

act as a creative force, for example in the generation of sequence variability of NBS-LRR resistance gene analogs (Wicker et al. 2007b). There, they can trigger the expansion of the leucine-rich repeat (LRR) domain which is responsible for the specific recognition of pathogens.

The processes of TE amplification and DNA removal drive a "genomic turnover" which is characterized by a balance between TE-driven duplication of DNA and deletions. This results in a permanent reshuffling of all intergenic sequences. Obviously, any alterations in the sequences essential for the immediate survival of the organism will be selected against. For example, if an important gene is disrupted by a TE insertion or partially deleted by illegitimate recombination, the offspring of that cell is not viable. However, parts of the genome which are not under selection pressure, namely TE sequences, can accumulate such rearrangements without negative effects on the fitness of the organism. Apparently, in plants this process is quite rapid and dynamic. As described above, between barley and wheat, intergenic sequences are completely reshuffled within a few million years. In fact, even between different *Triticum* species, only very limited conservation of intergenic sequences was found, although these species have diverged less than three million years ago (Wicker et al. 2003).

A first step toward unravelling the rate at which genomic turnover occurs was the introduction of a method for estimating the age of retrotransposons based on the divergence of their LTRs (SanMiguel et al. 1998). Because of the mechanism of reverse transcription, both LTRs are identical at the time of insertion (Lewin 1997). Since retrotransposons are largely free from selection pressure (Petrov 2001), they accumulate mutations at a background rate which was estimated to be 1.3×10^{-8} substitutions per site per year (Ma and Bennetzen 2004). Thus, the number of differences between the two LTR sequences is proportional to their age.

Numerous surveys have since studied age distributions of LTR retrotransposons in several species, including *Arabidopsis* (Pereira 2004), rice (Gao et al. 2004; Ma et al. 2004; Piegu et al. 2006; Wicker and Keller 2007), wheat (SanMiguel et al. 2002; Wicker and Keller 2007), maize (Du et al. 2006; Wolfgruber et al. 2009) and sorghum (Du et al. 2006). The finding common to all these studies was that hardly any retrotransposons older than 6–7 million years were found, indicating that the removal of repetitive sequences in plant genomes is rather efficient.

Genome-wide surveys of LTR retrotransposon age distributions showed that most retrotransposons are young and the older they get, the rarer they are. This finding suggested that intergenic sequences might be removed at a more or less constant rate from genomes. In fact, in rice and *Arabidopsis*, age distribution of *Copia* retrotransposons approximately follows a hyperbolic distribution, allowing to postulate a "half-life" value that describes the time it takes until half of the retrotransposons are at least partially removed from the genome (i.e. at least one LTR is deleted so that the time of insertion of the element can not be estimated anymore). Interestingly, this value was estimated to be 470,000 years in *Arabidopsis* (Pereira 2004) and 790,000 years in rice (Wicker and Keller 2007). This is consistent with the fact that rice has a significantly larger genome than *Arabidopsis*.

6.8 Dynamic Equilibrium of Genome Size

The findings on genomic turnover led to the emergence of the "increase/decrease" model for genome size evolution (Vitte and Panaud 2005) which describes genome size as a function of the rate of DNA increase through TE amplification and DNA decrease through TE removal. The balance of these two rates determines the current genome size. A change in one or both rates can therefore lead to an increase or decrease of genome size.

The rapid turnover of intergenic sequences makes them a perfect chronicle to study the background processes that shape genomes over time. Especially in large genomes like that of wheat or barley, the study of intergenic sequences allows detailed reconstruction of past events and gives an insight of the mechanisms at work. In the case of the *rym4* locus, evolutionary events could be traced back for approximately 7 million years (Wicker et al. 2005) and revealed a turbulent mixture of insertions, deletions and duplications. However, reconstruction of evolutionary events has its limitations due to high rate of genomic turnover. As mentioned above, within a few million years, intergenic sequences are completely reshuffled. Thus, detailed reconstruction of evolutionary events is not possible past that time frame.

6.9 The Molecular Basis for Gene Movement in Grass Genomes

As described above, the genomic turnover removes almost all sequence similarities outside of protein coding regions within a few million years. Therefore, among more distantly related species, sequence conservation is limited to regions which are under selection. We also discussed above that among grasses, many genetic markers are found in the same order in different species, reflecting the conserved chromosome structure of a common ancestor (Gale and Devos 1998). At the DNA sequence level, one finds that the majority of genes are in the same linear order across species, a finding that is commonly referred to as "synteny" or "colinearity". Comparison of the complete genomes of *Brachypodium*, rice and sorghum showed that at least 60–70 % of all genes are found in the same order in the three species (IBI 2010). This allows us to identify corresponding chromosomal regions with relative ease by comparing the positions of orthologous genes.

The more distantly related two species are, the fewer genes are found in colinear positions. This decrease of the number of collinear genes between species is generally attributed to "gene movement". The molecular mechanism of gene movement and the erosion of synteny that goes with it has been an unsolved riddle since the advent of comparative genomics. Several studies have shown that genes or gene fragments are sometimes captured by TEs and moved or copied to a different location (Wicker et al. 2003; Jiang et al. 2004; Lai et al. 2005; Morgante et al.

2005; Paterson et al. 2009). However, most of these captured pieces of DNA are very small and usually only contain fragments of genes. Thus, none of these studies could so far provide a robust explanation for the movement of large fragments which sometimes contain several genes.

A recent study investigated the molecular basis of the movement of large gene-containing fragments (Wicker et al. 2010). Three-way comparison of the genomes of *Brachypodium*, rice and sorghum identified genes which are specifically non-colinear in only one species (i.e. one could identify in which species the movement had occurred). This approach revealed evidence that gene movement is

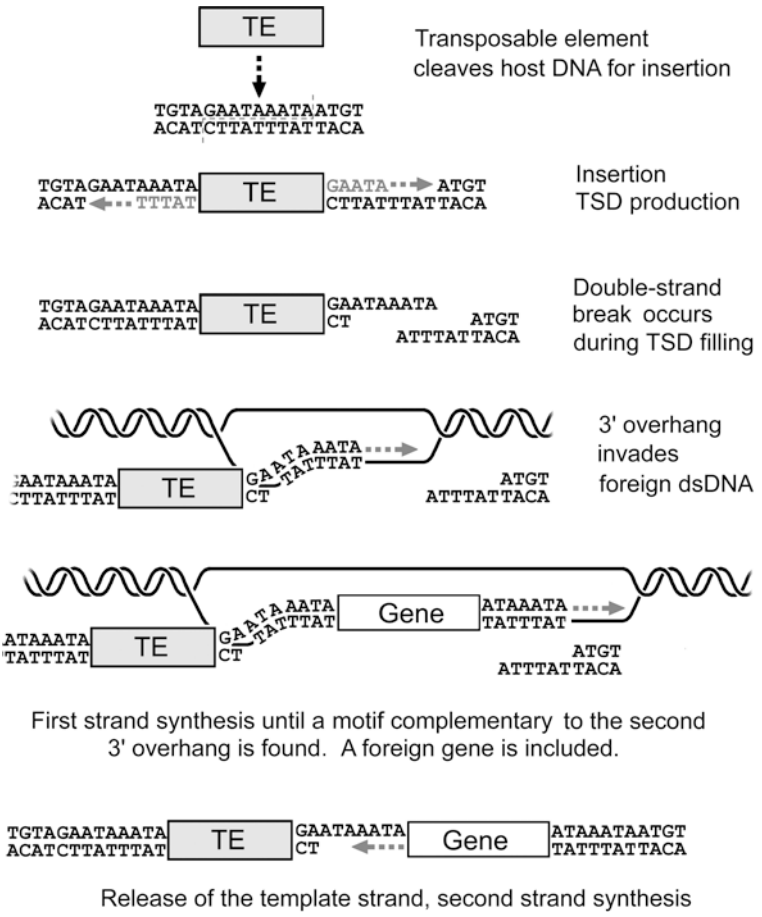

Fig. 6.6 Model for molecular events that lead to a duplication of a foreign gene. A DSB is introduced after the insertion of e *Mutator* element in the genome. A sequence fragment from elsewhere in the genome containing foreign gene is used as filler DNA to repair the DSB (adapted from Wicker et al. (2010))

mostly the result of double-strand breaks (DSBs) that are repaired by the "synthesis-dependent strand annealing" mechanism where a copy of the foreign fragment is used as "filler DNA" to repair the DSB (reviewed by Hartlerode and Scully 2009). Thus, gene movement is largely a copy-and-paste process. The duplicated fragments ranged from a few hundred bp to more than 50 kb and sometimes contained multiple genes. Most DSBs were apparently caused by TE insertions because we often found a TE immediately adjacent to the duplicated fragment (Wicker et al. 2010).

In several cases, highly diagnostic sequence motifs such as target site duplications of TEs on both sides of the duplicated fragments were found, strongly supporting the hypothesis that TE elements cause gene movement. A detailed example for the molecular events is provided in Fig. 6.6: In the first step, a *Mutator* element is inserted into the genome. The transposase cutting the host DNA creates the typical 9 bp overhangs bordering the termini of the element. Usually these gaps would be filled by cellular DNA repair enzymes, resulting in the characteristic target site duplication. We assume that during this process, a DSB can occur either precisely at the insertion point or a few bp away from it. The 3′ overhang produced by the transposase invades a complementary motif elsewhere in the genome. A filler strand is synthesized until a second matching motif is reached. The result is that the filler DNA is immediately adjacent to the TE insertion. Apparently, matching motifs of only a few bp in size are sufficient for strand invasion and priming of synthesis (Puchta 2005).

These findings suggest that gene movement is in fact the result of a rather routine process, namely the "patching up" of gaps in the genome. Besides TE insertions, there are several other mechanisms that can induce DSBs in genomes such as template slippage or unequal crossing-over (Wicker et al. 2010). In fact, recent analyses strongly suggest that the excision of TEs might also be a frequent source of DSBs (Buchmann et al. 2012).

The above observations indicate that more or less random fragments are used as filler DNA in "patching up" of gaps. If gene-containing segments are used to patch the gaps, most of these duplicated genes will probably degenerate as they are not under selection. In a few cases, the duplicated genes will gain a new function or be retained through genetic drift and therefore become established in a population.

6.10 Other Contributions of Transposable Elements to Evolution

Recent discoveries have supported the importance of TEs as a major evolutionary force (Biémont and Vieira 2006). There are a number of cases which clearly demonstrate a role of TEs in evolution. Very importantly, they create diversity in gene expression, either as a consequence of direct changes of the genome

sequence or by epigenetic mechanisms. A very good example (although not from a Triticeae species) of a direct change is the determination of grape color in grapevine: there, the insertion of a retrotransposon changes the grape color from blue to white by insertion into the promoter region of a *Myb* transcription factor gene which is involved in the control of anthocyanin production (Kobayashi et al. 2004; Morgante et al. 2007). The excision of this retrotransposon by unequal intrachromosomal recombination between the long terminal repeat sequences (LTR), resulted in one remaining LTR sequence in the *myb* promoter region. This (partial) excision formed a promoter which is only partially active, but the gene expression at a low level restores MYB activity sufficiently to allow the synthesis of some anthocyanins. This results in the production of red grapes, a nice example how retrotransposon activity is related to an economically highly relevant agronomical trait. As mentioned above, transposons can also change gene expression by epigenetic mechanisms. A classical example of such an epigenetic gene regulation based on a transposable element is the coat color of mice controlled by the *agouti* gene (Morgan et al. 1999). Finally, there is the suprising finding that six lineages of the *copia* retrotransposon show a surprising degree of conservation across phylogenetically different species such as rice and Arabidopsis, indicating some type of selection (Wicker and Keller 2007).

In an intriguing recent discovery, it was found that LINE (long interspersed nuclear element) retrotransposon activity is elevated in brain tissue vs. other somatic tissue in humans. The differential transposon activity in cells of the brain results in brain-specific genetic mosaicism. This brain-specific activity of LINE retrotransposons possibly has consequences on gene expression and neuronal function (Muotri et al. 2010; Singer et al. 2010). Whether this individual-specific diversity results in biologically significant traits remains to be determined.

An adaptive value of transposable elements was suggested by the findings in wild barley populations from Israel. There, two populations from very different, but geographically close, microclimates were analyzed for TE insertion patterns and copy number (Kalendar et al. 2000). The two populations were located in the so called evolution canyon in Northern Israel. This ecological site has two slopes which differ sharply in a number of ecologically important aspects. Specifically in wild barley plants harvested from the drier slope of the canyon, the genome was enriched with *BARE* 1 retrotransposons. Based on these data, the authors speculated about a possible adaptive selection for increased genome size caused by retrotransposon activity. There are some other data which can be viewed as being supportive for such a hypothesis, e.g. it was observed that individual families of the *copia* retrotransposon in rice and wheat were active at different time windows during evolution (Wicker and Keller 2007). These spikes of activity are possibly caused by the evolution of aggressively multiplying element followed by the evolution of efficient silencing mechanisms. However, it cannot be excluded that certain transposon families react to specific environmental conditions. Thus, one can speculate that transposon activity might leave a footprint of past environmental conditions in the genome, an intriguing and fascinating aspect of whole genome analysis in plants.

6.11 The Use of Transposons as Tools for Functional Studies

The ability of transposable elements to move to a new location in the genome has made them important tools for the analysis of gene function, particularly in plants. Insertion of a transposable element into a gene will mostly result in inactivation of this particular gene and the obtained insertion mutant can be used for further experimental studies (for a review on insertion mutagenesis in plants see Ramachandran and Sundaresan 2001). Both DNA transposons as well as retrotransposons have been used for insertional mutagenesis in plants in general, and cereals in particular. If the transposon moves quite frequently in a genome, it can create large sets of insertion mutants, ideally allowing identification of a mutant in any desired gene. Transposon insertion mutagenesis has been particularly important in maize and rice, but there is increasing interest in barley also.

As described above, the crop plant maize is at the origin of important transposable elements. The *Ac* transposon and derived *Ds* deletion variants were first isolated at the molecular level by Fedoroff et al. (1983). Based on the *Ac*/*Ds* elements as well as the *MuDr*/*Mu* transposons, saturation mutagenesis was established in maize (Walbot 2000; Fernandes et al. 2004). A large-scale study comparing the insertion patterns of *Ac*/*Ds* and *MuDR*/*Mu* revealed distinct and complementary target site preferences of the two systems (a review on the different systems available in maize was published by Weil and Monde (2007). The available collections of transposon insertion mutants in maize were recently summarized by Balyan et al. (2008).

One of the most studied and used insertional mutagenesis system applied in heterologous plant species lacking efficient endogenous transposons is based on the above mentioned *Ac*/*Ds* elements derived from maize. For instance, very efficient systems for transposon mutagenesis have been developed for rice using modified versions of these *Ac*/*Ds* elements (Qu et al. 2008). In this crop, starting with only 26 primary transformants, a total of 638 stable *Ds* insertions were identified, with a very wide distribution of the inserted sequences over the whole rice genome. Similarly, Kim et al. (2004) have shown the feasibility of using *Ds* elements for the generation of a large number of *Ds* insertion mutants in rice. A very large genetic resource based on *Ds* in rice was recently described in japonica rice cultivar Dongjin. In this study, 115,000 *Ds* insertion lines were produced, making it an excellent source for the identification of mutants (Park et al. 2009). A summary on the rice genetic resources with transposon insertions is found in a recent review (Balyan et al. 2008).

Interestingly, and very much in contrast to many other plant species, no highly active transposons have yet been identified in the economically important Triticeae species which include wheat, rye and barley. Thus, this very relevant tool for functional studies is not available in this important group of crop plants. Therefore, there are considerable efforts to establish a transgenic system based on the *Ac*/*Ds* system mostly in barley, with some first work done also in wheat. Barley can be relatively easily transformed by Agrobacterium transformation and is a diploid species, so that insertion mutants in this crop would be highly informative for

functional analysis for all Triticeae crops. Recently, several groups have described significant progress in establishing a transposon-tagging system in barley. In one study, more than 100 independent *Ds* insertions were identified and mapped. They were well distributed across the whole genome and integrated preferentially into gene-containing regions. These insertions can now be used as launch pads for further saturation of the genome with insertions (Zhao et al. 2006). Similarly, in an independent study a large number of single copy *Ds* insertions were generated and flanking sequences were determined (Singh et al. 2006). High frequencies of secondary and tertiary transpositions were observed, possibly allowing the development of large populations with independent insertions. More recently, Randhawa et al. (2009) have located single-copy *Ds* insertion events in barley by using wheat cytogenetic stocks. They concluded that it might be possible to target all genes by transposon tagging even with low transposition frequency in gene poor regions. The *Ac*/*Ds* system was also used for additional, more specific applications in barley. A gene trap approach was successfully implemented which will allow gene identification by expression studies as well as by forward and reverse genetics (Lazarow and Lütticke 2009). Along a similar line, an activation tagging system was developed in barley, based on a modified *Ds* element fused to the maize ubiquitin promoter (Ayliffe et al. 2007; Ayliffe and Pryor 2009). This system should allow identification of dominant over-expression phenotypes as done in several other plant species.

Large-scale collections of transposon insertion mutants were developed using DNA transposons, mostly the *Ac*/*Ds* and *Mu*/*MuDR* systems described above. However, in rice, a highly efficient approach was used for insertion mutagenesis based on a retrotransposon called *Tos17*. This element has a very low copy number in the rice genome, particularly if compared to other retrotransposon families. For instance, the well-studied cultivar Nipponbare with a completely sequenced genome contains only two copies. *Tos17* is activated specifically by tissue culture which is used to induce new insertions. In contrast to the *Ac*/*Ds* elements which preferentially insert into closely linked DNA, retrotransposons are mostly transposed to unlinked sequences. The molecular basis of this system and its applications for insertion-based mutagenesis in rice have been reviewed (Hirochika 2001; Kumar and Hirochika 2001). A recent summary of the research field and the complete overview on the available resources based on Tos17 insertions are also available in a recent review (Hirochika 2010).

Although *Tos17* is the only retrotransposon that has been used for large scale mutagenesis in plants, there are other retrotransposons in different plant species which are active and cause mutations. For instance, the spontaneous iron-inefficient mutant *fer* in tomato was recently shown to be caused by an insertion of the *Rider* retrotransposon into the first exon of the gene (Cheng et al. 2009). As this mutant was not derived from tissue culture, it must be assumed that it originates from spontaneous transposition in the plant. Interestingly, there is evidence that retrotransposons can also be used in a transgenic form in other species. The *Tnt1* retrotransposon was originally identified in tobacco and was subsequently used in a transgenic form in the heterologous system of lettuce. There, *Tnt1* gets frequently inserted into genes and the insertions were stably inherited (Mazier et al. 2007). As the lettuce genome

is large, it is tempting to speculate that *Tnt1* might also be useful as an active retrotransposon insertion system in the large genomes of barley and wheat. However, to our knowledge this has not yet been tested so far.

6.12 Summary and Outlook

Studying genome evolution is a complex and mostly theoretical field of research, because most theories and models can not be proved experimentally but have to be inferred between sequence and comparative analysis. Nevertheless, as described in this chapter, our understanding of the molecular mechanisms that shape genomes has greatly improved. The current knowledge opens up many new possible areas of research, some of which are outlined here. We have seen that DNA repair is a major driving force for genomic rearrangements, but we are only beginning to understand what is its actual impact on genome evolution. It will be fascinating to further explore the causes of DSBs, the role of TEs in causing DSBs and the various ways in which they are repaired. Quantitative analyses will be necessary to determine the average size ranges of filler DNA and deletions that are introduced during DSB repair. This will allow conclusions on the magnitude of the impact of DSB repair on genome evolution. Of particular interest is the question why gene colinearity erodes much more rapidly in plants than in animals. Are animal genes less likely to be moved because of their much larger size or is their large size an adaptation that prevents their movement? An increasing number of available eucaryotic genome sequences that are becoming publicly available will allow targeted comparative analyses to address these questions.

One of the central themes of this chapter is the role of TEs in genome evolution. It is essential to study further several fundamental aspects of TEs and their interaction with their host genome. Besides being a frequent source of DSBs, different types of TEs seem to be confined to specific “genomic compartments”. For example, miniature inverted-repeat transposable elements (MITEs) are almost exclusively found in or near genes and their sequence composition is very similar to that of con-coding regions of genes (i.e. promoters, introns and downstream regions). It is therefore perceivable that many if not most gene promoters and regulatory sequences are actually derived from such TE sequences. This would provide an elegant explanation for the observation that non-coding parts of genes are almost completely divergent even between closely related plant species.

In contrast to MITEs, some retrotransposon families are specifically and exclusively found in centromeres of grasses. It is suspected that specific protein domains encoded by the retrotransposon are responsible for guiding the insertion of the DNA copy to specific locations in the genome. This suggests that some of these centromeric elements indeed play a vital role in centromere function.

TEs have proven extremely useful as agents for mutagenesis as well as in gene tagging systems. The more we expand our knowledge of TEs, the more we will discover useful properties that can expand our set of molecular tools to investigate

and manipulate target organisms. Most likely, future studies will help to differentiate our general perception that TEs are purely selfish genetic elements but have, at least to some degree, been recruited by the host to fulfil specific functions.

References

AGI (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 408:796–815

Ayliffe MA, Pallotta M, Langridge P, Pryor AJ (2007) A barley activation tagging system. Plant Mol Biol 64:329–347

Ayliffe M, Pryor A (2009) Transposon-based activation tagging in cereals. Funct Plant Biol 36:915–921

Balyan H, Sreenivasulu N, Lizarazu OR, Azhaguvel P, Kianian S (2008) Mutagenesis and high throughput functional genomics. In: Sparks DL (ed) Cereal crops: current status. Academic Press, New York, pp 357–414

Bennett MD, Smith JB (1976) Nuclear DNA amounts in angiosperms. Philos Trans R Soc Lond B Biol Sci 274:227–274

Bennetzen JL, Coleman C, Liu R, Ma J, Ramakrishna W (2004) Consistent over-estimation of gene number in complex plant genomes. Curr Opin Plant Biol 7:732–736

Bennetzen JL, Kellogg EA (1997) Do plants have a one-way ticket to genomic obesity? Plant Cell 9:1509–1514

Berkman PJ, Skarshewski A, Lorenc MT, Lai K, Duran C, Ling EYS, Stiller J, Smits L, Imelfort M, Manoli S, McKenzie M, Kubaláková M, Šimková H, Batley J, Fleury D, Doležel J, Edwards D (2011) Sequencing and assembly of low copy and genic regions of isolated Triticum aestivum chromosome arm 7DS. Plant Biotech J 9:768–775

Biderre C, Pagès M, Méténier G, Canning EU, Vivarès CP (1995) Evidence for the smallest nuclear genome (2.9 Mb) in the microsporidium *Encephalitozoon cuniculi*. Mol Biochem Parasitol 74:229–231

Biémont C, Vieira C (2006) Genetics: junk DNA as an evolutionary force. Nature 443:521–524

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

Britten RJ, Graham DE, Neufeld BR (1974) Analysis of repeating DNA sequences by reassociation. Method Enzymol 29:363–418

Buchmann JP, Matsumoto T, Stein N, Keller B, Wicker T (2012) Interspecies sequence comparison in Brachypodium reveals how transposon activity corrodes colinearity. Plant J 71(4):550–563

Chain PSG, Grafham DV, Fulton RS, Fitzgerald MG, Hostetler J, Muzny D, Ali J, Birren B, Bruce DC, Buhay C, Cole JR, Ding Y, Dugan S, Field D, Garrity GM, Gibbs R, Graves T, Han CS, Harrison SH, Highlander S, Hugenholtz P, Khouri HM, Kodira CD, Kolker E, Kyrpides NC, Lang D, Lapidus A, Malfatti SA, Markowitz V, Metha T, Nelson KE, Parkhill J, Pitluck S, Qin X, Read TD, Schmutz J, Sozhamannan S, Sterk P, Strausberg RL, Sutton G, Thomson NR, Tiedje JM, Weinstock G, Wollam A, Consortium GS, Detter JC (2009) Genomics. Genome project standards in a new era of sequencing. Science 326:236–237

Chalupska D, Lee HY, Faris JD, Evrard A, Chalhoub B, Haselkorn R, Gornicki P (2008) *Acc* homoeoloci and the evolution of wheat genomes. Proc Natl Acad Sci USA 105:9691–9696

Chao S, Sharp PJ, Gale MD (1988) In: Miller TE, KoebnerK RMD (ed) Proceedings of the 7th international wheat genetics symposium. IPSR, Cambridge Laboratory, Cambridge

Chao S, Sharp PJ, Worland AJ, Warham EJ, Koebner RMD, Gale MD (1989) RFLP-based genetic maps of wheat homoeologous group 7 chromosomes. Theor Appl Genet 78:495–504

Cheng X, Zhang D, Cheng Z, Keller B, Ling H-Q (2009) A new family of *Ty1*-copia-like retrotransposons originated in the tomato genome by a recent horizontal transfer event. Genetics 181:1183–1193

Choulet F, Wicker T, Rustenholz C, Paux E, Salse J, Leroy P, Schlub S, Paslier M-CL, Magdelenat G, Gonthier C, Couloux A, Budak H, Breen J, Pumphrey M, Liu S, Kong X, Jia

J, Gut M, Brunel D, Anderson JA, Gill BS, Appels R, Keller B, Feuillet C (2010) Megabase level sequencing reveals contrasted organization and evolution patterns of the wheat gene and transposable element spaces. Plant Cell 22:1686–1701

Devos KM (2005) Updating the 'crop circle'. Curr Opin Plant Biol 8:155–162

Devos KM, Atkinson MD, Chinoy CN, Francis HA, Harcourt RL, Koebner RMD, Liu CJ, Masojć P, Xie DX, Gale MD (1993) Chromosomal rearrangements in the rye genome relative to that of wheat. Theor Appl Genet 85:673–680

Devos KM, Brown JKM, Bennetzen JL (2002) Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. Genome Res 12:1075–1079

Doolittle WF, Sapienza C (1980) Selfish genes, the phenotype paradigm and genome evolution. Nature 284:601–603

Du C, Swigonová Z, Messing J (2006) Retrotranspositions in orthologous regions of closely related grass species. BMC Evol Biol 6:62

Fedoroff N, Wessler S, Shure M (1983) Isolation of the transposable maize controlling elements *Ac* and *Ds*. Cell 35:235–242

Fernandes J, Dong Q, Schneider B, Morrow DJ, Nan G-L, Brendel V, Walbot V (2004) Genome-wide mutagenesis of Zea mays L. using RescueMu transposons. Genome Biol 5:R82

Flavell RB, Bennett MD, Smith JB, Smith DB (1974) Genome size and the proportion of repeated nucleotide sequence DNA in plants. Biochem Genet 12:257–269

Freeling M, Subramaniam S (2009) Conserved noncoding sequences (CNSs) in higher plants. Curr Opin Plant Biol 12:126–132

Gale MD, Devos KM (1998) Comparative genetics in the grasses. Proc Natl Acad Sci USA 95:1971–1974

Gallego F, Feuillet C, Messmer M, Penger A, Graner A, Yano M, Sasaki T, Keller B (1998) Comparative mapping of the two wheat leaf rust resistance loci *Lr*1 and *Lr*10 in rice and barley. Genome 41:328–336

Gao L, McCarthy EM, Ganko EW, McDonald JF (2004) Evolutionary history of *Oryza sativa* LTR retrotransposons: a preliminary survey of the rice genome sequences. BMC Genomics 5:18

Goff SA, Ricke D, Lan T-H, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, Hadley D, Hutchison D, Martin C, Katagiri F, Lange BM, Moughamer T, Xia Y, Budworth P, Zhong J, Miguel T, Paszkowski U, Zhang S, Colbert M, lin Sun W, Chen L, Cooper B, Park S, Wood TC, Mao L, Quail P, Wing R, Dean R, Yu Y, Zharkikh A, Shen R, Sahasrabudhe S, Thomas A, Cannings R, Gutin A, Pruss D, Reid J, Tavtigian S, Mitchell J, Eldredge G, Scholl T, Miller RM, Bhatnagar S, Adey N, Rubano T, Tusneem N, Robinson R, Feldhaus J, Macalma T, Oliphant A, Briggs S (2002) A draft sequence of the rice genome (Oryza sativa L. ssp. japonica). Science 296:92–100

Gregory TR (2001) Coincidence, coevolution, or causation? DNA content, cell size, and the C-value enigma. Biol Rev Camb Philos Soc 76:65–101

Guyot R, Yahiaoui N, Feuillet C, Keller B (2004) In silico comparative analysis reveals a mosaic conservation of genes within a novel colinear region in wheat chromosome 1AS and rice chromosome 5S. Funct Integr Genomics 4:47–58

Hartlerode AJ, Scully R (2009) Mechanisms of double-strand break repair in somatic mammalian cells. Biochem J 423:157–168

Hirochika H (2001) Contribution of the *Tos*17 retrotransposon to rice functional genomics. Curr Opin Plant Biol 4:118–122

Hirochika H (2010) Insertional mutagenesis with *Tos*17 for functional analysis of rice genes. Breeding Sci 60:486–492

IBI (2010) Genome sequencing and analysis of the model grass *Brachypodium distachyon*. Nature 463:763–768

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

Jaillon O, Aury J-M, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C, Vezzi A, Legeai F, Hugueney P, Dasilva C, Horner D, Mica E, Jublot D, Poulain J, Bruyère C, Billault A, Segurens B, Gouyvenoux M, Ugarte E, Cattonaro F, Anthouard V, Vico V, Fabbro CD, Alaux M, Gaspero GD, Dumas V, Felice N, Paillard S,

Juman I, Moroldo M, Scalabrin S, Canaguier A, Clainche IL, Malacrida G, Durand E, Pesole G, Laucou V, Chatelet P, Merdinoglu D, Delledonne M, Pezzotti M, Lecharny A, Scarpelli C, Artiguenave F, Pè ME, Valle G, Morgante M, Caboche M, Adam-Blondon A-F, Weissenbach J, Quétier F, Wincker P (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449:463–467

Jiang N, Feschotte C, Zhang X, Wessler SR (2004) Using rice to understand the origin and amplification of miniature inverted repeat transposable elements (MITEs). Curr Opin Plant Biol 7:115–119

Kalendar R, Tanskanen J, Immonen S, Nevo E, Schulman AH (2000) Genome evolution of wild barley (*Hordeum spontaneum*) by BARE-1 retrotransposon dynamics in response to sharp microclimatic divergence. Proc Natl Acad Sci USA 97:6603–6607

Katinka MD, Duprat S, Cornillot E, Méténier G, Thomarat F, Prensier G, Barbe V, Peyretaillade E, Brottier P, Wincker P, Delbac F, Alaoui HE, Peyret P, Saurin W, Gouy M, Weissenbach J, Vivarès CP (2001) Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. Nature 414:450–453

Kim CM, Piao HL, Park SJ, Chon NS, Je BI, Sun B, Park SH, Park JY, Lee EJ, Kim MJ, Chung WS, Lee KH, Lee YS, Lee JJ, Won YJ, Yi G, Nam MH, Cha YS, Yun DW, Eun MY, deok Han C (2004) Rapid, large-scale generation of *Ds* transposant lines and analysis of the *Ds* insertion sites in rice. Plant J 39:252–263

Kobayashi S, Goto-Yamamoto N, Hirochika H (2004) Retrotransposon-induced mutations in grape skin color. Science 304:982

Kumar A, Hirochika H (2001) Applications of retrotransposons as genetic tools in plant biology. Trends Plant Sci 6:127–134

Lai J, Li Y, Messing J, Dooner HK (2005) Gene movement by *Helitron* transposons contributes to the haplotype variability of maize. Proc Natl Acad Sci USA 102:9068–9073

Lazarow K, Lütticke S (2009) An *Ac*/*Ds*-mediated gene trap system for functional genomics in barley. BMC Genomics 10:55

Leitch IJ, Beaulieu JM, Cheung K, Hanson L, Lysak MA, Fay MF (2007) Punctuated genome size evolution in *Liliaceae*. J Evol Biol 20:2296–2308

Lewin B (1997) Genes VI. Oxford University Press, New York

Lovett ST (2004) Encoded errors: mutations and rearrangements mediated by misalignment at repetitive DNA sequences. Mol Microbiol 52:1243–1253

Ma J, Bennetzen JL (2004) Rapid recent growth and divergence of rice nuclear genomes. Proc Natl Acad Sci USA 101:12404–12410

Ma J, Devos KM, Bennetzen JL (2004) Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. Genome Res 14:860–869

Mayer KFX, Martis M, Hedley PE, Simková H, Liu H, Morris JA, Steuernagel B, Taudien S, Roessner S, Gundlach H, Kubaláková M, Suchánková P, Murat F, Felder M, Nussbaumer T, Graner A, Salse J, Endo T, Sakai H, Tanaka T, Itoh T, Sato K, Platzer M, Matsumoto T, Scholz U, Dolezel J, Waugh R, Stein N (2011) Unlocking the barley genome by chromosomal and comparative genomics. Plant Cell 23:1249–1263

Mazier M, Botton E, Flamain F, Bouchet J-P, Courtial B, Chupeau M-C, Chupeau Y, Maisonneuve B, Lucas H (2007) Successful gene tagging in lettuce using the *Tnt*1 retrotransposon from tobacco. Plant Physiol 144:18–31

McClintock B (1950) The origin and behavior of mutable loci in maize. Proc Natl Acad Sci USA 36:344–355

McClintock B (1984) The significance of responses of the genome to challenge. Science 226:792–801

Moore G, Devos KM, Wang Z, Gale MD (1995a) Cereal genome evolution. Grasses, line up and form a circle. Curr Biol 5:737–739b

Moore G, Foote T, Helentjaris T, Devos K, Kurata N, Gale M (1995b) Was there a single ancestral cereal chromosome? Trends Genet 11:81–82a

Morgan HD, Sutherland HG, Martin DI, Whitelaw E (1999) Epigenetic inheritance at the *agouti* locus in the mouse. Nat Genet 23:314–318

Morgante M, Brunner S, Pea G, Fengler K, Zuccolo A, Rafalski A (2005) Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize. Nat Genet 37:997–1002

Morgante M, Paoli ED, Radovic S (2007) Transposable elements and the plant pan-genomes. Curr Opin Plant Biol 10:149–155

Muotri AR, Marchetto MCN, Coufal NG, Oefner R, Yeo G, Nakashima K, Gage FH (2010) L1 retrotransposition in neurons is modulated by *MeCP*2. Nature 468:443–446

Park D-S, Park S-K, Han S-I, Wang H-J, Jun N-S, Manigbas N, Woo Y-M, Ahn B-O, Yun D-W, Yoon U-H, Kim Y-H, Lee M-C, Kim D-H, Nam M-H, Han C-D, Kang H-W, Yi G (2009) Genetic variation through Dissociation (*Ds)* insertional mutagenesis system for rice in Korea: progress and current status. Mol Breeding 24:1–15

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, ur Rahman M, Ware D, Westhoff P, Mayer KFX, Messing J, Rokhsar DS (2009) The *Sorghum bicolor* genome and the diversification of grasses. Nature 457:551–556

Pereira V (2004) Insertion bias and purifying selection of retrotransposons in the *Arabidopsis thaliana* genome. Genome Biol 5:R79

Petrov DA (2001) Evolution of genome size: new approaches to an old problem. Trends Genet 17:23–28

Piegu B, Guyot R, Picault N, Roulin A, Saniyal A, Kim H, Collura K, Brar DS, Jackson S, Wing RA, Panaud O (2006) Doubling genome size without polyploidization: dynamics of retrotransposition-driven genomic expansions in *Oryza australiensis*, a wild relative of rice. Genome Res 16:1262–1269

Puchta H (2005) The repair of double-strand breaks in plants: mechanisms and consequences for genome evolution. J Exp Bot 56:1–14

Qu S, Desai A, Wing R, Sundaresan V (2008) A versatile transposon-based activation tag vector system for functional genomics in cereals and other monocot plants. Plant Physiol 146:189–199

Ramachandran S, Sundaresan V (2001) Transposons as tools for functional genomics. Plant Physiol Biochem 39:243–252

Randhawa HS, Singh J, Lemaux PG, Gill KS (2009) Mapping barley *Ds* insertions using wheat deletion lines reveals high insertion frequencies in gene-rich regions with high to moderate recombination rates. Genome 52:566–575

Rensing SA, Lang D, Zimmer AD, Terry A, Salamov A, Shapiro H, Nishiyama T, Perroud P-F, Lindquist EA, Kamisugi Y, Tanahashi T, Sakakibara K, Fujita T, Oishi K, Shin-I T, Kuroki Y, Toyoda A, Suzuki Y, Hashimoto S-I, Yamaguchi K, Sugano S, Kohara Y, Fujiyama A, Anterola A, Aoki S, Ashton N, Barbazuk WB, Barker E, Bennetzen JL, Blankenship R, Cho SH, Dutcher SK, Estelle M, Fawcett JA, Gundlach H, Hanada K, Heyl A, Hicks KA, Hughes J, Lohr M, Mayer K, Melkozernov A, Murata T, Nelson DR, Pils B, Prigge M, Reiss B, Renner T, Rombauts S, Rushton PJ, Sanderfoot A, Schween G, Shiu S-H, Stueber K, Theodoulou FL, Tu H, de Peer YV, Verrier PJ, Waters E, Wood A, Yang L, Cove D, Cuming AC, Hasebe M, Lucas S, Mishler BD, Reski R, Grigoriev IV, Quatrano RS, Boore JL (2008) The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. Science 319:64–69

van Rijk A, Bloemendal H (2003) Molecular mechanisms of exon shuffling: illegitimate recombination. Genetica 118:245–249

Salse J, Feuillet C (2011) Palaeogenomics in cereals: modeling of ancestors for modern species improvement. C R Biol 334:205–211

SanMiguel P, Gaut BS, Tikhonov A, Nakajima Y, Bennetzen JL (1998) The paleontology of intergene retrotransposons of maize. Nat Genet 20:43–45

SanMiguel PJ, Ramakrishna W, Bennetzen JL, Busso CS, Dubcovsky J (2002) Transposable elements, genes and recombination in a 215-kb contig from wheat chromosome $5A^m$. Funct Integr Genomics 2:70–80

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, Gill N, Joshi T, Libault M, Sethuraman A, Zhang X-C, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht J, Grimwood J, Rokhsar D, Stacey G, Shoemaker RC, Jackson SA (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178–183

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, Minx P, Reily AD, Courtney L, Kruchowski SS, Tomlinson C, Strong C, Delehaunty K, Fronick C, Courtney B, Rock SM, Belter E, Du F, Kim K, Abbott RM, Cotton M, Levy A, Marchetto P, Ochoa K, Jackson SM, Gillam B, Chen W, Yan L, Higginbotham J, Cardenas M, Waligorski J, Applebaum E, Phelps L, Falcone J, Kanchi K, Thane T, Scimone A, Thane N, Henke J, Wang T, Ruppert J, Shah N, Rotter K, Hodges J, Ingenthron E, Cordes M, Kohlberg S, Sgro J, Delgado B, Mead K, Chinwalla A, Leonard S, Crouse K, Collura K, Kudrna D, Currie J, He R, Angelova A, Rajasekar S, Mueller T, Lomeli R, Scara G, Ko A, Delaney K, Wissotski M, Lopez G, Campos D, Braidotti M, Ashley E, Golser W, Kim H, Lee S, Lin J, Dujmic Z, Kim W, Talag J, Zuccolo A, Fan C, Sebastian A, Kramer M, Spiegel L, Nascimento L, Zutavern T, Miller B, Ambroise C, Muller S, Spooner W, Narechania A, Ren L, Wei S, Kumari S, Faga B, Levy MJ, McMahan L, Buren PV, Vaughn MW, Ying K, Yeh C-T, Emrich SJ, Jia Y, Kalyanaraman A, Hsia A-P, Barbazuk WB, Baucom RS, Brutnell TP, Carpita NC, Chaparro C, Chia J-M, Deragon J-M, Estill JC, Fu Y, Jeddeloh JA, Han Y, Lee H, Li P, Lisch DR, Liu S, Liu Z, Nagel DH, McCann MC, SanMiguel P, Myers AM, Nettleton D, Nguyen J, Penning BW, Ponnala L, Schneider KL, Schwartz DC, Sharma A, Soderlund C, Springer NM, Sun Q, Wang H, Waterman M, Westerman R, Wolfgruber TK, Yang L, Yu Y, Zhang L, Zhou S, Zhu Q, Bennetzen JL, Dawe RK, Jiang J, Jiang N, Presting GG, Wessler SR, Aluru S, Martienssen RA, Clifton SW, McCombie WR, Wing RA, Wilson RK (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112–1115

Shirasu K, Schulman AH, Lahaye T, Schulze-Lefert P (2000) A contiguous 66-kb barley DNA sequence provides evidence for reversible genome expansion. Genome Res 10:908–915

Singer T, McConnell MJ, Marchetto MCN, Coufal NG, Gage FH (2010) LINE-1 retrotransposons: mediators of somatic variation in neuronal genomes? Trends Neurosci 33:345–354

Singh J, Zhang S, Chen C, Cooper L, Bregitzer P, Sturbaum A, Hayes PM, Lemaux PG (2006) High-frequency Ds remobilization over multiple generations in barley facilitates gene tagging in large genome cereals. Plant Mol Biol 62:937–950

Soleimani VD, Baum BR, Johnson DA (2006) Quantification of the retrotransposon BARE-1 reveals the dynamic nature of the barley genome. Genome 49:389–396

Swigonová Z, Lai J, Ma J, Ramakrishna W, Llaca V, Bennetzen JL, Messing J (2004) Close split of sorghum and maize genome progenitors. Genome Res 14:1916–1923

Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, Schein J, Sterck L, Aerts A, Bhalerao RR, Bhalerao RP, Blaudez D, Boerjan W, Brun A, Brunner A, Busov V, Campbell M, Carlson J, Chalot M, Chapman J, Chen G-L, Cooper D, Coutinho PM, Couturier J, Covert S, Cronk Q, Cunningham R, Davis J, Degroeve S, Déjardin A, Depamphilis C, Detter J, Dirks B, Dubchak I, Duplessis S, Ehlting J, Ellis B, Gendler K, Goodstein D, Gribskov M, Grimwood J, Groover A, Gunter L, Hamberger B, Heinze B, Helariutta Y, Henrissat B, Holligan D, Holt R, Huang W, Islam-Faridi N, Jones S, Jones-Rhoades M, Jorgensen R, Joshi C, Kangasjärvi J, Karlsson J, Kelleher C, Kirkpatrick R, Kirst M, Kohler A, Kalluri U, Larimer F, Leebens-Mack J, Leplé J-C, Locascio P, Lou Y, Lucas S, Martin F, Montanini B, Napoli C, Nelson DR, Nelson C, Nieminen K, Nilsson O, Pereda V, Peter G, Philippe R, Pilate G, Poliakov A, Razumovskaya J, Richardson P, Rinaldi C, Ritland K, Rouzé P, Ryaboy D, Schmutz J,

Schrader J, Segerman B, Shin H, Siddiqui A, Sterky F, Terry A, Tsai C-J, Uberbacher E, Unneberg P, Vahala J, Wall K, Wessler S, Yang G, Yin T, Douglas C, Marra M, Sandberg G, de Peer YV, Rokhsar D (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science 313:1596–1604

Vicient CM, Kalendar R, Anamthawat-Jónsson K, Schulman AH (1999) Structure, functionality, and evolution of the BARE-1 retrotransposon of barley. Genetica 107:53–63

Vitte C, Panaud O (2005) LTR retrotransposons and flowering plant genome size: emergence of the increase/decrease model. Cytogenet Genome Res 110:91–107

Walbot V (2000) Saturation mutagenesis using maize transposons. Curr Opin Plant Biol 3:103–107

Weil CF, Monde R-A (2007) Induced mutations in maize Israel. J Plant Sci 55:183–190

Wicker T, Buchmann JP, Keller B (2010) Patching gaps in plant genomes results in gene movement and erosion of colinearity. Genome Res 20:1229–1237

Wicker T, Keller B (2007) Genome-wide comparative analysis of copia retrotransposons in Triticeae, rice, and Arabidopsis reveals conserved ancient evolutionary lineages and distinct dynamics of individual copia families. Genome Res 17:1072–1081

Wicker T, Krattinger SG, Lagudah ES, Komatsuda T, Pourkheirandish M, Matsumoto T, Cloutier S, Reiser L, Kanamori H, Sato K, Perovic D, Stein N, Keller B (2009a) Analysis of intraspecies diversity in wheat and barley genomes identifies breakpoints of ancient haplotypes and provides insight into the structure of diploid and hexaploid triticeae gene pools. Plant Physiol 149:258–270a

Wicker T, Matthews D, and Keller B (2002) TREP: a database for Triticeae repetitive elements. Trends Plant Sci 7:561–562

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, Paux E, SanMiguel P, Schulman AH (2007a) A unified classification system for eukaryotic transposable elements. Nat Rev Genet 8:973–982a

Wicker T, Stein N, Albar L, Feuillet C, Schlagenhauf E, Keller B (2001) Analysis of a contiguous 211 kb sequence in diploid wheat (*Triticum monococcum* L.) reveals multiple mechanisms of genome evolution. Plant J 26:307–316

Wicker T, Taudien S, Houben A, Keller B, Graner A, Platzer M, Stein N (2009b) A whole-genome snapshot of 454 sequences exposes the composition of the barley genome and provides evidence for parallel evolution of genome size in wheat and barley. Plant J 59:712–722b

Wicker T, Yahiaoui N, Guyot R, Schlagenhauf E, Liu Z-D, Dubcovsky J, Keller B (2003) Rapid genome divergence at orthologous low molecular weight glutenin loci of the A and A^m genomes of wheat. Plant Cell 15:1186–1197

Wicker T, Yahiaoui N, Keller B (2007b) Illegitimate recombination is a major evolutionary mechanism for initiating size variation in plant resistance genes. Plant J 51:631–641b

Wicker T, Zimmermann W, Perovic D, Paterson AH, Ganal M, Graner A, Stein N (2005) A detailed look at 7 million years of genome evolution in a 439 kb contiguous sequence at the barley *Hv-eIF4E* locus: recombination, rearrangements and repeats. Plant J 41:184–194

Wolfgruber TK, Sharma A, Schneider KL, Albert PS, Koo D-H, Shi J, Gao Z, Han F, Lee H, Xu R, Allison J, Birchler JA, Jiang J, Dawe RK, Presting GG (2009) Maize centromere structure and evolution: sequence analysis of centromeres 2 and 5 reveals dynamic Loci shaped primarily by retrotransposons. PLoS Genet 5:e1000743

Yang YW, Lai KN, Tai PY, Li WH (1999) Rates of nucleotide substitution in angiosperm mitochondrial DNA sequences and dates of divergence between Brassica and other angiosperm lineages. J Mol Evol 48:597–604

Yu J, Hu S, Wang J, Wong GK-S, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, Cao M, Liu J, Sun J, Tang J, Chen Y, Huang X, Lin W, Ye C, Tong W, Cong L, Geng J, Han Y, Li L, Li W, Hu G, Huang X, Li W, Li J, Liu Z, Li L, Liu J, Qi Q, Liu J, Li L, Li T, Wang X, Lu H, Wu T, Zhu M, Ni P, Han H, Dong W, Ren X, Feng X, Cui P, Li X, Wang H, Xu X, Zhai W, Xu Z, Zhang J, He S, Zhang J, Xu J, Zhang K, Zheng X, Dong J, Zeng W, Tao L, Ye J, Tan J, Ren X, Chen X, He J, Liu D, Tian W, Tian C, Xia H, Bao Q, Li G, Gao H, Cao T, Wang J, Zhao

W, Li P, Chen W, Wang X, Zhang Y, Hu J, Wang J, Liu S, Yang J, Zhang G, Xiong Y, Li Z, Mao L, Zhou C, Zhu Z, Chen R, Hao B, Zheng W, Chen S, Guo W, Li G, Liu S, Tao M, Wang J, Zhu L, Yuan L, Yang H (2002) A draft sequence of the rice genome *Oryza sativa*. Science 296:79–92

Zhao T, Palotta M, Langridge P, Prasad M, Graner A, Schulze-Lefert P, Koprek T (2006) Mapped *Ds*/T-DNA launch pads for functional genomics in barley. Plant J 47:811–826

Chapter 7
Functional Annotation of Plant Genomes

Vindhya Amarasinghe, Palitha Dharmawardhana, Justin Elser and Pankaj Jaiswal

7.1 Introduction

The recent introduction of highly-efficient next-generation sequencing platforms (Roche 454, Illumina, PacBio, Life Technologies SOLiD, etc.) has lead to an increased number of sequenced plant genomes. Compared to the time (`5–7 years) taken to complete *Arabidopsis thaliana* (AGI 2000) and rice genomes (Goff, Ricke et al. 2002; IRGSP 2005), the time taken to sequence a genome of comparable size can now be measured in weeks rather than years. The recently sequenced plant genomes for which we have draft versions include angiosperms like cacao (Couch, Zintel et al. 1993), cucumber (Huang, Li et al. 2009), papaya (Ming, Hou et al. 2008), strawberry (Shulaev, Sargent et al. 2011), *Jatropha* (Sato, Hirakawa et al. 2011), date palm (Al-Dous, George et al. 2011), poplar (Tuskan, Difazio et al. 2006), soybean (Schmutz, Cannon et al. 2010), grape (Jaillon, Aury et al. 2007), apple (Velasco, Zharkikh et al. 2010), sorghum (Paterson, Bowers et al. 2009), *Brachypodium* (Vogel 2010), maize (Schnable, Ware et al. 2009), bryophyte *Physcomitrella* (Rensing, Lang et al. 2008), lycophyte *Seilaginella* (Banks, Nishiyama et al. 2011) and algae *Chlamydomonas* (Merchant, Prochnik et al. 2007) and *Ectocarpus* (Cock, Sterck et al. 2010). Once the data generated by genome sequencers is pruned and assembled, a genome sequencing exercise is merely a report of the sequential arrangement of nucleotides of a particular organism's chromosomes. These nucleotide sequences are therefore just the starting point of a genome sequencing project. An organism's nucleotide sequence is informative to the researchers only when it is interpreted in the context of biology. This includes, phylogenetic position in the "tree of life" taxonomy, response to various biotic and

V. Amarasinghe · P. Dharmawardhana · J. Elser · P. Jaiswal (✉)
Deparment of Botany and Plant Pathology, Oregon State University,
2082 Cordley Hall, Corvallis, OR 97331-2902, USA
e-mail: jaiswalp@science.oregonstate.edu

P. K. Gupta and R. K. Varshney (eds.), *Cereal Genomics II*,
DOI: 10.1007/978-94-007-6401-9_7,

abiotic stresses besides adaptation to its natural/introduced growth environment and the genetic variation found within the species that contributes to natural variation in the phenotype. Such interpretations are termed annotations.

Genome annotation proceeds in four major stages: (1) Structural annotation, which begins with chromosome assembly and gene identification and continues with the marking of repeat regions, which amount to a large percentage of the genome, (2) identification of gene homologues from different species (also called orthologs) and gene duplications within the same genome (paralogs), (3) functional annotation, where biological functions and processes associated with the encoded proteins are predicted, and (4) mapping genetic variation and markers for the elucidation of genotype-phenotype associations. In this chapter, we will describe the structural annotation of genes in brief and emphasize on gene homology and functional annotation analysis of plant genomes. Such gene annotations can be accomplished via manual extraction and interpretation of information from the published scientific literature. However, given the size and complexity of plant genomes which include genome duplication, ploidy and excessive amounts of repeat representations, computationally driven annotation methods are the most practical starting point for a novel genome.

7.2 Structural Annotation

7.2.1 Masking Repetitive DNA

In the current chapter, as in many public genome annotation pipelines, we will define the genome as masked without repeats or unmasked with repeats. The masked region includes highly repetitive DNA elements such as transposable elements, LINEs, LTRs, MULEs and SINEs and these accounts for a major part of the DNA sequence in the genomes of most higher-order organisms. For example, *Brachypodium*, rice, sorghum and wheat have repetitive fractions of 21, 26, 54 and over 80 % respectively (Vogel 2010). Prior to computational identification of protein coding genes, such highly repetitive and low complexity DNA sequences are analyzed and identified by screening against publicly available repeat sequence libraries such as the Repeat Element Database-MIPS-REdat (Spannagl, Noubibou et al. 2007). MIPS-REdat has an exhaustive collection of plant repetitive elements that is compiled from Repbase (Jurka, Kapitonov et al. 2005), TIGR repeats (Ouyang and Buell 2004) and Triticeae Repeat Sequence Database (TREP, http://wheat.pw.usda.gov/ITMI/Repeats/index.shtm). The genome sequence to be annotated is screened against the repeat databases using Repeat Masker (repeatmasker.org), a program that screens DNA sequences for interspersed repeats and low complexity DNA sequences (Tarailo-Graovac and Chen 2009), Dust, a program for filtering low complexity regions from nucleic acid sequence and Tandem Repeat Finder (TRF, http://tandem.bu.edu/trf/trf.html) (Benson 1999), used to locate tandem repeats.

7.2.2 Gene Prediction Using Empirical and Ab Initio Approaches

Empirical evidence-based methods are dependent on existing evidence available for the encoded mRNA (EST/cDNA/RNA-seq) and or protein sequence. In this approach, the first option is to run a BLASTX (Camacho, Coulouris et al. 2009) against the GenBank non-redundant (NR) database that could identify the best hit cluster loci within the genome assembly. Another option is to use known RNA (EST/cDNA/RNA-seq) and/or peptide sequences that are already available for the genome of interest, or a closely-related genome. In both methods, genes can be directly called using BLAST. Since the expression of genes is highly dependent on cell type, developmental, temporal and environmental conditions, many genes in an empirical dataset may or may not be represented in EST, cDNA or RNA-seq libraries and are considered incomplete even with extensive RNA-sequencing. Therefore, when annotating the genome of a newly-sequenced species, initial round of *ab initio* gene prediction methods that systematically scan the genomic sequence for nucleotide *signals* predicting the existence of a gene in the unmasked regions of the genome are considered a necessity.

Ab initio gene prediction in eukaryotic genomes is a complex process. Unlike prokaryotes, the eukaryotic Open Reading Frame (ORF) is not continuous and is not part of operons. The eukaryote gene ORFs often contains alternate blocks of exon (protein-coding region) and intron (non-protein coding) sequences. Transcribed regions in the genome are at a much lower density compared to the non-transcribed regions. The splice junctions between exon and interon sequence boundaries may have only weak signatures, and very short exons may be buried within very long introns. Aside from the transcribed regions of genes, genomes also have promoter and other regulatory sequences that are complex and not well understood. All the above attributes complicate the distinctive telltale nucleotide signatures of eukaryotic genomes and make it more difficult to decipher than prokaryotic genomes. Consequently, *ab initio* programs depend on probabilistic models such as Hidden Markov Models (HMM) and use combined sets of signal or sensor information to predict gene boundaries and internal gene structures such as introns, exons and regulatory elements. Some of the frequently used *ab initio* gene prediction programs include FGENESH (Solovyev, Kosarev et al. 2006), GeneID (Blanco, Parra et al. 2007; Blanco and Abril 2009), Augustus (Stanke and Morgenstern 2005) and SNAP (Korf 2004). These programs can generally be "trained" by developers for different organisms using high-confidence reference gene sets or expressed sequence tag (EST)-based gene models.

One of the most widely used gene finding programs, FGENESH, is utilized in the standard analysis pipelines of genome sequencing projects and genomic databases such as Gramene. Gramene (www.gramene.org) runs FGENESH against each core genome database in its collection and is implemented via the Ensembl pipeline (Potter, Clarke et al. 2004). Gramene has also created their own version of an evidence-based gene prediction pipeline (the Gramene pipeline) for plants

(Liang, Mao et al. 2009), which can use transcriptional evidence across related species. FGENESH is an HMM-based program and can be trained using various high-confidence EST gene sets. There are many variations of FGENESH software, namely, for the analysis of genomic DNA, cDNA and protein-based gene predictions as well as for the prediction of multiple (alternatively spliced) variants of potential genes in genomic DNA.

7.3 Gene Homology Analysis

Gene homology analysis is a comparative procedure where evolutionary relationships between the genes encoded by a genome of interest are compared to genes in a well-annotated genome of another species. The choice for the second species is often a model organism such as *Arabidopsis thaliana* or *Oryza sativa* in plants. There are two reasons to perform this analysis. First, it is helpful in elucidating the phylogenetic (evolutionary) relationships of genes in the genome of interest with respect to genes from other known species. Second, considering that the genome of interest being pursued for sequencing is a less-studied model, many of the genes have no empirical evidence of known function. Therefore, the homology-based phylogenetic relationship to the "known" genes with empirical evidence from other species (preferably a model plant) can serve as a reference to project annotations onto genes from the new species.

The basic premise is that genomes in current living organisms have evolved from common ancestral genomes. During this process, they may continue to harbor ancestral genes or may have acquired/created novel ones by DNA rearrangements, natural variation and/or horizontal transfer as deemed fit for their survival and adaptation. If a specific set of gene sequences in divergent species has *originated* from a single ancestral gene in the last-shared common ancestor, the members of that set are termed *orthologs* of each other. Such orthologous genes in two species that have evolved from a single gene in the last common ancestor are likely to have a similar function. Since functionally annotating all newly sequenced genomes by empirical means (by experimental evidence) is by no means practical or feasible, empirical evidence-based functional information already available in an annotated genome (e.g. a model organism) can be mapped to the orthologous genes in a newly sequenced genome. As such, gene orthology analysis is a crucial step in functional annotation.

Finding the true functional orthologs between two species, however, is not trivial. In most cases the genes would have duplicated both before and following speciation. Consequently, there is more than one ortholog within a single species, and these are termed *paralogs*. In homology analysis it is important to distinguish between the paralogs that arose before speciation from the ones that arose after speciation. Remm et al. (2001) have introduced the terms *out-paralogs* (paralogs arising from gene duplication before the speciation event) and in-paralogs (paralogs that are arising from gene duplication after the speciation

event). By definition of orthologs (above) the out-paralogs are not true orthologs. In orthology analysis functional equivalence in two species is recognized between orthologs and *in-paralogs*. Out-paralogs are avoided since they are more ancient, and had been subjected to different selection pressures and therefore may have assumed different functions and may lack functional equivalence. It is an important but challenging task to identify the orthologs and in-paralogs during homology analysis. Phylogenetic gene trees may be what first come to mind in identifying orthologs since orthologs by definition are related by evolution. However, building phylogenetic trees is computationally intensive. In our group, we use an all-vs-all sequence comparison method called InParanoid (Remm, Storm et al. 2001; O'Brien, Remm et al. 2005; Berglund, Sjolund et al. 2008; Ostlund, Schmitt et al. 2010). The InParanoid program has been developed to identify clusters of orthologs which are in-paralogs, while avoiding the inclusion of out-paralogs.

7.3.1 InParanoid Gene Orthology Analysis

This algorithm makes an all-versus-all comparison between the complete set of protein sequences of one species (i.e. a newly-sequenced genome) and another species (i.e. a model species). It also provides the option to use a third group as an out group. The program requires two FASTA format protein sequence file (e.g. file-A from species-A and file-B from species-B). All-versus-all BLAST search is run and sequence pairs with reciprocal best hits are detected. Sequences from out group species are optionally used to detect cases of selective ortholog loss. The species A–B sequence pairs are eliminated if either sequence—A from species-A or B from species-B—scores higher to the outgroup sequence than they score to each other. In-paralogs are clustered together with each remaining pair of potential orthologs. Finally, an optional bootstrapping technique can be used to estimate the probability that a given pair of orthologs have a mutual best score only by chance.

Pairwise Similarity Comparisons

Ortholog detection is initiated by calculating the similarity scores between all studied sequences. For example all protein sequences of the new genome (species-A) versus all protein sequences of a model species (Species-B). Generally this is done with BLAST or with any other pair-wise alignment program. Similarity scores are calculated between the all vs all comparisons that include against self and another species (A $\rightarrow$ B$\rightarrow$B $\rightarrow$ A$\rightarrow$ A). User customized score cut off and an overlap cut off can be applied to avoid insignificant hits and short domain level matches. This is particularly critical since orthologous sequences are expected to maintain homology along majority of the sequence length.

Creating Homologue Gene Clusters

As explained above ortholog detection starts with calculating the similarity scores. From the similarity scores best scoring ortholog pair is identified and is marked as the main ortholog pair or the seed ortholog pair. In-paralogs are added around the main ortholog pair for each species separately. The program runs through all main ortholog pairs and adds in-paralogs from the two datasets (two species). If two ortholog groups overlap they are either merged or deleted according to pre-set rules (Remm, Storm et al. 2001).

Confidence Values for Recognizing In-Paralogs

All in-paralogs should be orthologous to the main ortholog but some are more similar to the main ortholog while others are too dissimilar. InParanoid uses a confidence value calculation to call in-paralog to the main ortholog" within a species. This value varies from 100 % for the main ortholog to 0 % to a sequence at or below the minimum similarity score required to be an in-paralog.

This is calculated by the following method (Remm, Storm et al. 2001).

$$\text{Confidence for Ap} = 100\ \% \times (\text{score}AA\text{p}) - \text{score}AB\ /\ (\text{score}AA - \text{score}AB)$$
$$\text{Confidence for Bp} = 100\ \% \times (\text{score}BB\text{p} - \text{score}AB\ /\ (\text{score}BB - \text{score}AB).$$

Ap is an in-paralog from dataset A, Bp is an in-paralog from dataset B, A is the main ortholog from dataset A, B is the main ortholog from dataset B, score is the similarity between the two proteins (Remm, Storm et al. 2001). As such each member of an ortholog cluster receives an in-paralog score which reflects the relative distance to the seed in-paralog.

InParanoid Implementation

There are two options for implementing InParanoid, i.e. (1) local installation of InParanoid or (2) online analysis using the web interface. Prior to whole genome orthology analysis using locally installed InParanoid one needs to locally install, BLAST software available from NCBI and a data set of fasta format sequence files. In our group we use an in-house installed data set of 45 species downloaded and regularly updated from multi species databases, Phytozome, Gramene, Ensemble Genomes and from model species specific databases. Small datasets, on the other hand, can be analyzed using the InParanoid web interface (http://inparanoid.sbc.su.se/cgi-bin/index.cgi). InParanoid 7 at the time of writing this article has a database of eukaryotic ortholog groups for 100 organisms. These genomes are considered as completely sequenced with >6X coverage with <1 % unknown proteins. The data sources are Ensemble, NCBI and MODs (model organism databases). The BLAST tool within this web interface enables one to find ortholog clusters for hitherto non annotated protein sequences (e.g. expression analysis derived sequences of a non annotated genome

or a subset of a newly sequenced genome) by doing a BLAST search against the above protein data set. On the other hand, a known protein of interest can be analyzed using a gene identifier, a protein identifier or a text search.

Creating Super Clusters from Ortholog Clusters of Multiple Pair Wise Comparisons

Comparative gene homology analysis of a newly sequenced genome with multiple already annotated genomes provides additional information such as, how many orthologous gene clusters of a new genome are (a) shared with other species (b) shared with which species (c) unique to the newly sequenced species (e.g. Fig. 7.1) and (d) how many are in-paralogs. A multi species view would thus provide information on phylogenetic relationships between multiple species and provide significant functional links to corroborate orthology based functional annotation of the

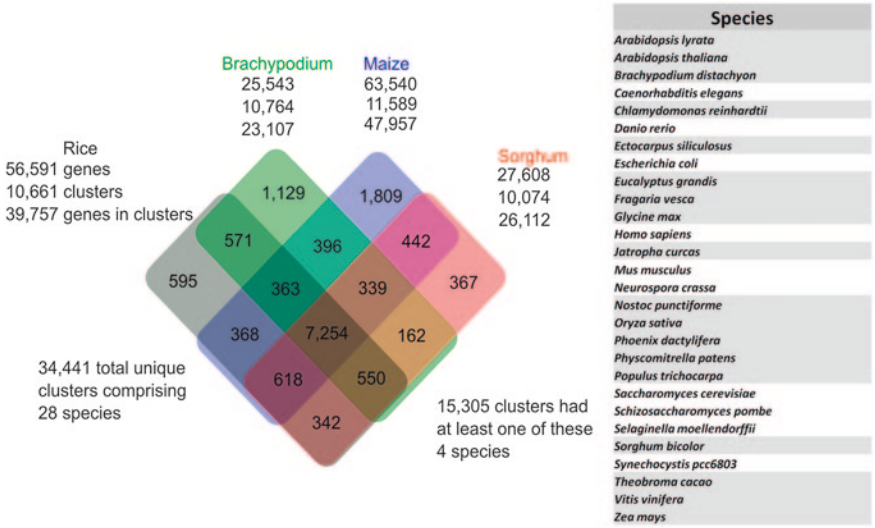

Fig. 7.1 A Venn diagram illustrating unique and shared gene families between rice, maize, sorghum, and *Brachypodium*: The relationships of gene families among the different species were derived from InParanoid analysis of a total of 28 species listed in the figure (Plant species are highlighted in *grey colored* cells). Comparison of the 4 species illustrated here reveals 595, 1129, 1809 and 367 gene clusters (note that these are not number of genes) being unique to rice, maize, sorghum and *Brachypodium* respectively. The Venn diagram illustrates that, in our InParanod analysis of *Brachypodium* (with comparison to 28 other species from tree of life), out of the 25,543 *Brachypodium* genes we have assigned 23,107 genes (~90 %) to the gene (protein) clusters, while 2,436 genes of the total 25,543 genes did not fall in clusters and therefore, are unique to *Brachypodium*. Note that at least two genes/proteins from different species are needed to form a cluster. Analysis of the all the 34,441 clusters and comparing the relative uniqueness between rice, *Brachypodium*, maize and *Sorghum* we find 1,129 clusters unique to *Brachypodium*

new species. InParanoid however, performs only pairwise comparisons. Therefore, additional methods such as MultiParanoid (Alexeyenko, Tamas et al. 2006) is used to generate multi species comparisons. The MultiParanoid program chains together the overlapping pairwise groups generated by InParanoid. To reduce false positives, relatively strict cutoffs for InParanoid scores are used within MultiParanoid. The means of the Inparanoid scores are used to calculate the confidence scores for the members of the MultiParanoid clusters. Attempting to lower confidence value cutoff of MultiParanoid to tighten the 'MultiParanoid cluster' and eliminate low confidence orthologs can lead to increased false negatives. While doing this analysis it is important to note that to get true ortholog clusters in MultiParanoid the proteomes should have diverged at a somewhat similar time point. If not the program will not be able to find an ancestral node to represent the last common ancestor for clustering true orthologs (Alexeyenko, Tamas et al. 2006).

Parallel to MultiParanoid in our group we have developed an algorithm to construct super clusters from InParanoid clusters (Shulaev, Sargent et al. 2011). This algorithm is developed with the premise that a gene can only be in one super cluster. When super clusters are created from InParanoid pairwise clusters, some genes cluster into super clusters because they are so called vicarious orthologies (i.e. they are connected to each other by common pairwise comparisons). For example if there are three genes A, B, and C from three different species. InParanoid may find a high degree of homology between genes A–B, B–C, but does not have a direct relationship between A–C. If it is just assumed that these can be simply grouped together, there is a low confidence that A–C are actually homologous. It could have been that gene B was very long and overlapped with A and C, but in different regions. To try to minimize the impact of these "vicarious" orthologies, a simple "voting" algorithm was implemented. This algorithm counts how many times a gene is found to have similar paralogs, and then puts those genes in the same super cluster. Over the course of many species pairings, this should reduce the super cluster assignment to genes that had a high BLAST score to each other. In an example analysis we compared 28 species across many clades and four Kingdoms (Fig. 7.1). Based on the super cluster analysis when we query all the super clusters to compare data sets between the four species, rice, *Brachypodium*, maize and sorghum one can easily find that 7,264 super clusters are shared by all the four species, and rice has unique genes (among these species) that are listed in 595 super clusters. This super cluster analysis is somewhat naïve compared to MultiParanoid, and does not include scoring in its analysis, but has the advantage of working across clades without a common ancestor and the requirement that all the species diverged at the same time.

7.3.2 Other Gene Homology Analysis Methods

OrthoMCL provides a scalable method for constructing orthologous groups across Eukaryotic taxa (Li, Stoeckert et al. 2003; Chen, Mackey et al. 2006).

This approach is similar to InParanoid but is applicable to multiple species, distinguishes between divergence and functional redundancy, and requires recent paralogs to be more similar to each other than to any sequence in other species. To resolve orthologous relationships between genes of multiple species, OrthoMCL applies a Markov Cluster algorithm (Li, Stoeckert et al. 2003). Another algorithm provided by Ensemble, the Compara GeneTrees (Vilella, Severin et al. 2009) provides explicit tree topology and takes into account homology relationships and duplication consistency score to generate the trees. Compara driven phylogenetic tree based clustering method enables researchers to explore evolution of gene families and recognize ancestral gene duplications which in turn may be associated with positive selection. The Gramene Ensemble Compara pipeline for plants (Youens-Clark, Buckler et al. 2011) provides homology datasets on select plant genomes and pan genome (non-plant) for comparisons.

7.4 Functional Annotation of Genomes Using an Integrative Approach

7.4.1 Discovering Protein Signatures Using InterPro

Once the ORF of a gene is identified by *ab initio* methods, the potential peptide sequence—the primary structure of the protein—has been determined. This is fundamentally a string of amino acids. This primary structure provides a framework for the protein to fold and form three-dimensional secondary, tertiary and quaternary structures which in turn provide functionality to the protein. The three-dimensional structure is manifested by different structural domains or signatures which in turn are brought forth by the string(s) of amino acid sequences, combined with other characteristics such as side chains and modifications. The different structural signatures therefore provide crucial information on physiochemical properties which determines specific functional characteristics of the protein. The proteins of a newly sequenced genome thus can be functionally annotated by identifying these predictive signatures within each protein.

In the past decade, multiple databases have been built to predict functionality of proteins via recognition of predictive signatures. These databases have been created independently and may use different methodologies for signature recognition. Their domain and feature libraries for associated protein sequences are drawn from a mixed bag of functional domain evidences, and computationally-driven consensus sequence models. These databases possess individual strengths and weaknesses but share redundancies which complement each other in many ways. Therefore, the best outcome in searching for predictive signatures and in projecting the function based on best hit to a consensus/known domain sequence may be achieved by combined searches. A strategy to combine multiple databases was implemented by InterPro (an Integrated resource of Protein Domains

and Functional sites) database (Mulder and Apweiler 2007). InterPro is a collaborative project which provides an integrated layer upon the most commonly-used suites of signature databases (Pfam, PROSITE, PRINTS, ProDom, SMART, TIGRFAMMs). The collection of domain identities assigned by various resources is integrated by InterPro through a curational effort that includes manual and computational efforts. If different signatures originating from different member databases match the same set of proteins and falls on the same place in the experimental/seed sequence they (the set of signatures) are presumed to describe the same functional family, domain or site and are placed into a single InterPro entry (or InterPro record: IPR). For example, over 50 % of approximately 58,000 of the current signatures in the member databases had been manually incorporated into an InterPro entry and have empirical evidence of associated function that had been published in a peer reviewed journal. Once an InterPro entry is created it is annotated with attributes such as: a unique accession number, name, a descriptive abstract and cross references to other resources including Gene Ontology terms (GO terms).

There are different InterPro entry or IPR types, viz. Family, Domain, Region, Repeat or Site. Sites are sub-classified into conserved sites, active sites, binding sites or Post-translational Modifications (PTMs). They are described in Table 7.1. InterPro database also maintains links to other databases including MEROPS protease resource: (Rawlings, Tolle et al. 2004), IntAct (protein interaction database: (Hermjakob, Montecchi-Palazzi et al. 2004), CluSTr (protein cluster: (Kriventseva, Fleischmann et al. 2001) and PDB (the 3D protein structure database:(Yeats, Lees et al. 2011). For proteins which have a solved 3D structure

Table 7.1 InterPro (IPR) types

Domain	Biological units with defined boundaries, which have structural and functional domains and may include sub-domains
Region	Identifies signatures that cannot be classified either as a family or domain
Repeats	A signature generally, <50 amino acids and can be repeated many times within a sequence
Sites	Conserved site (including motifs) is a short sequence that contain a unique residue/s but cannot be classified under active/binding/or post translation
Active sites	Catalytic pockets of enzymes, where distant parts of a protein may come together to form the catalytic site which may cover one or more signatures. To be included here, empirical data on mutational inactivation data is required
Binding sites	Bind to chemical compounds to facilitate a reaction but are not reaction substrates, i.e. a cofactor, a site needed for electron transport or for protein structure modification, biding is reversible and mutational activation study reports are required
Post translation modifications	Provides activation or de-activation functionality to the protein, e.g. a result of glycosylation, phosphorylation, sulphasion etc., the modification is either permanent or reversible.

in PDB, MODBASE : (Pieper, Webb et al. 2011) or SWISS-MODEL: (Kopp and Schwede 2006), that information is also displayed in the graphical display of the InterPro web interface.

Relationships Between InterPro Entries

Taking into account the protein set mentioned above, if one signature matches with only a subset of that protein set as compared to another signature which matches with the complete set it is likely that the first signature is functionally (and taxonomically) more specific than the second signature. In such a case these two signatures are considered related and the first or the more specific signature is assigned as a child class of the second parent signature. InterProScan entries are developed according to set rules that are described in detail in the InterPro user manual (Hunter, Apweiler et al. 2009). Relationships between InterPro entries provide depth to protein annotation. For example, parent/child relationships described here in gene families can be used to indicate Super families or subfamilies. Parent/child relationships are also permitted for Entry classes Repeats and Sites but not others. Another relationship class Contains/Found-in is applied to Regions, Domains as well as Repeats and Sites and describes the composition of the InterPro entry.

InterPro Implementation

InterProScan, a tool available from InterPro, can be run on a web-based interface (http://ebi.ac.uk/interpro/scan.html) or via a local installation. The current web version requires a protein sequence as the input file (Fig. 7.2). Single sequences may be provided via text box entry, or multiple peptide sequences can be uploaded in any of the standard peptide sequence file formats (i.e. GCG, FASTA, EMBL, GenBank, PIR, NBRF, PHYLIP or UniProtKB/Swiss-Prot) for batch analysis. If working with a DNA sequence the researcher first needs to translate the DNA sequence to its amino acid counterpart using a six fame translation tool such as Transeq (http://www.ebi.ac.uk/Tools/emboss/transeq) and pick the ORF for analysis. In general the query protein sequences used are more than 80 amino acid residues long, since shorter sequences are unlikely to have any signature matches. Once the protein sequence is entered into the input window, the web interface allows the user to select the complete set (or a subset) of member databases to search against (Fig. 7.2). The signature recognition approaches used by current member application databases are described in detail in Table 7.2. As pointed out earlier, diagnostically, these databases have different areas of optimum application owing to the different underlying analysis methods. While all of the methods share a common interest in protein sequence classification, their individual motivations vary amongst divergent domains (e.g. Pfam), functional sites (e.g. PROSITE), and hierarchical family definitions (e.g. PRINTS).

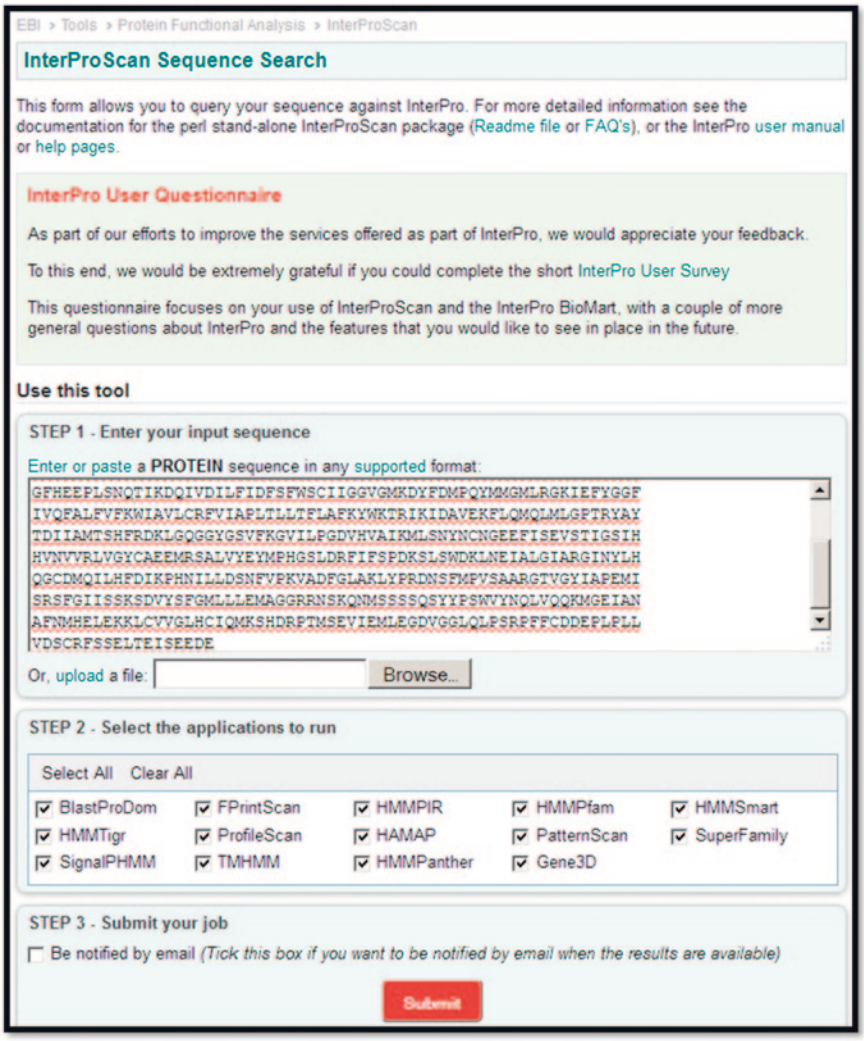

Fig. 7.2 The sequence input page of the InterProScan web interface: In this example the translated sequence (amino acid sequence) of the rice protein LOC-Os01g02350.1 is uploaded into the input window. The applications or the databases against which the amino acid sequence will be scanned is picked by selecting the appropriate boxes at the *bottom* of the window and the job is submitted

Retrieving InterPro Results

On the web interface InterProScan will return the results in a graphical "Picture View" format (Fig. 7.3), and can be viewed or retrieved in Table View, Raw Output, or XPML output. It is important to note that InterProScan calculates a

Table 2 Signature recognition approaches used by current InterProScan member databases

Program name	Description	Abbreviation
BlastProDom	Scans the families in the ProDom database. ProDom is a comprehensive set of protein domain families automatically generated from the UniProtKB/Swiss-Prot and UniProtKB/TrEMBL sequence databases using psi-blast. In InterProScan the blastpgb program is used to scan the database. BLASTPGP performs gapped blastp searches and can be used to perform iterative searches in PSI-BLAST and PHI-BLAST mode	BLASTPRODOM
FPrintScan	Scans against the fingerprints in the PRINTS database. These fingerprints are groups of motifs that together are more potent than single motifs by making use of the biological context inherent in a multiple motif method	FPRINTSCAN
HMMPIR	Scans the hidden markov models (HMMs) that are present in the PIR Protein Sequence Database (PSD) of functionally annotated protein sequences, PIR-PSD	HMMPIR
HMMPfam	Scans the hidden markov models (HMMs) that are present in the PFAM Protein families database.	HMMPFAM
HMMSmart	Scans the hidden markov models (HMMs) that are present in the SMART domain/domain families database	HMMSMART
HMMTigr	Scans the hidden markov models (HMMs) that are present in the TIGRFAMs protein families database	HMMTIGR
ProfileScan	Scans against PROSITE profiles. These profiles are based on weight matrices and are more sensitive for the detection of divergent protein families	PROFILESCAN
HAMAP	Scans against HAMAP profiles. These profiles are based on weight matrices and are more sensitive for the detection of divergent bacterial, archaeal and plastid-encoded protein families	HAMAP
PatternScan	PatternScan is a new version of the PROSITE pattern search software which uses new code developed by the PROSITE team	PATTERNSCAN
SuperFamily	SUPERFAMILY is a library of profile hidden Markov models that represent all proteins of known structure	SUPERFAMILY
SignalPHMM	Predicts the presence and location of signal peptide cleavage sites in amino acid sequences from different organisms	SIGNALP
TMHMM	Predicts transmembrane helices in proteins	TMHMM
HMMPanther	Is a large collection of protein families that have been subdivided into functionally related subfamilies, using human expertise. These subfamilies model the divergence of specific functions within protein families, allowing more accurate association with function (human-curated molecular function and biological process classifications and pathway diagrams), as well as inference of amino acids important for functional specificity. Hidden Markov models (HMMs) are built for each family and subfamily for classifying additional protein sequences. PANTHER is publicly available without restriction	HMMPANTHER

Table 2 (continued)

Program name	Description	Abbreviation
Gene3D	Describes protein families and domain architectures in complete genomes. Protein families are formed using a Markov clustering algorithm, followed by multi-linkage clustering according to sequence identity. Mapping of predicted structure and sequence domains is undertaken using hidden Markov models libraries representing CATH and Pfam domains. Functional annotation is provided to proteins from multiple resources. Functional prediction and analysis of domain architectures is available from the Gene3D website	GENE3D

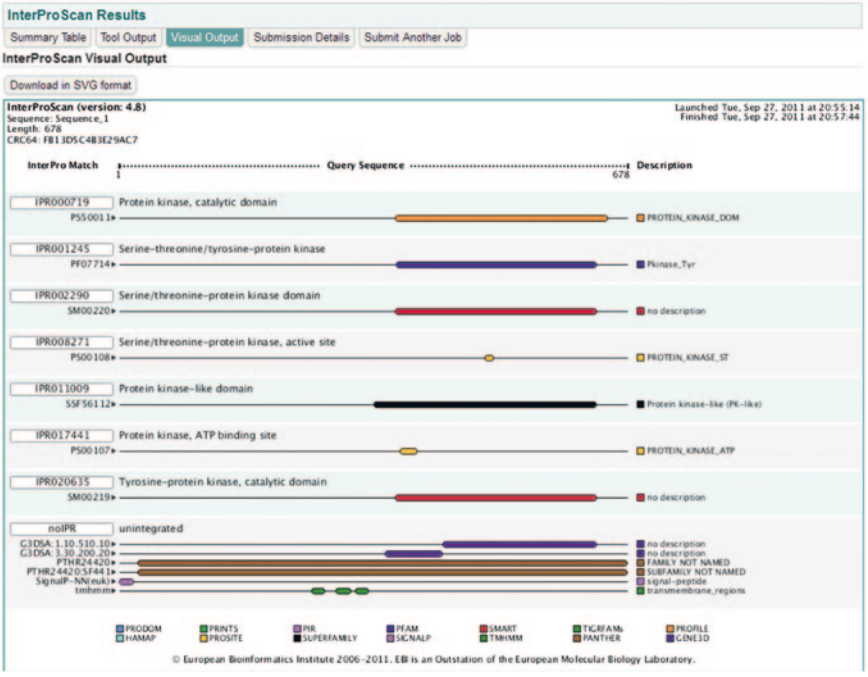

Fig. 7.3 The graphical output from the web interface version of the InterProScan for the rice protein LOC_Os01g02350.1: InterPro entries (number on the left-hand side of the image) and InterPro name (right-hand side of the image) is returned for seven different applications (domain databases). PANTHER, GENE3D,SIGNALP and TMHMM are not included in the integrated InterProScan and therefore no InterPro entries are returned for these applications, but if the corresponding domains are recognized they are marked on the sequence (*bottom* of the window)

checksum for your query sequence and uses it to look up precomputed tables to find an equivalent signature (that equals that checksum). If the particular signature is not integrated, annotation corresponding to that particular signature will not be

returned and therefore you will need to run additional searches for the non integrated signature databases. For bulk data sets such as in a whole genome, a standalone downloadable version is available from ftp.ebi.ac.uk/pub/databases/interpro/iprscan/. In our group we frequently carryout such whole genome level interpro scans for newly sequenced genomes (Shulaev, Sargent et al. 2011). Running InterProScan is computationally intensive and the success of the run is based on the number of input sequences, choice of parameters, data libraries to run against and the availability of compute cluster hardware. A sample output from the stand alone command line local applications is shown in Fig. 7.4.

7.4.2 *Functional Annotation Using Gene Ontology Assignments*

As soon as the genomes of model organisms were sequenced in late 1990s researchers tried to find mechanisms for assigning functions to genes of their newly sequenced organisms and to compare those assignments with that from other organisms. The only common mechanism available at that time was the use of Enzyme commission numbers (EC) (International Union of Biochemistry and Molecular Biology. Nomenclature Committee and Webb 1992). This limited the comparison only to proteins having an enzymatic (catalytic) function and covered only 5–10 % of the genes in the genome. On the other hand, vast majority of proteins are involved in gene regulation, transport, development, growth and response to abiotic and biotic stresses etc., Therefore, it became very apparent that

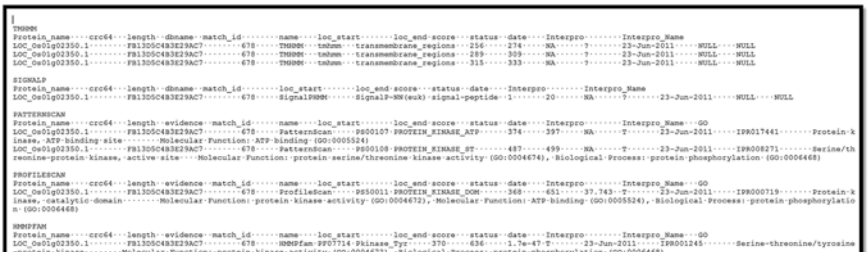

Fig. 7.4 The screen shot of the tabular output from an InterProScan analysis run on an in-house installed, stand- alone version of the program: The rice protein LOC-Os01g02350.1 was analyzed using a locally installed InterProScan tool and domain databases. In this InterProScan run the domain databases PATTERNSCAN, PROFILESCAN, HMMPFAM has recognized functional domains and assigned IPR entries and Interpro names, i.e. IPR017441-Protein Kinase as well as IPR008271-serine/threonine (from PATTERNSCAN), IPR000719-Protei n kinase (from PROFILESCAN), and IPR001245_serine-threonine/tyrosine-protein kinase (from HMMPFAM). These also provide GO assignments for Biological Process and Molecular function. In addition, the domain databases THMM and SIGNALP recognize transmembrane regions and a signal peptide respectively for this protein. Note that our locally installed InterPro tool has only a selected set of five applications as opposed to the web version

the community needed a controlled vocabulary that spanned the breadth of protein functions in order to assign functionality to all genes/proteins. Furthermore functions of these genes also depended on the cellular location of the encoded protein which the EC was unable to address. In response, the Gene Ontology (GO: http://www.geneontology.org/) (Ashburner, Ball et al. 2000) was formed, and a structured, controlled vocabulary (ontology) was developed. Initially it was based on three model organism databases [i.e. FlyBase, Saccharomyces Genome Database and the Mouse Genome Database (Ashburner, Ball et al. 2000)] and since then GO consortium has grown to include many plant, animal and microbial databases. This controlled vocabulary is structured to describe gene products (RNA/Protein) in terms of their role in a biological process (BP), their location in a cellular component (CC), and their molecular function (MF). The GO assignments to genes/gene products are carried out by a set of computational and empirical methods described as follows.

GO Assignments Based on InterPro Assignments

As described before, proteins with similar signatures identified by unique InterPro entries are considered to have common function(s) across all organisms. Protein domains associated with a multiplicity of genes carrying similar experimentally-validated signatures can serve as a reference for assigning function. The InterPro database maintains and shares a list of GO terms assigned to majority of InterPro entries. These mappings include knowledge integrated from genes sharing the same signatures, and therefore they can predict the molecular function (MF), the location (CC) and the role (BP) of a particular protein. However, if the InterPro entries are determined based on computational analysis only they may/may not have GO assignments to them. Mappings between InterPro entries and GO terms, also called *interpro2go* have been generated by the InterPro project team and are available at (http://www.geneontology.org/external2go/interpro2go). Using the *interpro2go* mappings and the InterPro assignments (Fig. 7.4), one can infer GO assignments for the genes of the genome of interest. GO database maintains several such useful mappings listed at http://www.geneontology.org/GO.indices.shtml that can be used for enriching the functional annotation of genes in a genome. An example is the *ec2go* mapping that is used often in assigning EC enzyme role to genes based on GO or vice versa.

GO Assignments Based on Gene Homology

As described above in the gene homology section, in genome annotation there is an option to use the homology based phylogenetic guide trees and gene clusters to project/transfer the GO assignments from a well characterized gene of another species to the gene of interest of the new species. Within the plant kingdom most often *Arabidopsis thaliana* and/or *Oryza sativa* (rice) genes have the most

studied and characterized information on molecular function, expression and the resulting phenotype. Therefore, if the gene cluster has any one of the genes from Arabidopsis and/or rice one has the option to pursue it as a reference and to project similar function to the genes of interest. However, we should note that it is a mere projection of putative function based on sequence similarity. It is therefore denoted with the evidence code to indicate that it is inferred by sequence similarity (ISS) based on the GO evidence code resource list (http://www.geneontology.org/GO.evidence.tree.shtml).

7.5 Use of GO Annotations to Understand the Biology of an Organism

7.5.1 Why is GO Enrichment Analysis Useful?

Initial analysis of most high-throughput genomic or proteomic experiments yields long lists of genes or proteins that are up regulated/down regulated or co-expressed during a treatment or developmental state. In interpreting such long gene lists, manual gene-by-gene interrogation is impractical. On the other hand, GO assignments to genes via MF, BP, CC gives the opportunity to investigate which GO categories are most represented or enriched (Fig. 7.5) under a certain set of conditions (e.g. a treatment or a developmental stage). This is termed GO enrichment analysis. The enrichment analysis therefore, provides the first line of data/dimensionality reduction and biological interpretation capacity for transcriptomic and proteomic data. In

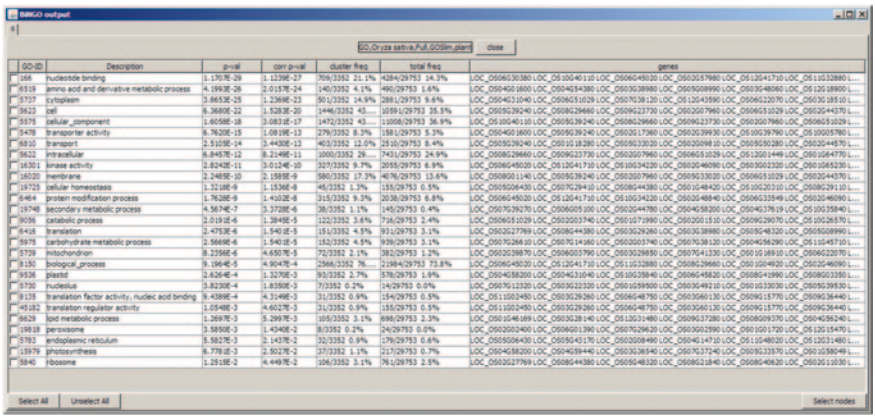

Fig. 7.5 Screen shot of the BinGO tabular output showing statistics of enriched GO categories: Information on over-represented GO function categories include, the p value, adjusted p value, gene number mapping the GO in the query list and background, the total number in query list and background and the IDs of genes within each category

essence enrichment analysis is a statistical method that identifies biological process, cellular compartment and molecular function GO assignments for the genes that may be over or under-represented in a list of genes that is being analyzed.

7.5.2 GO Enrichment Analysis Methods

Over the years many enrichment analysis tools and software packages have been developed (Berriz, Beaver et al. 2009; Tipney H 2010). Gene Ontology consortium website lists links to most of the currently available tools (http://www.geneontology.org/GO.tools). Publicly available applications such as BiNGO (Maere, Heymans et al. 2005), GOStat (Beissbarth and Speed 2004), EasyGO/AgriGO (Du, Zhou et al. 2010) and GOEAST (Zheng and Wang 2008), based on singular enrichment analysis that iteratively tests GO terms one at a time against a list of genes for enrichment, are some of the most popular tools.

7.5.3 Example of GO Enrichment Analysis

As an example here we describe the enrichment analysis of a gene set identified in an experiment on circadian control in rice (Filichkin, Breton et al. 2011). The cycling genes from a time series global expression profiling study were determined using a model-based pattern-matching algorithm (Mockler, Michael et al. 2007). To determine which GO categories are overrepresented in the gene set identified as diurnally cycling between the dark/night and light/day photoperiods, we used the open-source Java tool BiNGO(Maere, Heymans et al. 2005). BiNGO is implemented as a "plugin" to the versatile molecular interaction and network visualization program Cytoscape (Smoot, Ono et al. 2011). This plugin allows the mapping of significant functional categories within the gene set of interest and presents the results on a graphical GO hierarchy.

To analyze against a reference set of GO annotations, BiNGO provides GO annotation for a range of sequenced organisms primarily extracted form GO information available at NCBI (http://www.ncbi.nlm.nih.gov/Ftp/). In this example GO annotations for rice were obtained from Gramene project (http://www.gramene.org/download/index.html) database release 31. If one is analyzing a newly annotated organism or an improved annotation that you want to use, the annotation file created by a user/genome-project by compiling all the GO-gene assignments of your genome of interest can be loaded as a custom annotation in the organism selection menu. BiNGO provides two statistical tests (hypergeometric or binomial) for determining the enrichment degree of a specific GO term in the test gene list. The hypergeometric test, with sampling without replacement gives a probability value for groups of genes belonging to a functional category. The binomial test on the other hand samples with replacement, requires less computation but only provides an approximate P value. To account for multiple hypothesis/statistical tests performed in a single analysis of

majority of high-throughput data, the obtained *P value* has to be corrected for type-1 or false positive error rate. The standard multiple testing correction Bonferroni family-wise error rate correction (the probability of making at least a single type-1 error) was used in this example. There is also the option of using the popular Benjamini and Hochberg false discovery rate (FDR) correction (Benjamini 1995).

Once the statistical test, multiple testing corrections, significance level cutoff (0.05 default), ontology type and the annotation file for the organism is selected, BiNGO calculation will be displayed as a table (Fig. 7.5) and a graphical display as shown in Fig. 7.6). In this specific example GO biological process categories overrepresented in the rice gene set that is regulated by circadian control include, lipid metabolism, carbohydrate metabolic processes, photosynthesis, nucleotide binding, translation, amino acid metabolism and nucleotide metabolism. Similarly, the molecular function categories for kinase activity, transporter and nucleotide binding functions are significantly enriched. The color intensity of the circles are

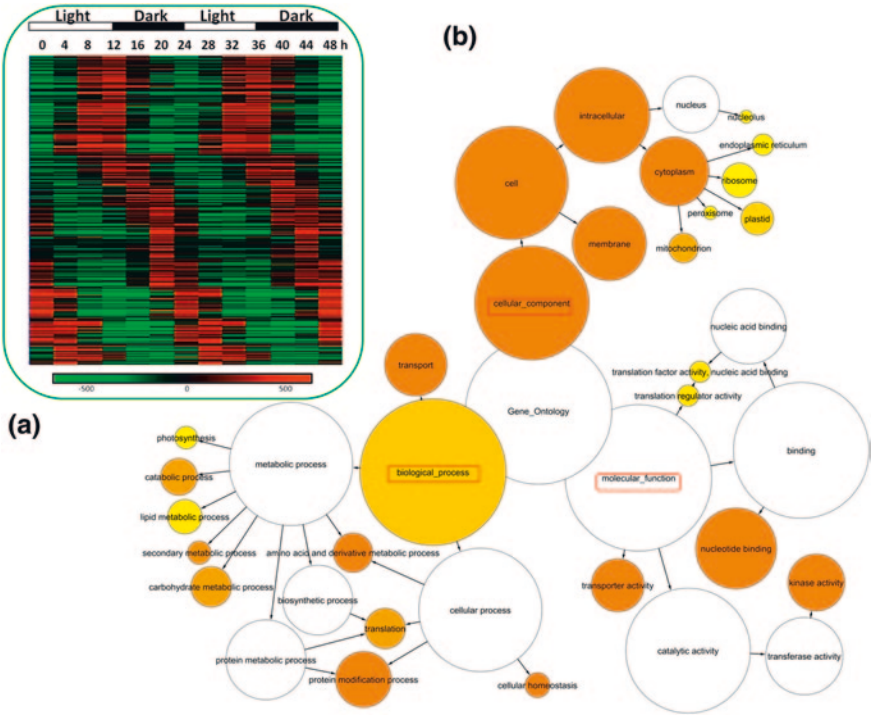

Fig. 7.6 Gene Ontology (*GO*) enrichment analysis: **a** The heat map represents a hierarchical cluster of all cycling genes in rice showing the rhythmic upregulation of genes at different phases of the light/dark cycle. Mean centered expression levels are depicted in *red* for high expression and *green* for low expression. **b** GO categories over-represented in rice cycling genes were analyzed using BinGO, a plugin within Cytoscape. The color intensity of the *circles* indicate the degree of overrepresentation/statistical significance (categories with FDR = 0.05 are shown in *yellow*). The radius of each circle denotes the number of genes in each category

based on the significance level (yellow = FDR below 0.05), and the radius of each circle denotes the number of genes in each category.

7.6 Summary and Outlook

In this chapter, we discussed the structural and functional annotation of plant genomes supported by various empirical and computational methods. By undertaking an integrated approach on plant genome annotation, we think that the context in which the gene function and its structure is annotated adds more confidence with reference to the environment responses, growth, phenotype, cellular and tissue specificity. The approaches such as those listed and described by us confirm that the plant genome is not just a catalog of genes but much more than the sum of its parts and derivatives which the researchers have just started to investigate for answering complex biological questions. We would also like to say that though every effort is made to suggest a putative function using an integrative and comparative approach, there is a lot of dependency on the development, curation, maintenance and updating the library of high quality reference annotations and the algorithms adopted by various annotation tools which continue to evolve. Therefore, no annotation is final unless tested and confirmed by some laboratory based experiment.

Acknowledgments The authors VA, PD, JE and PJ are supported by the Gramene project award (# IOS:0703908) and the Plant Ontology project (# IOS:0822201) from the National Science Foundation (NSF) of USA. The Jaiswal lab is also supported by the startup funds provided to PJ by the Oregon State University (OSU), Corvallis, OR, USA. Authors would also like to thank Rajani Raja of OSU for the InterProScan tabular output (Fig. 7.4); Justin Preece of OSU for editorial comments; and Sarah Hunter, InterPro team and the European Bioinformatics Institute for giving the permission to use screen shots of the InterProScan web interface (Figs. 7.2, 7.3).

References

AGI (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 408(6814):796–815

Al-Dous EK, George B et al (2011) De novo genome sequencing and comparative genomics of date palm (Phoenix dactylifera). Nat Biotechnol 29(6):521–527

Alexeyenko A, Tamas I et al (2006) Automatic clustering of orthologs and inparalogs shared by multiple proteomes. Bioinformatics 22(14):e9–e15

Ashburner M, Ball CA et al (2000) Gene ontology: tool for the unification of biology. The gene ontology consortium. Nat Genet 25(1):25–29

Banks JA, Nishiyama T et al (2011) The selaginella genome identifies genetic changes associated with the evolution of vascular plants. Science 332(6032):960–963

Beissbarth T, Speed TP (2004) GOstat: find statistically overrepresented gene ontologies within a group of genes. Bioinformatics 20(9):1464–1465

Benjamini YH, Yosef (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Roy Stat Soc 57(1):289–300

Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res 27(2):573–580

Berglund AC, Sjolund E et al (2008) InParanoid 6: eukaryotic ortholog clusters with inparalogs. Nucleic Acids Res 36(Database issue):D263–266

Berriz GF, Beaver JE et al (2009) Next generation software for functional trend analysis. Bioinformatics 25(22):3043–3044
Blanco E, Abril JF (2009) Computational gene annotation in new genome assemblies using GeneID. Methods Mol Biol 537:243–261
Blanco E, Parra G et al (2007) Using geneid to identify genes. Curr Protoc Bioinformatics Chapter 4: Unit 4 3
Camacho C, Coulouris G et al (2009) BLAST+: architecture and applications. BMC Bioinf 10:421
Chen F, Mackey AJ et al (2006) OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. Nucleic Acids Res 34(Database issue):D363–368
Cock JM, Sterck L et al (2010) The Ectocarpus genome and the independent evolution of multicellularity in brown algae. Nature 465(7298):617–621
Couch JA, Zintel HA et al (1993) The genome of the tropical tree Theobroma cacao L. Mol Gen Genet 237(1–2):123–128
Du Z, Zhou X et al (2010) AgriGO: a GO analysis toolkit for the agricultural community. Nucleic Acids Res 38(Web Server issue):W64–W70
Filichkin SA, Breton G et al (2011) Global profiling of rice and poplar transcriptomes highlights key conserved circadian-controlled pathways and cis-regulatory modules. PLoS ONE 6(6):e16907
Goff SA, Ricke D et al (2002) A draft sequence of the rice genome (Oryza sativa L. ssp. japonica). Science 296(5565):92–100
Hermjakob H, Montecchi-Palazzi L et al (2004) IntAct: an open source molecular interaction database. Nucleic Acids Res 32(Database issue):D452–D455
Huang S, Li R et al (2009) The genome of the cucumber, cucumis sativus L. Nat Genet 41(12):1275–1281
Hunter S, Apweiler R et al (2009) InterPro: the integrative protein signature atabase. Nucleic Acids Res 37(Database issue):D211–D215
International Union of Biochemistry and Molecular Biology. Nomenclature Committee. and E. C. Webb (1992) Enzyme nomenclature 1992: recommendations of the nomenclature committee of the international union of biochemistry and molecular biology on the nomenclature and classification of enzymes. Published for the International Union of Biochemistry and Molecular Biology by Academic Press, San Diego
IRGSP (2005) The map-based sequence of the rice genome. Nature 436(7052):793–800
Jaillon O, Aury JM et al (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449(7161):463–467
Jurka J, Kapitonov VV et al (2005) Repbase Update, a database of eukaryotic repetitive elements. Cytogenet Genome Res 110(1–4):462–467
Kopp J, Schwede T (2006) The SWISS-MODEL repository: new features and functionalities. Nucleic Acids Res 34(Database issue):D315–D318
Korf I (2004) Gene finding in novel genomes. BMC Bioinf 5:59
Kriventseva EV, Fleischmann W et al (2001) CluSTr: a database of clusters of SWISS-PROT+TrEMBL proteins. Nucleic Acids Res 29(1):33–36
Li L, Stoeckert CJ Jr et al (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res 13(9):2178–2189
Liang C, Mao L et al (2009) Evidence-based gene predictions in plant genomes. Genome Res 19(10):1912–1923
Maere S, Heymans K et al (2005) BiNGO: a cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics 21(16):3448–3449
Merchant SS, Prochnik SE et al (2007) The chlamydomonas genome reveals the evolution of key animal and plant functions. Science 318(5848):245–250
Ming R, Hou S et al (2008) The draft genome of the transgenic tropical fruit tree papaya (carica papaya Linnaeus). Nature 452(7190):991–996
Mockler TC, Michael TP et al (2007) The DIURNAL project: DIURNAL and circadian expression profiling, model-based pattern matching, and promoter analysis. Cold Spring Harb Symp Quant Biol 72:353–363
Mulder N, Apweiler R (2007) InterPro and InterProScan: tools for protein sequence classification and comparison. Methods Mol Biol 396:59–70

O'Brien KP, Remm M et al (2005) Inparanoid: a comprehensive database of eukaryotic orthologs. Nucleic Acids Res 33(Database issue):D476–D480

Ostlund G, Schmitt T et al (2010). InParanoid 7: new algorithms and tools for eukaryotic orthology analysis." Nucleic Acids Res 38(Database issue):D196–D203

Ouyang S, Buell CR (2004) The TIGR plant repeat databases: a collective resource for the identification of repetitive sequences in plants. Nucleic Acids Res 32(Database issue):D360–D363

Paterson AH, Bowers JE et al (2009) The sorghum bicolor genome and the diversification of grasses. Nature 457(7229):551–556

Pieper U, Webb BM et al (2011) ModBase, a database of annotated comparative protein structure models, and associated resources. Nucleic Acids Res 39(Database issue): D465–D474

Potter SC, Clarke L et al (2004) The Ensembl analysis pipeline. Genome Res 14(5):934–941

Rawlings ND, Tolle DP et al (2004) MEROPS: the peptidase database. Nucleic Acids Res 32(Database issue):D160–D164

Remm M, Storm CE et al (2001) Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. J Mol Biol 314(5):1041–1052

Rensing SA, Lang D et al (2008) The physcomitrella genome reveals evolutionary insights into the conquest of land by plants. Science 319(5859):64–69

Sato S, Hirakawa H et al (2011) Sequence analysis of the genome of an oil-bearing tree, Jatropha curcas L. DNA Res 18(1):65–76

Schmutz J, Cannon SB et al (2010) Genome sequence of the palaeopolyploid soybean. Nature 463(7278):178–183

Schnable PS, Ware D et al (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326(5956):1112–1115

Shulaev V, Sargent DJ et al (2011) The genome of woodland strawberry (Fragaria vesca). Nat Genet 43(2):109–116

Smoot ME, Ono K et al (2011) Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics 27(3):431–432

Solovyev V, Kosarev P et al (2006) Automatic annotation of eukaryotic genes, pseudogenes and promoters. Genome Biol 7 Suppl 1:S10 11–12

Spannagl M, Noubibou O et al (2007) MIPSPlantsDB–plant database resource for integrative and comparative plant genome research. Nucleic Acids Res 35(Database issue):D834–D840

Stanke, M. and B. Morgenstern (2005). "AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints." Nucleic Acids Res **33**(Web Server issue): W465-467

Tarailo-Graovac M, Chen N (2009) Using repeatmasker to identify repetitive elements in genomic sequences. Curr Protoc Bioinformatics Chapter 4: Unit 4 10

Tipney HHL (2010) An introduction to effective use of enrichment analysis software. Hum Genomics 4(3):202

Tuskan GA, Difazio S et al (2006) The genome of black cottonwood, Populus trichocarpa (Torr. and Gray). Science 313(5793):1596–1604

Velasco R, Zharkikh A et al (2010) The genome of the domesticated apple (Malus x domestica Borkh.). Nat Genet 42(10):833–839

Vilella AJ, Severin J et al (2009) Ensembl Compara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. Genome Res 19(2):327–335

Vogel (2010) Genome sequencing and analysis of the model grass Brachypodium distachyon. Nature 463(7282):763–768

Yeats C, Lees J et al (2011) The Gene3D Web Services: a platform for identifying, annotating and comparing structural domains in protein sequences. Nucleic Acids Res 39(Web Server issue):W546–W550

Youens-Clark K, Buckler E et al (2011) Gramene database in 2010: updates and extensions. Nucleic Acids Res 39(Database issue): D1085–D1094

Zheng Q, Wang XJ (2008). GOEAST: a web-based software toolkit for gene ontology enrichment analysis. Nucleic Acids Res 36(Web Server issue): W358–W363

Chapter 8
Different Omics Approaches in Cereals and Their Possible Implications for Developing a System Biology Approach to Study the Mechanism of Abiotic Stress Tolerance

Palakolanu Sudhakar Reddy and Nese Sreenivasulu

8.1 Introduction

Cereals comprise a number of crops including rice, wheat, maize, barley, rye and sorghum. In the form of starch and proteins, the cereal grains provide nearly 60 % of the calories consumed globally as food and fodder. There is a growing challenge to meet the global demand of food security for a human population of 9 billion expected by the year 2050 (Royal 2009; Sreenivasulu and Schnurbusch 2012). Current predicted climatic conditions such as prolonged drought and heat episodes pose a serious threat for the agricultural production world-wide, affecting yield losses estimated at billions of dollars (Mittler 2006; IPCC 2007; Battisti and Naylor 2009). Hence, increasing crop productivity in view of escalating population as well diminishing cultivable land and natural resources in such challenging environmental conditions has become a matter of urgency. Although much research has been conducted to evaluate the effects of global warming due to a variety of human activities (Smit et al. 1988), efforts to search specific and practical approaches to improve adaptability of plants to the climate change have only begun recently (Charng et al. 2006; Montero-Barrientos et al. 2010).

Abiotic stresses lead to a series of changes in the plant that affect molecular, biochemical, physiological and phenological processes eventually affecting the performance of plant growth and development impacting overall yield (Wang et al. 2003; Sreenivasulu et al. 2007). Plants that successfully withstand stresses are constantly monitoring their external milieu and are redefining the appropriate cellular response. It depends on the ability of the plants to be equipped with intricate gene regulatory mechanisms leading to the appropriate physiological adaptation to survive harsh challenging conditions. Therefore, understanding plant abiotic stress responses is

P. S. Reddy · N. Sreenivasulu (✉)
Interdiciplinary Center for Crop Plant Research (IZN) Research Group Stress Genomics, Leibniz-Institute of Plant Genetics and Crop Plant Research (IPK), Corrensstraße 3, 06466 Gatersleben, Germany
e-mail: srinivas@ipk-gatersleben.de

P. K. Gupta and R. K. Varshney (eds.), *Cereal Genomics II*,
DOI: 10.1007/978-94-007-6401-9_8,

now thought to be one of the most important topics in plant science. Different omics-approaches have been used to elucidate some of the key regulatory pathways in plant responses to abiotic stresses. The plant physiological and molecular responses to abiotic stresses have been investigated using various genomics strategies (Vij and Tyagi 2007; Collins et al. 2008; Hu et al. 2009), which include transcriptomics (Rostoks et al. 2005; Mohammadi et al. 2007; Zeller et al. 2009), proteomics (Qureshi et al. 2007; Caruso et al. 2009) and metabolomics (Shulaev et al. 2008). For a comprehensive understanding of global response we need to integrate these responses at a systems level and need to build integrative platforms to derive knowledge, which may facilitate development of stress tolerance in crop plants.

A systems biology/omics approach is a new upcoming field in plant biology, which allows not only a better understanding of molecular processes and cellular function (Kitano 2000), but also to identify the molecular targets for crop improvement (Cramer et al. 2011). One of the key challenges of systems biology is to integrate the different omics information to give a more complete picture of living organisms. Such an integrated approach would unravel the complex interplay or cross-talk between the different components and to understand the dynamic activities of a tissue/organ/organism in different environments (Cramer et al. 2011). The availability of these data in model species not only allowed a comprehensive understanding of responses against abiotic stresses, but eventually will make the way forward to identify key targets for engineering abiotic stress tolerance in cereals.

8.2 Status of Genome Sequences in Cereals

The genome sequence, often referred to as the genetic blueprint, provides a foundation for connecting the information from the genome to the phenome via structural and functional genomics with an extended approach of systems biology. The development of genomic resources has progressed in a number of plant species, thus creating the gold standard reference genomes in several crops of the grass family including rice, maize, Brachypodium and sorghum. Despite variation in genetic diversity, genome size and chromo-some number, there is substantial conservation in gene order between the grasses which is explored through the study of synteny and collinearity. Extensive data on all aspects of cereal genomics are now available at GrainGenes (http://wheat.pw.usda.gov/) and Gramene (http://gramene.org/), the latter having a major emphasis on rice genome and its syntenic relationship with other cereal genomes. Here, we briefly review the current status of available genomic sequences for cereal crop species (Table 8.1).

8.2.1 Rice

Among cereals, the first draft sequence is released in rice in two sub-species japonica and indica (the two subspecies of Asia) by the commercial effort from Syngenta, USA (Goff et al. 2002) and public academic effort by Beijing

Table 8.1 Whole genome sequencing projects in cereals

Common Name	Species	Genome size(Mbp)	Database	Sequencing group
Rice	*Oryza sativa* (japonica)	420	http://rgp.dna.affrc.go.jp/IRGSP/index.html	IRGSP
Rice	*Oryza sativa* (indica)	466	http://rice.genomics.org.cn/rice/index2.jsp	Beijing Genomics Institute
Barley	*Hordeum vulgare*	5,500	http://www.public.iastate.edu/~imagefpc/IBSC%20Webpage/IBSC%20Template-home.html	Consortium (IBSC)
Barley	*Hordeum vulgare*		http://webblast.ipk-gatersleben.de/barley/index.php	
Purple false brome	*Brachypodium*	272	http://www.brachypodium.org/	JGI, Consortium (IBI)
Maize	*Zea mays*	2,500	http://www.maizegdb.org/	Consortium
Wheat	*Triticum aestivum*	16,500	http://www.wheatgenome.org/ http://urgi.versailles.inra.fr/index.php/urgi/Projects/3BSeq http://www.cshl.edu/genome/wheat http://www.cerealsdb.uk.net/	Consortium (IWGSC)

Genomics Institute, China (Yu et al. 2002), respectively. This effort resulted in generating whole-genome shotgun (WGS) genome draft of japonica and indica, covering more than 90 % of the 420 megabase (Mb) genome and also suggested that genome size is increased by >6 % and >2 %, respectively compared to the common ancestor. Also these two sub-species showed a genetic divergence through a detection of numerous SNPs, indels within both the unique (coding) and the repetitive regions. In 2005, IRGSP released high quality map-based draft sequence in the public domain by providing indexing of 37,544 protein coding genes (International Rice Genome Sequencing Project 2002). Gene predictions developed by the Plant Genomics Group at TIGR (http://rice.plantbiology.msu.edu/) and RAP-DB released the rice genome annotation for the public use (Tanaka et al. 2008). The Rice FOX (full-length cDNA overexpressor) gene hunting system is a resource of gain-of-function mutants where 13,000 full-length rice cDNA clones are overexpressed in Arabidopsis (rice FOX Arabidopsis lines, http://ricefox.psc.riken.jp/) to characterize gene functions in a heterologous system (Kondou et al. 2009; Sakurai et al. 2011). By this way, several full-length cDNAs from rice were shown to represent function of orthologous genes in Arabidopsis as a FOX line mutant collection with interesting phenotypes (Sakurai et al. 2011).

8.2.2 Maize

Maize is an important model C_4 cereal crop that is predominantly a cross-pollinating, a feature that has contributed to its broad morphological variability and geographical adaptability. Maize genome size is estimated to be 2,500 Mb, which is six times bigger than the rice genome, owing to the expansion of families of transposable elements, particularly retrotransposons (Berhan et al. 1993). The maize genome size has expanded dramatically (up to 2.3 Gb) over the last ~3 million years via a proliferation of long terminal repeats of retrotransposons (SanMiguel et al. 1998). Comparative analysis of grass genomes also reveals conservation of gene order but some local rearrangements interrupt collinearity at molecular level (Feuillet and Keller 2002). These rearrangements often prevent maize gene cloning using other cereals genome sequence information as a reference. Thus, having completed maize genome sequencing is extremely beneficial to better understand gene and genome structure of rice and maize, and to understand the evolution of complex grass genomes. The draft genome of maize B73 has been sequenced (Schnable et al. 2009) using a minimum tiling path of bacterial artificial chromosomes (BACs) (16,848) and fosmid (63) clones derived from an integrated physical and genetic map (Wei et al. 2009), augmented by comparisons with an optical map (Zhou et al. 2009). Shotgun sequenced clones covered up to 4–6 fold genome and followed by automated and manual sequence improvement of the unique regions only, which resulted in the B73 reference genome version 1(B73 RefGen_v1). This B73 RefGen_v1 contains 855 families of DNA transposable elements that make

up 8.6 % of the genome. From the genome sequence information 32,540 protein-encoding genes and 150 microRNA (miRNA) genes were predicted from assembled B73 RefGen_v1. Exon sizes of maize genes were similar to that of their orthologous genes in rice and sorghum, but maize genes contained larger introns because of insertion of repetitive elements (Wei et al. 2009; Haberer et al. 2005). In future, exploring intraspecific gene variability and a study of the role of epigenetics and retrotransposons will remain an important exercise to resolve the hybrid vigour and plant performance in maize.

8.2.3 Brachypodium

The whole genome sequence of Brachypodium reveals that relative to other grass genomes, Brachypodium genome is compact (272 Mb), with retrotransposons concentrated at the centromeres and at the collinearity breakpoints. A total of 25,532 protein-coding genes were predicted in the v1.0 annotation. This is in the same range as sorghum (27,640) (Paterson et al. 2009). Between 77 and 84 % gene families are shared among the three grass subfamilies represented by Brachypodium, rice and sorghum, reflecting a relatively recent common origin (The International Brachypodium Initiative 2010). The similarities in gene content and gene family structure between Brachypodium, rice and sorghum support the value of Brachypodium as a functional genomics model for all grasses. The relatively small genome of Brachypodium contains many active retroelement families, but recombination between these retroelements keeps genome expansion in check. Because of small size and rapid life cycle, and its genetic proximity to tribe Triticeae, Brachypodium has several advantages. The small size of some accessions makes it convenient for cultivation in a small space. This has led to the development of highly efficient transformation systems for a range of Brachypodium genotypes (Vain et al. 2008; Vogel and Hill 2008; Alves et al. 2009). Also several important resources have been developed, which includes germplasm collections (Vogel and Hill 2008; Filiz et al. 2009; Vogel et al. 2009), genetic markers (Vogel et al. 2009), a genetic linkage map (Garvin et al. 2010), bacterial artificial chromosome (BAC) libraries (Huo et al. 2006, 2008), physical maps (Gu et al. 2009), large-scale collection of T-DNA tagged lines termed 'the BrachyTAG program' mutant collections (Thole et al. 2010), microarrays and databases (Table 8.2). These resources are facilitating the use of Brachypodium by the research community, and will allow Brachypodium to be used as a powerful functional genomics resource for grasses. Since Brachypodium is more closely related to the Triticeae (wheat, barley) than to the other cereals, Brachypodium genome also helps in the genome analysis and gene identification in the large and complex genomes of Triticeae tribe (wheat and barley), which are among the world's most important crops. It is also an important advance in grass structural genomics permitting for the first time, whole-genome comparisons between members of the three most important grass subfamilies.

Table 8.2 Omics related resources in cereals and the corresponding URLs links

S.No	Omic resources	URLs	Species
1	Transcriptomics	https://www.genevestigator.com/gv/	Plant species
		http://bar.utoronto.ca/welcome.htm	Cereals
		http://www.plexdb.org/index.php	Cereals
		http://mpss.udel.edu	Cereals
		http://contigcomp.acpfg.com.au	Wheat, barley
		http://bioinformatics.med.yale.edu/riceatlas/	Rice
2	Metabolomics	http://mapman.gabipd.org/web/guest	Cereals
		http://pathway.gramene.org/	Brachypodium, rice, maize
		http://www.genome.jp/kegg/pathway.html	Brachypodium, rice, maize
		https://www.metabolome-express.org/	Rice, wheat
		http://www.cbib.u-bordeaux2.fr/MERYB/	Rice, maize
		http://www.plantcyc.org/	Rice, maize
3	Proteomics	http://ppdb.tc.cornell.edu/	Maize, rice
		http://www.p3db.org/	Rice, maize
		https://database.riken.jp/sw/en/Plant_Phosphoproteome_Database/ria102i/	Rice
		http://cdna01.dna.affrc.go.jp/RPD/main_en.html	Rice
4	Phenomics	http://wheat.pw.usda.gov/GG2/germplasm.shtml#collections	Wheat, barley
		http://tilling.ucdavis.edu/index.php/Main_Page	Rice, wheat
		http://www.postech.ac.kr/life/pfg/risd/	Rice
		http://tos.nias.affrc.go.jp/	Rice
		http://ricefox.psc.riken.jp/	Rice
		http://orygenesdb.cirad.fr/	Rice
		http://www.shigen.nig.ac.jp/rice/oryzabase/top/top.jsp	Rice
		http://barley.ipk-gatersleben.de/ebdb/	Barley

(continued)

Table 8.2 (continued)

S.No	Omic resources	URLs	Species
		http://ace.untamo.net/	Barley
		http://barleygenomics.wsu.edu/mut-4-3-2.html	Barley
		http://www.maizegdb.org/stock.php	Maize
		http://www.brachytag.org	Brachypodium
5	Integrative omics	http://prime.psc.riken.jp/	
		http://www.phytozome.net/	Plant species
		http://www.plantgdb.org/	Plant species
		http://plants.ensembl.org/index.html	Plant species
		http://chloroplast.cbio.psu.edu/	Plant species
		http://www.genome.jp/kegg/plant/	Plant species
		http://pathway.gramene.org/expression.html	Gramineae
		http://wheat.pw.usda.gov/GG2/index.shtml	Triticeae, avena
		http://kpv.kazusa.or.jp/kappa-view/	Rice
		http://rice.plantbiology.msu.edu/	Rice
		http://rapdb.dna.affrc.go.jp/	Rice
		http://www.maizegdb.org/	Maize
7	Full-length cDNA	http://www.shigen.nig.ac.jp/barley/	Barley
		http://cdna01.dna.affrc.go.jp/cDNA/	Rice (japonica)
		http://www.ncgr.ac.cn/ricd	Rice (indica)
		http://trifldb.psc.riken.jp/	Wheat
		http://www.maizecdna.org/	Maize

8.2.4 Barley

Barley (*Hordeum vulgare* L.) ranks fourth among the cereals in worldwide production and due to its broad stress tolerance adaptability, high genetic variability and close relationship to wheat and rye, barley is considered as an excellent model C_3 crop of Triticeae (Koornneef et al. 1997; Hayes et al. 2003; Sreenivasulu et al. 2008a). Barley genome comprises seven chromosomes with estimated genome size of 5,100 Mb (12 times that of rice) of which 80 % of genome is composed of repetitive DNA, which is presently a major challenge to decipher the complete genome. The systematic efforts for sequencing the whole barley genome were initiated in 2006 by International Barley Sequencing Consortium (IBSC) (http://www.public.iastate.edu/~imagefpc/IBSC%20Webpage/IBSC%20Template-home.html) and the cultivar Morex was recommended as a reference genome. Several approaches are being used to unlock the gene content in the whole genome by next-generation sequencing of sorted chromosomes, sequencing of gene-rich BAC clones and full-length cDNA collections (Sreenivasulu et al. 2008b; Mayer et al. 2011; Schulte et al. 2011). As a result, Barley Sequencing Consortium is continuously generating voluminous sequencing data that is accessible from the website (http://webblast.ipk-gatersleben.de/barley/index.php). A novel analytical platform is also available for genome-wide SNP genotyping (9 K Infinium array) for barley and has been used to survey genomic variations among barley germplasm and to evaluate chromosomal distribution of introgressed segments of near-isogenic lines. Also several transcriptome platforms are available to generate genome wide transcriptome atlas (Druka et al. 2006, 2011; Sreenivasulu et al. 2006, 2008a). Natural variants among barley collections were used to investigate the associations between nucleotide haplotypes and growth habits that are witnessed in different geographical distribution (Saisho and Takeda 2011; Pasam et al. 2012).

8.2.5 Wheat

Wheat is the most widely grown and important staple cereal crop, which occupies more arable land (17 % of all crop area) and possesses more market share ($31 billion) than any other cereal crop (Gupta et al. 2008; Safar et al. 2010). Wheat is a hexaploid, with A, B and D subgenomes, the entire genome being 40-fold larger than the rice genome (Arumuganathan and Earle 1991) and each individual subgenome being ~5,500 Mb in size. The large genome size, hexaploid nature and a high proportion of repetitive DNA creates significant challenges in elucidating its genome sequence and to connect genome sequences to the phenotypic variance of agronomic traits (Chantret et al. 2005; Paux et al. 2008; Wanjugi et al. 2009). International wheat genome sequencing consortium (IWGSC) has began to target a complete high quality genome

sequence, by adopting a chromosome-based strategy to construct physical BAC clone maps and subsequently to sequence each of the individual chromosomes (Dolezel et al. 2007). In this context, around 68,000 BAC clones of a 3B chromosome-specific BAC library (Safar et al. 2004; Paux et al. 2008) have been fingerprinted at the French National Sequencing Centre and the sequencing of these BAC clones is under progress (http://urgi.versailles.inra.fr/index.php/urgi/Projects/3BSeq). Several approaches have been initiated to sequence the complex wheat genome. For instance, the consortium from UK produced 5X sequence of the bread wheat genome using Roche 454 technology (http://www.cerealsdb.uk.net/), and also produced a draft wheat genome assembly from the donor species of the wheat D genome, *A. tauschii* (http://www.cshl.edu/genome/wheat). Sequences from individual flow-sorted bread wheat chromosome arms are also piling up gradually (Berkman et al. 2011; Wicker et al. 2011). With the increased availability of wheat genome sequence data, it is necessary to provide resources that can integrate wheat-specific sequence information to become useful for crop improvement (Edwards and Batley 2010). Since wheat genome sequencing is still in progress, and a high quality genome sequence is expecting by 2015, one can foresee the possibilities of launching systems biology approaches even in barley and wheat. These systematic attempts to move from genomic to post-genomic strategies greatly facilitate researchers who wish to use this information to improve this valuable crop. The update about the genome sequencing project information and other genetic resources are listed in the Table 1.

Evaluating the impact of genome organization, monitoring dynamic alteration of retrotransposons, assessing the impact of epigenetic hallmarks by covering genome wide DNA methylation and omics driven systems biology approaches are all part of genome dynamic applications. In this review we focus on transcriptome, proteome and metabolome data available in cereals and other model species. Further we discuss the future needs of implementing systems biology applications to derive work flow to identify key target genes for crop improvement.

8.3 Omics Revolution by High Throughput Approaches

Major progress made in the last decade is through the use of new high-throughput techniques not only in the field of whole genome sequencing but also through characterization of genes through functional genomics. Systematic use of different omics approaches such as transcriptomics, proteomics, metabolomics, fluxome and a way forward to connect the global data to the phenotypic variance (generated through phenomics) have led to expand the area towards systems biology for elucidating the mechanisms underlying the expression of agronomic traits. System-based approaches based on a combination of multiple omics analyses has been an efficient approach to determine the global picture of cellular systems and to reveal the plant responses and adaptation to a specific stress. In this context, the integrated approaches with multiple-omics data should contribute greatly to the

identification of key regulatory steps and to characterize the pathway interaction in various processes. These illustrative examples demonstrate the power of multi-omics-based systems analysis for understanding the key components of cellular systems underlying various plant functions. The integration of a wide spectrum of omics datasets from various plant species is then essential to promote translational research to engineer plant systems in response to the challenges of emerging climate change.

8.3.1 Transcriptomics

Genome-wide transcriptome profiling is a powerful approach to assemble a transcriptome atlas of expressed genes involved in various biological phenomena and to reveal the molecular cross-talk of gene regulatory networks of responses to various abiotic stresses. Microarray analysis is known to be an important approach to elucidate the molecular basis of the plant stress response (Van Baarlen et al. 2008; Deyholos 2010). The investigation of gene expression related to several physiological and agronomical traits have been reported in different cereals. These responses include the following: responses to hormones (Seki et al. 2002b; Rabbani et al. 2003), various stress responses (Kreps et al. 2002; Rabbani et al. 2003; Takahashi et al. 2004), including drought (Kreps et al. 2002; Oono et al. 2003; Rabbani et al. 2003), cold (Kreps et al. 2002; Rabbani et al. 2003; Yamaguchi et al. 2004), high light (Rossel et al. 2002; Kimura et al. 2003), hyperosmolarity, oxidative stress (Takahashi et al. 2004), and iron deficiency (Thimm et al. 2001).

More detailed and comprehensive gene expression studies have been conducted in the model species like *Arabidopsis* and rice, and the resulting knowledge can be used in cereals through comparative gene networks. In case of cereals, several data repositories have been created to store the raw data and normalized expression values generated from GeneChip arrays including Affymetrix 57 K from Rice, 61 K Wheat, 22 K Barley1, full-genome Brachypodium and Maize arrays. Furthermore, these databases not only allow storage of data from Affymetrix platform but also allow storing data from Agilent and NimbleGen platforms (Sreenivasulu et al. 2010). These databases include PLEXdb, GEO, Genevestigator, UniProt, PlantGDB (Bombarely et al. 2011) Gramene (Youens-Clark et al. 2011), TAIR (Swarbreck et al. 2008) and MaizeGDB (Schaeffer et al. 2011).

Transcriptome studies have also been carried out in cereals and other model plants but mainly applying single stress at a time such as drought, salinity, cold or heat during the vegetative state (for recent reviews see Ingram and Bartels 1996; Sreenivasulu et al. 2004a, 2007; Kishor et al. 2005; Vij and Tyagi 2007; Fleury et al. 2010). Interestingly, unique stress responsive pathways such as osmolyte metabolism, antioxidant machinery, dehydrin and LEA proteins, chaperones and gene machinery involved in protection of cell integrity are preferentially upregulated in both dicots and monocots (Xue et al. 2006; Ergen et al. 2009; Fleury et al. 2010; Sreenivasulu et al. 2010). However, within osmolyte metabolism, wide

array of biochemical pathways are known to activate preferentially in a species and genotype specific manner, which corresponds to compounds proline, mannitol, myo-innositol, trehalose, glycine metabolism, accumulation of sugar alcohols and free sugars including fructose metabolism. Additionally, some studies have identified abundance of various transcripts during heat treatment, including genes encoding for galactinol synthase and enzymes in the raffinose oligosaccharide pathway, and antioxidant enzymes (Lim et al. 2006; Xu et al. 2007). Comparison of transcript profiles between tolerant and susceptible lines under various stress responses has revealed differences in stress-responsive pathways reflecting difference in physiological response and adaptation behavior. Transcriptome analysis also revealed some unexpected results such as a decrease in the expression of glutathione-related genes following withholding of water in a tolerant synthetic wheat line (Mohammadi et al. 2007), or the accumulation of proline in a drought-sensitive emmer wheat line (Ergen and Budak 2009), suggesting that some pathways/mechanisms are dependent upon genotype, the duration, intensity, and type of stress applied. There are some reports, which show decrease in transcript abundance related to programmed cell death, basic metabolism, and biotic stress responses (Larkindale and Vierling 2008) under heat stress conditions. Recently, Pinheiro and Chaves (2011) reviewed 450 research papers on drought-mediated changes in photosynthesis.

Until now most of the transcriptome responses have been studied in vegetative tissues and recently few attempts were made to reveal the transcriptome alterations in developing seeds to understand the yield stability. In case of cereals, transcriptome analyses were recently applied to analyze rice developing caryopses under high temperature conditions (Yamakawa and Hakata 2010) and seed developmental alterations in barley under drought (Worch et al. 2011). Overall, several extensive attempts have been made to identify several genes/pathways in a number of cereal crops including rice (Amudha and Balasubramani 2011; Hadiarto and Tran 2011; Yang et al. 2010). However, any deeper and/or new insights into mechanisms of the function of genes were missing. In combination with these reviews, the present review of literature based on transcriptome studies should present a pertinent update on genes involved in abiotic stress tolerance in crop plants. The effort is to lend a perspective on how different pieces may fit into the complicated puzzle and to present the integrated view on abiotic stress tolerance.

8.3.2 Proteomics

Although transcriptomics data provides an useful overview of global gene expression regulation, proteomics is often used as a complementary technique that provides the actual state of the condition of cell response to stress. Moreover, proteomics is considered as an essential bridge between the transcriptome and the metabolome (Wasinger et al. 1995; Zhu et al. 2003). Compared to transcriptome

analysis, proteomics approach has a close relationship to phenotype because of their direct action on several biochemical processes. This approach is important in evaluating stress responses since the mRNA levels may not always correlate with protein accumulation (Gygi et al. 1999) and moreover several regulatory proteins are subjected to proteolysis to fine tune the dynamics of transcribed machinery. Despite this strategic importance, compared to transcriptomics analysis, plant proteome response to abiotic and biotic stresses is still limited.

In the last decade, good progress has been made in the separation of proteins and their identification by mass spectrometry. Studies have evaluated changes in protein levels of plant tissues in response to stresses (Canovas et al. 2004; Kim et al. 2003). However, these studies have mainly focused on model species such as Arabidopsis and rice (Canovas et al. 2004). Implication of proteomic studies in cereals is mainly based on rice as a model species (Agrawal and Rakwal 2006, 2011; Komatsu and Yano 2006). A proteomic analysis of drought and salt-stressed rice plants found that around 3000 proteins could be detected in a single gel and over 1,000 could be analyzed (Salekdeh et al. 2002). The effect of salt stress on young rice panicles has been investigated by the same group (Dooki et al. 2006). The proteomic analysis of rice leaf sheaths during drought stress identified 10 up-regulated and two down-regulated proteins. Among the up-regulated proteins, one was an actin depolymerizing factor present at high levels in the leaves of non-stressed drought-resistant cultivars (Ali and Komatsu 2006). Proteome reference maps have been compiled for maize (Mechin et al. 2004) and wheat (Vensel et al. 2005) endosperm and for barley grain (Finnie et al. 2002) during the processes of grain filling and maturation. The effect of heat stress on the grain of hexaploid wheat has been thoroughly studied at the protein level and down-regulation of several proteins involved in the starch metabolism and the induction of HSPs was reported (Majoul et al. 2003, 2004). The effect of drought on the wheat grain proteome, involved 121 proteins that exhibited significant changes in response to the stress; 57 of these 121 proteins could be identified (Hajheidari et al. 2007). Two-thirds of the identified proteins turned out to be thioredoxin targets, revealing the link between drought and oxidative stresses. Changes in the protein complement have been monitored in maize under progressive water deficit and several genes/proteins were reported to be involved in the drought response (Riccardi et al. 1998). The high level of genetic variability observed at the proteome level for the drought response in maize (de Vienne et al. 1999) allowed identification of *Asr*1 (ABA/water-stress/ripening-related1) gene as a candidate for genetic improvement (Jeanneau et al. 2002). Apart from this, some proteomics resources are also available for grasses, such as the plant proteome database (http://ppdb.tc.cornell.edu/) which provides information on the maize and *Arabidopsis* proteomes. RIKEN Plant Phosphoproteome Database (RIPP-DB, http://phosphoproteome.psc.database.riken.jp) was updated with a data set of large-scale identification of rice phosphorylated proteins (Nakagami et al. 2010, 2012). The OryzaPG-DB was launched as a rice proteome database based on shotgun proteomics (Helmy et al. 2011). Although only a handful of studies have been carried out in cereal

crops, it is expected to have a significant increase in the implementation of these techniques in cereal crops to study genome wide protein–protein interactions.

8.3.3 Metabolomics and Fluxome

Metabolomics is one of the important component of functional genomics. It defines the quantitative metabolite signatures present in a cell/tissue under a given set of physiological conditions (Oliver et al. 1998; Kell et al. 2005; Jordan et al. 2009). Higher plants have the remarkable ability to synthesize a vast array of compounds that differ in the chemical complexity, structure and biological activity, playing indispensable roles in chemical defenses against biotic and abiotic stresses (Verpoorte and Memelink 2002; Dixon and Strack 2003; Schwab 2003). Moreover, under various stress conditions, crop species are known to modulate the primary metabolism due to the impaired photosynthesis and respiration events. The main advantage of metabolomics is that it allows one to measure the impact of metabolism and to interlink the key metabolic signatures to the phenotype.

Study of metabolic regulation during stressful conditions has been facilitated through mass spectrometry-based analytical methods resulting in the detection and identification of diverse metabolites (Sawada et al. 2009). Metabolite profiling deals with detection of a wide range of metabolites in diverse concentrations, which makes their analysis more complicated. Therefore, more comprehensive coverage can only be achieved by using multi-parallel complementary extraction and detection technologies subjected to chemical analysis using liquid and gaseous chromatography-mass spectrometry (LC–MS and GC–MS), nuclear magnetic resonance (NMR) and Fourier transform-infrared spectrometer (FT-IR).

Metabolome analyses of model plants have markedly increased in the recent decade and helped to understand the plant response to various stresses. To obtain deeper view into cellular conditions under abiotic stresses, metabolomic investigations have been performed initially in model species like *Arabidopsis* and other plant species (Schauer and Fernie 2006). From the genome sequence information of the *A. thaliana*, it is evident that plants appear to re-organize their metabolic network in order to adapt to such conditions (Kaplan et al. 2004). Therefore, metabolomics plays a key role in understanding cellular functions and decoding the functions of genes under challenging abiotic stress conditions (Fiehn 2002; Bino et al. 2004; Oksman-Caldentey and Saito 2005; Hall 2006; Schauer and Fernie 2006; Hagel and Facchini 2008; Saito et al. 2008). Metabolic adjustments in response to different stress conditions are dynamic and multifaceted because of their intensity and nature of the stress, but it also depends on the cultivar and the type of plant species. This approach also covers the extensive comprehensive metabolite analyses, illustrating the complexity of metabolic adjustments to different abiotic stresses (Rizhsky et al. 2004; Urano et al. 2009) including salinity (Cramer et al. 2007; Kempa et al. 2007; Sanchez et al. 2008; Janz et al. 2010; Lugan et al. 2010), and temperature stress (Cook et al. 2004; Rizhsky et al. 2004;

Kaplan et al. 2007; Usadel et al. 2008; Espinoza et al. 2010; Caldana et al. 2011). Some metabolic changes are common to salt, drought, and temperature stress, whereas others are specific to particular stress (Gong et al. 2005; Cramer et al. 2007; Gagneul et al. 2007; Kempa et al. 2008; Sanchez et al. 2008; Usadel et al. 2008; Urano et al. 2009; Lugan et al. 2010). Metabolomic profiles illustrate that plants have developed a wide range of strategies to adapt their metabolism to unfavorable growth conditions and that enhanced stress resistance is not restricted to a single compound or mechanism. Several metabolites/metabolic pathways that contribute to stress acclimation also play a role in development (Hanzawa et al. 2000; Samach et al. 2000; Eastmond et al. 2002; Palanivelu et al. 2003; Imai et al. 2004; van Dijken et al. 2004; Alcazar et al. 2005; Gupta and Kaur 2005; Satoh-Nagasawa et al. 2006; Mattioli et al. 2008, 2009; Szekely et al. 2008; Deeb et al. 2010; Zhang et al. 2011).

Surprisingly, metabolomic research has made a limited progress in cereals. A recent metablolome study in rice identified 88 metabolites from the extract of leaves. It was found that sugar and amino acid metabolism is dynamically altered under stress treatment (Sato et al. 2008). Metablolome study from maize kernels showed wide range of natural variability based on the influence of genetic background and growing season (Reynolds et al. 2005), developmental stages (Seebauer et al. 2004) and environment (Harrigan et al. 2007). Metabolome study of diverse maize genotypes recently explored and highlighted the importance of grain fatty acid methyl esters, free fatty acid methyl esters, free amino acids. Around 167 metabolites were identified from 300 distinct analytes by using GC–MS approach (Rohlig et al. 2009). Integrated metabolome and transcriptome analysis has also been applied to investigate changing metabolic systems in plants growing in field conditions, such as the rice Os-GIGANTEA (Os-GI) mutant and transgenic barley (Kogel et al. 2010; Izawa et al. 2011). The application of metabolomics in cereals has just begun, and its full potential will be realized only in future. Large-scale metabolic analyses are therefore necessary to observe the metabolic networks important for plant growth and development under a range of environmental conditions.

Measurement of metabolism-wide fluxes through steady-state metabolic flux balance analysis (MFA or FBA) by measuring ^{13}C redistribution signatures within the primary metabolism at subcellular compartment level, and the information about the biomass composition and growth rate generate data, which is collectively described as Fluxome. Predicted flux maps is an important part of metabolic engineering (Becker et al. 2007). Recently, several methods are refined to predict metabolic networks that determine the fluxes, which directly report on cellular physiology. The most widely used approaches for fluxome analysis are based on GC–MS measurement of labelling pattern of metabolites from the tracer studies. This approach is optimized and applied to move from gaining information of static metabolic signatures to end products. A recent approach to the fluxome consists of the comprehensive determination of enzyme activities from cyclic robotic assays and determination of the activity of each reaction step in the metabolic pathway (Gibon et al. 2006; Osuna et al. 2007). The most direct information

of metabolic regulations can be obtained through the determination of an actual metabolic flux. This method also allows gaining precise knowledge of metabolic physiology and its engineering (Christensen and Nielsen 2000; Des Rosiers et al. 2004). A range of different MFA methods has been applied to plant systems, resulting in identifying unique insights into the operation of plant metabolic networks. Implementing the emerging MFA methods for plant studies faces considerable hurdles because of the greater complexity of plant metabolic networks and our ignorance of understanding the biochemical pathway and kinetics at sub-cellular compartment levels (Sweetlove et al. 2008). For metabolic flux calculation, the different labelling data obtained are usually utilized to globally fit the unknown flux parameters by a computer flux model combining isotopomer and metabolite balancing strategies (Wiechert et al. 2001; Kiefer et al. 2004; Wittmann et al. 2004; Frick and Wittmann 2005). It has been recognized that better optimization of experimental designs is essential for distinguishing activities between parallel metabolic pathways operative in distinct cellular compartments, such as cytosol and plastids (Allen et al. 2007; Kruger et al. 2007; Li et al. 2008). Overall, MFA and dynamic labeling methods are instrumental for quantifying metabolic fluxes of plant responses under ambient and challenging environments (Roscher et al. 2000; Boatright et al. 2004; Matsuda et al. 2005; Ratcliffe and Shachar-Hill 2006; Matsuda et al. 2009). Recently, genome wide metabolic fluxes have been predicted in *Arabidopsis* for high temperature and hyperosmotic stress, so that it was possible to identify key signatures such as severe reduction in carbon-use efficiency through reduction in PEP flux and increased TCA cycle for altered growth rate (Williams et al. 2010). Fewer studies have applied MFA in cereals. In maize, fast-growing excised root tips were used to study the central carbon metabolism by keeping them for 12–18 h in a medium containing ^{13}C-labeled glucose (Dieuaide-Noubhani et al. 1995; Edwards et al. 1998), and then analyzing the most abundant labeled free intracellular metabolites (i.e., sugars and amino acids) by NMR or MS; large flux maps of central carbon metabolism were derived in this manner (Dieuaide-Noubhani et al. 1995; Alonso et al. 2005). In other studies ^{13}C labeled glucose was used to label maize kernels and barley caryopsis, and label was analyzed in both glucose (derived from starch) and amino acids (derived from proteins) available in the starchy endosperm (Glawischnig et al. 2001, 2002; Grafahrend-Belau et al. 2009; Rolletschek et al. 2011).

8.3.4 Role of Hormones

Abiotic stress response involves a trigger of similar set of transcription factors involved in both ABA-dependent and ABA-independent manner in both dicotyledonous and monocotyledonous plants (reviewed by Sreenivasulu et al. 2007). Genes differentially regulated in *Arabidopsis* and rice in response to drought, salinity and cold stress comprise gene-sets enriched with DRE-related and ABRE core motifs. Therefore both ABA-dependent and ABA-independent

signaling pathways are important in regulating the transcriptome responses (Seki et al. 2002a, 2001; Gomez-Porras et al. 2007). Abscisic acid (ABA) remains the best-studied hormone for plant stress response. However, other hormones such as cytokinins, auxins, gibberellins, brassinosteroids, strigolactones, jasmonic acid, salicylic acid as well as the gaseous hormones, ethylene and nitric oxide are being studied for their role in abiotic stress response in the recent past. Hence, we need to understand the manipulation of the phytohormone synthesis and action across the plant life-cycle, which is an attractive avenue to understand and engineer abiotic stress tolerance. In barley, the response to salinity stress includes the synthesis and the induction of the jasmonate signalling transduction pathways (Walia et al. 2006, 2007). Recently, modification of cytokinin expression, with the critical difference in the use of a stress and maturation-induced promoter in rice resulted in elevating drought tolerance to produce higher yield under stress (Peleg et al. 2011). The observed differences in the content of other phytohormones in the cytokinin-modulated transgenic rice lines also suggested synergistic or antagonistic interactions between auxins, ethylene, cytokinins and ABA in regulating stomatal behavior. Furthermore, gibberellins and brassinosteroids have a strategic importance in tolerance to a variety of abiotic stresses (Peleg and Blumwald 2011). Critical alteration in the ratio of cytokinins and abscisic acid and its antagonistic responses is known to alter the growth dynamics under abiotic stress response (Nishiyama et al. (2011). Also, the effects of three different phytohormones auxin, ABA and cytokinins on the single trait of nitrogen acquisition were reported in a recent review (Kiba et al. 2011). Nitrogen acquisition and remobilization is an important trait to be considered in abiotic stress tolerance to fine tune source-sink relationships in enhancing grain yield (Seiler et al. 2011; Kohli et al. 2012).

8.3.5 Phenomics

Phenomics involves comprehensive capture of a plant's phenotype that helps to explore the germplasm. Unfortunately, there is a large gap in our understanding of events that may occur when genotype is translated into phenotype; there is an urgent need to fill this gap (Zamboni and Sauer 2004; Furbank and Tester 2011). Plant genomes possess great plasticity in the genomes for producing various types of phenotypes. However, the genetic variability that may prove useful for developing stress tolerant lines is limited.

There are large number of initiatives launched (IPPN: International Plant Phenomics Network; DPPN: Deutsches Pflanzen Phänotypisierungs Netzwerk; EPPN: European Plant Phenotyping Network; APPF: Australian Plant Phenomics Facility) to create phenotyping facilities to screen populations, GMO material and mutant collections by employing high end image capture technologies in the phytotrons and glass houses. The Plant Accelerator (Lemnatech scan analyzer 3D) which takes non-destructive measurements of plant biomass (Finkel 2009) can also be

used. The core of the Plant Accelerator's phenotyping facility also measures level of watering and nutrient supplementation control, managing plant movement and tracking, and records images of plants in a range of different wavelengths, thus providing enormous information about the diversity of phenotype. Visible cameras quantify overall plant morphology, size, colour, shoot mass and other physical characteristics; near infra-red cameras detect water content of the leaves and soil; far infra-red provides information about leaf temperature and transpiration rate. While UV detects chlorophyll fluorescence, the GFP fluorescence will be helpful to monitor transgene expression. The first phenomics study was the use of quantitative phenotypic assays to measure salt tolerance traits such as osmotolerance Na^+ exclusion and Na^+ tissue tolerance in the diploid wheat *T. monococcum* (Rajendran et al. 2009). The advantage of this approach is that it is non-invasive, allowing other omics approaches to analyze cell products from the same plant. Also other non-invasive techniques such as magnetic resonance imaging, high resolution based nuclear magnetic resonance and positron emission tomography are implemented to gain insights into structure–function relationship (reviewed by Mir et al. 2012). To fully explore the genotype dependent tolerance mechanisms within the breeding programs, field-based high-throughput phenotyping platforms are essential to monitor the canopy temperature using infrared thermography. Furthermore, implementations of remote sensing technologies are essential to fully explore phenotypic plasticity at the field level. To explore the key agronomic traits for the improvement of sustainable agriculture, one needs to expand the systematic phenotyping to explore allelic variation in mapping populations, breeding programs and large scale mutants and GMO collections.

8.4 Integrative Systems Biology

Integration of the different omic approaches in the area of abiotic stress tolerance allows more robust identifications of molecular targets for future biotechnological applications in crop plants. Manipulating plant metabolism to better serve the future needs requires an improved understanding of the links between genotype and phenotype. Therefore, the massive omics data created from multifaceted platforms of genome, transcriptome, proteome, metabolome, flux and enzyme kinetics (Table 8.2 and Fig. 8.1) need to be interlinked to the cellular phenotype to understand the cellular physiological status under perturbed environmental conditions (Sauer et al. 1999). To address the missing links between molecules and physiology, different approaches of systems biology are implemented which includes "top–down" and "bottom–up" strategies. The major strengths of top-down systems biology are to gain an integrative view of the huge collection of omic data sets like transcriptomics and/or proteomics, metabolomics and fluxomics (Westerhoff and Palsson 2004). Top–down systems biology identifies molecular interaction networks on the basis of correlated molecular behavior observed through genome-wide 'omics' studies. Also, bottom-up systems biology deduces the functional properties that could emerge from a subsystem that has been characterized to a

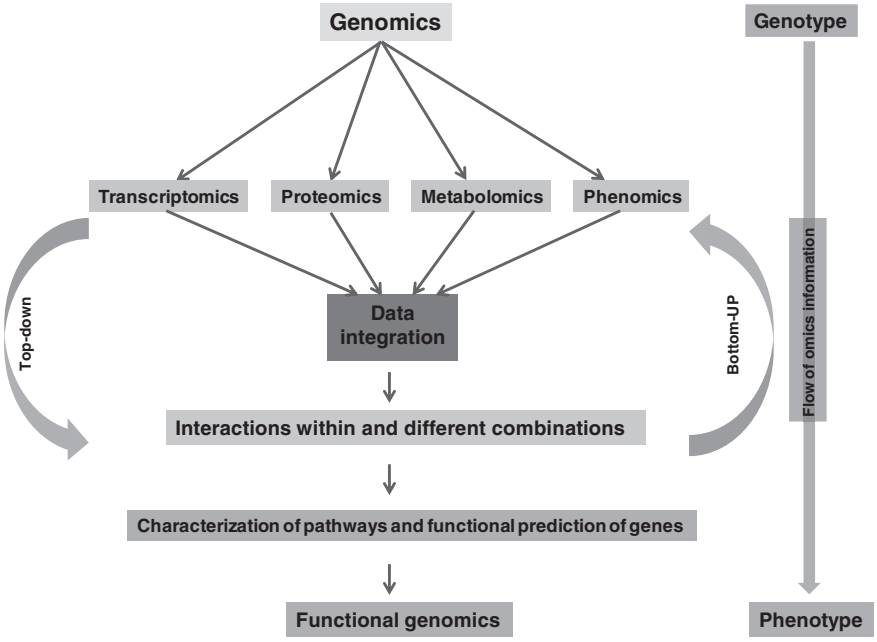

Fig. 8.1 Schematic representation of omics integration pipeline used in systems biology srategies

high level of mechanistic detail using molecular methods but focuses at the cellular level. By employing the systems biology tools in plant science, many abiotic stress-inducible genes were identified and their functions were precisely characterized in the model species.

Also, top–down systems biology concerns the identification of the structure of the molecular network that underlies system behavior that is, 'reverse engineering' from system data alone. Top–down approach starts by (re)constructing a possible topology of the network at a low level of complexity and provides a broad overview of the system at low resolution. Transcriptional networks through reverse engineering methods from the collections of gene expression data have been well pioneered on single-cell organisms, but have increasingly been applied to higher order organisms including plants where applications of systems biology methods are now emerging (Carro et al. 2010; Carrera et al. 2009; Needham et al. 2009). The available network models are mainly based on Boolean, Relevance, or Bayesian networks or association rules (Hache et al. 2009). These network inference methods are categorized into (1) those that aim to influence the genes in the general manner to influence the expression of other genes by forming gene regulatory networks (Bansal et al. 2007; Marbach et al. 2010) and (2) those, which aims to having physical interaction between transcription factors and the regulatory genes/motifs by forming the gene regulatory networks (Styczynski and Stephanopoulos 2005). The metabolic

network reconstructions that are normally done at the genome-scale are the key factors to characterize the genotype to phenotype relationships using all sequence and functional annotation data that is available in public databases combined with manual curation using the available literature and experimental data (Feist et al. 2009). Most systems biology studies have been implemented in the model plant *Arabidopsis*, where large transcriptomics programs have generated adequate quantities of high-quality data to enable systems analysis (Krishnan et al. 2009). The resulting knowledge can be used in cereals through comparative gene networks. Therefore, it is important to perform parallel studies in cereals with other characteristics, as well as to develop methods to allow use of data from the *Arabidopsis* system to conduct studies in other plant species.

Integration of different multiple 'omics' data is required to reconstruct complex networks that characterize the phenotypes in the cell (Moles et al. 2003; Kremling et al. 2004). In particular, transcriptome co-expression analysis for delimiting genes of interest has been implemented more efficiently using publicly available large transcriptome datasets such as AtGenExpress (Schmid et al. 2005; Goda et al. 2008) and NASCArrays (Craigon et al. 2004), which contain data from >1000 microarrays from model species alone. This kind of in-depth data is yet to be generated among cereal species to elucidate gene regulatory networks. The current status of available resources related to integrated databases for cereal crop species are listed in the Table 2. Transcriptome data sets are now available for co-expression analysis of the transcriptome in cereal crops; for instance, RiceArrayNet and OryzaExpress databases provide web-accessible co-expression data for rice (Lee et al. 2009; Hamada et al. 2011). The ATTED-II database also provides co-expression data sets for rice in addition to those for *Arabidopsis* (Obayashi et al. 2007; 2011). A co-expressed barley gene network was recently generated and then applied to comparative analysis to discover potential Triticeae- -specific gene expression networks (Mochida et al. 2011). PlaNet (http://aranet.mpimp-golm.mpg.de/), a database of co-expression networks for *Arabidopsis* and six plant crop species, uses a comparative network algorithm, NetworkComparer, to estimate similarities between network structures (Mutwil et al. 2011). This platform integrates gene expression patterns, associated functional annotations and MapMan term-based ontology, and facilitates knowledge transfer from *Arabidopsis* to crop species for the discovery of conserved co-expressed gene networks. The KEGG PLANT Resource (KEGG; http://www.genome.jp/kegg/) is one of the most widely established integrated database which provide information on primary metabolism of biosynthetic pathways. It aims to integrate genomic information resources with the biosynthetic pathways of natural plant products (Masoudi-Nejad et al. 2007). Another information resource for biosynthetic pathways, PlantCyc platform has been used for a number of plant species to analyze the computational analysis of the genes, enzymes, compounds, reactions and pathways involved in developmental and stress response. The pathways section in the Gramene databases provides RiceCyc, MaizeCyc, BrachyCyc and SorghumCyc, for rice, maize, Brachypodium and sorghum, respectively (http://www.gramene.org/pathway/). These resources will enable cereal workers to focus on active analysis of regulatory networks that

may be involved in different biological functions (de la Fuente et al. 2002; Vlad et al. 2004; Kholodenko et al. 2002).

Generally, a preselected set of genes designated as guide genes or bait genes for the core part of the network modules is computed for co-expression with other genes for the generation of co-expression networks (Horan et al. 2008). If a network frame is formed between unknown and known genes, it is presumed that these genes share a common regulatory system and thus are involved in the same pathway. This approach was applied for identification of genes involved in several biochemical pathways such as cellulose synthesis (Persson et al. 2005), aliphatic glucosinolate biosynthesis (Hirai et al. 2007), glucosinolate biosynthetic pathway (Hansen et al. 2007; Geu-Flores et al. 2009) and hormone metabolism (Goda et al. 2008). In case of cereals, the integrated analysis of metabolome and transcriptome was recently conducted to analyze rice caryopses developing under high temperature conditions (Yamakawa and Hakata 2010); molecular events underlying pollination-induced and pollination-independent fruit sets were also examined (Wang et al. 2009). Integrated analysis of metabolome and transcriptome has also been applied to investigate changing metabolic systems in field grown plants of rice Os-GIGANTEA (Os-GI) mutant and transgenic barley lines (Kogel et al. 2010; Izawa et al. 2011). An integrated analysis of proteome and metabolome was also used to compare the differences in response to anoxia between rice and wheat coleoptiles (Shingaki-Wells et al. 2011). Furthermore, an integrated analysis of transcriptome, proteome and metabolome was conducted to characterize the cascading changes in UV-B-mediated responses in maize (Casati et al. 2011). In this context, the integrated approaches with multiple omics data should contribute greatly to the identification of key regulatory steps and to characterize the pathways for various processes. Following these successful efforts, multi-omics-based systems analyses have improved our understanding of plant cellular systems by integrating metabolome analysis with genome and transcriptome resources (Hirai et al. 2004; Saito et al. 2008; Okazaki et al. 2009). The URLs of each integrative database in plant genomics are listed in Table 2.

The main objective of the above strategy is to discover new molecular mechanisms using an iterative cycle that starts with experimental data, followed by data analysis and data integration to determine correlations between the molecules. As an end process, the formulation of hypotheses concerning co- and inter-regulation of groups of those molecules will be revealed. The omics data obtained under a specific condition such as stress response from a given gene knockouts are used for integrated omics analysis in this strategy. Such an analysis allowed the prediction of functional relevance of key genes involved in stress-specific regulons determining tolerance. This approach has become the key to decipher the functional analysis of the genes identified from the whole genome sequencing of the plants. Alternatively, it also helps to identify ubiquitous stress regulated pathways. However, more attention is now focused in the creation of mutants and screening the response to abiotic stress using multi-layered omics strategy. To date, more than half a million T-DNA mutants have been developed for rice and *Arabidopsis* (An et al. 2005; O'Malley and Ecker 2010). In other cereals, like Brachypodium, a large-scale collection of T-DNA tagged lines termed 'the BrachyTAG program' have been developed and

used to investigate gene functions (Thole et al. 2010). A collection of several knockout mutants in cereals has been generated to assess the function of genes involved in abiotic stress. The rice full-length cDNA overexpressed *Arabidopsis* mutant database (Rice FOX Database, http://ricefox.psc.riken.jp/) was a new information resource for the FOX line (Sakurai et al. 2011). The system was also used to screen salt stress-resistant lines in the T_1 generation produced by the transformation of 43 focused stress-inducible transcription factors of *Arabidopsis* (Fujita et al. 2007). Then, the system was applied to a set of full-length rice cDNA clones aiming for *in planta* high-throughput screening of rice functional genes, with *Arabidopsis* as the host species (Kondou et al. 2009). Thus, the FOX hunting system is capable of the high-throughput characterization of gene functions. Furthermore, in rice, the endogenous retrotransposon *Tos*17, which is activated in particular conditions, is also available for the study of the insertion mutant lines of a *japonica* rice cultivar, Nipponbare (Miyao et al. 2007). Several mutants were isolated in wheat, which showed increased resistance towards biotic stress tolerance. In wheat, heat tolerant (Mullarkey and Jones 2000) and salt tolerant plants (Huo et al. 2004) have already been charecterized to study the genetic basis of stress tolerance. Additionally, the maize *Enhancer*/*Suppressor Mutator* (En/*Spm*) element has also been used as an effective tool for the study of functional genomics in plants (Kumar et al. 2005). Other approach to study the gain-of-function of mutations by activation tagging have been developed and performed in *Arabidopsis*, rice and soybean (Weigel et al. 2000; An et al. 2005; Kuromori et al. 2009). The current status of available resources related to mutants database for cereal crop species were listed in the Table 8.2.

8.5 Identification of Key Candidate Genes for Tolerance to Abiotic Stress and Validation of their Functions Using Transgenic Approaches

One of the key challenges facing agriculture today is the acute water shortage and high temperature caused by worldwide climate change and the increasing world population. Fulfilling the needs of this growing population is quite difficult from the limited arable land area available on the globe. Although there are legal, social and political barriers to the utilization of biotechnology, advances made in this field have great potential to substantially improve agricultural productivity under challenging environments. Both non-GMO and GMO strategies have been implemented to improve tolerance in crop plants. Genetic engineering is thus being intensively explored to improve plant tolerance to various abiotic stresses, and transgenic crop genotypes with improved stress resistance have actually been produced (Bartels and Sunkar 2005; Vinocur and Altman 2005; Umezawa et al. 2006; Pennisi 2008; Wan et al. 2009). In case of maize, drought tolerance transgenics are also undergoing field trails in Africa, and some other drought tolerant genotypes are also being used by the farmers for commercial cultivation. Performance of a number of other events in maize and other crops are being subjected to field trials. Partial drought

tolerance has been achieved in the vegetative phase through gene transfer by altering the accumulation of osmoprotectants, production of chaperones, protection of cell integrity by expression of LEA proteins and improved superoxide radical scavenging mechanisms (see reviews by Hasegawa et al. 2000; Kishor et al. 2005; Sangam et al. 2005; Sreenivasulu et al. 2004b, 2007; Vij and Tyagi 2007). In addition, overexpression of the key regulators ABF2, ABF3 and ABF4 of Arabidopsis involved in ABA-dependent signaling as well as constitutive expression of the Arabidopsis DREB1A, DREB1B, DREB1C and DREB2A transcription factors participating in ABA-independent signaling pathways have been shown to be effective in engineering drought tolerance (see reviews by Agarwal et al. 2006; Umezawa et al. 2006; Sreenivasulu et al. 2007). Genetic engineering strategy has been successfully applied to increase tolerance against a number of other abiotic stresses also. In this context, a variety of crops from cereals (rice, maize, barley, Brachypodium and wheat etc.,) have been engineered for enhanced resistance to a multitude of stresses, each individually, or in combination of biotic and abiotic stresses. Enhancing plant tolerance to abiotic stresses involves multiple mechanisms and therefore involves manipulation of different physiological and biochemical pathways (Wang et al. 2003; Zhang et al. 2009).

8.6 Summary and Outlook

The availability of complete genome sequence information of model species like *Arabidopsis thaliana*, *Oryza sativa* and other cereal plants has made valuable contributions in dissecting the stress response at the level of transcriptional regulation, post-transcriptional, post-translational modifications and epigenetic regulation. Using high throughput modern techniques like transcriptomics, metabolomics and proteomics, stress-responsive pathway genes have been identified. These strategies enabled us to identify key stress regulators by deriving complicated regulatory network. Employing the systems biology tools in plant science, many abiotic stress-inducible genes were identified and their functions were precisely characterized in the model species.

The identification of stress-regulators gave rise to the idea that plants have developed flexible cellular response mechanisms to efficiently respond to various abiotic stresses. Numerous genes that are induced by various abiotic stresses have been identified using various microarray systems and these gene products are classified into two groups. The first group includes proteins functioning in direct abiotic stress tolerance; these include the following: chaperones, LEA proteins, osmotin, antifreeze proteins, mRNA-binding proteins, key enzymes for osmolyte biosynthesis such as proline, water channel proteins, sugar and proline transporters, detoxification enzymes, and enzymes involved in fatty acid metabolism, proteinase inhibitors, ferritin, and lipid-transfer proteins. The second group includes factors involved in regulatory function related to signal transduction, hormonal response and transcription factors, which are responsive to various stress factors. These transcription factors could regulate various stress inducible genes cooperatively or independently, and may constitute gene networks.

Under drought, photosynthesis is affected by decreased intake and diffusion of CO_2 due to modulation of stomatal opening by phytohormones. In response to altered carbon intake, the changed leaf sugar status acts as a metabolic signal. In concert with other phytohormones, it inhibits growth, which further alters the carbon: nitrogen ratio. The stress conditions generated by severe drought and nutrient deprivation triggers energy imbalances, as well further loop-in alteration between growth promoting and growth retarding phytohormones (Sreenivasulu et al. 2012), generation of reactive oxygen species (ROS) and second messengers such as calcium to affect transcriptional regulation of numerous genes. Their meta-analysis indicated that variables on the time and severity of stress and plant species made it difficult to find a general trend in relating molecular responses to the physiological status of the plant. Functional characterization of stress inducible transcription factors should provide more information in the complex regulatory gene networks that are involved in responses to drought, high temperature, and high salinity stresses. At present, the functions of many of these genes are not fully characterized. Some attempts at analyzing large scale high throughput data allows us to bring the different elements together, suggesting that the integration of stress cues into development and plant growth in dealing with crop yield under stress is rather complicated. Such diversity in needs, approaches, opinions and indeed results has led to generation of massive literature, which needs to be systematically reviewed to derive proper strategies for understanding the stress tolerance mechanisms. Therefore methods implied in systems biology approaches remain pivotal to systematically reveal the function of these stress-responsive pathways.

Acknowledgments NS is thankful to research funding obtained through BMBF (IND 09/526), BLE grant 511-06.01-28-1-45.041-10, BMZ grant 81131833 and from the Ministry of Education Saxony-Anhalt (IZN). PS acknowledge the Leibniz-DAAD post doctoral fellowship award (Number: A/11/94309) from Germany Academic Exchange programme (DAAD), Germany. We acknowledge the help of Prof. P.K. Gupta for the editorial changes which helped to improve the manuscript.

References

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

Agrawal GK, Rakwal R (2006) Rice proteomics: a cornerstone for cereal food crop proteomes. Mass Spectrom Rev 25:1–53

Agrawal GK, Rakwal R (2011) Rice proteomics: a move toward expanded proteome coverage to comparative and functional proteomics uncovers the mysteries of rice and plant biology. Proteomics 11:1630–1649

Alcazar R, Garcia-Martinez JL, Cuevas JC, Tiburcio AF, Altabella T (2005) Overexpression of ADC2 in Arabidopsis induces dwarfism and late-flowering through GA deficiency. Plant J 43:425–436

Ali GM, Komatsu S (2006) Proteomic analysis of rice leaf sheath during drought stress. J Proteome Res 5:396–403

Allen DK, Shachar-Hill Y, Ohlrogge JB (2007) Compartment-specific labeling information in 13C metabolic flux analysis of plants. Phytochemistry 68:2197–2210

Alonso AP, Vigeolas H, Raymond P, Rolin D, Dieuaide-Noubhani M (2005) A new substrate cycle in plants. Evidence for a high glucose-phosphate-to-glucose turnover from in vivo steady-state and pulse-labeling experiments with [13C] glucose and [14C] glucose. Plant Physiol 138:2220–2232

Alves SC, Worland B, Thole V, Snape JW, Bevan MW, Vain P (2009) A protocol for agrobacterium-mediated transformation of brachypodium distachyon community standard line Bd21. Nat Protoc 4:638–649

Amudha J, Balasubramani G (2011) Recent molecular advances to cambat abiotic stress tolerance in crop plants. Biotech Mol Bio Rev 6:31–58

An GH, Lee S, Kim SH, Kim SR (2005) Molecular genetics using T-DNA in rice. Plant Cell Physiol 46:14–22

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

Bansal M, Belcastro V, Ambesi-Impiombato A, di Bernardo D (2007) How to infer gene networks from expression profiles. Mol Syst Biol 3(1):78

Bartels D, Sunkar R (2005) Drought and salt tolerance in plants. Crit Rev Plant Sci 24:23–58

Battisti DS, Naylor RL (2009) Historical warnings of future food insecurity with unprecedented seasonal heat. Science 323:240–244

Becker J, Klopprogge C, Herold A, Zelder O, Bolten CJ, Wittmann C (2007) Metabolic flux engineering of L-lysine production in corynebacterium glutamicum: over expression and modification of G6P dehydrogenase. J Biotechnol 132:99–109

Berhan AM, Hulbert SH, Butler LG, Bennetzen JL (1993) Structure and evolution of the genomes of sorghum-bicolor and zea-mays. Theor Appl Genet 86:598–604

Berkman PJ, Skarshewski A, Lorenc MT, Lai K, Duran C, Ling EY, Stiller J, Smits L, Imelfort M, Manoli S, McKenzie M, Kubalakova M, Simkova H, Batley J, Fleury D, Dolezel J, Edwards D (2011) Sequencing and assembly of low copy and genic regions of isolated triticum aestivum chromosome arm 7DS. Plant Biotechnol J 9:768–775

Bino RJ, Hall RD, Fiehn O, Kopka J, Saito K, Draper J, Nikolau BJ, Mendes P, Roessner-Tunali U, Beale MH, Trethewey RN, Lange BM, Wurtele ES, Sumner LW (2004) Potential of metabolomics as a functional genomics tool. Trends Plant Sci 9:418–425

Boatright J, Negre F, Chen XL, Kish CM, Wood B, Peel G, Orlova I, Gang D, Rhodes D, Dudareva N (2004) Understanding in vivo benzenoid metabolism in petunia petal tissue. Plant Physiol 135:1993–2011

Bombarely A, Menda N, Tecle IY, Buels RM, Strickler S, Fischer-York T, Pujar A, Leto J, Gosselin J, Mueller LA (2011) The sol genomics network (solgenomics.net): growing tomatoes using perl. Nucleic Acids Res 39:D1149–D1155

Caldana C, Degenkolbe T, Cuadros-Inostroza A, Klie S, Sulpice R, Leisse A, Steinhauser D, Fernie AR, Willmitzer L, Hannah MA (2011) High-density kinetic analysis of the metabolomic and transcriptomic response of Arabidopsis to eight environmental conditions. Plant J 67:869–884

Canovas FM, Dumas-Gaudot E, Recorbet G, Jorrin J, Mock HP, Rossignol M (2004) Plant proteome analysis. Proteomics 4:285–298

Carrera J, Rodrigo G, Jaramillo A, Elena SF (2009) Reverse-engineering the Arabidopsis thaliana transcriptional network under changing environmental conditions. Genome biol 10(9):R96

Carro MS, Lim WK, Alvarez MJ, Bollo RJ, Zhao XD, Snyder EY, Sulman EP, Anne SL, Doetsch F, Colman H, Lasorella A, Aldape K, Califano A, Iavarone A (2010) The transcriptional network for mesenchymal transformation of brain tumours. Nature 463:318–368

Caruso G, Cavaliere C, Foglia P, Gubbiotti R, Samperi R, Lagana A (2009) Analysis of drought responsive proteins in wheat (Triticum durum) by 2D-PAGE and MALDI-TOF mass spectrometry. Plant Sci 177:570–576

Casati P, Morrow DJ, Fernandes JF, Walbot V (2011) Rapid Maize leaf and immature ear responses to UV-B radiation. Front Plant Sci 2:33

Chantret N, Salse J, Sabot F, Rahman S, Bellec A, Laubin B, Dubois I, Dossat C, Sourdille P, Joudrier P, Gautier MF, Cattolico L, Beckert M, Aubourg S, Weissenbach J, Caboche M, Bernard M, Leroy P, Chalhoub B (2005) Molecular basis of evolutionary events that shaped the hardness locus in diploid and polyploid wheat species (Triticum and aegilops). Plant Cell 17:1033–1045

Charng YY, Liu HC, Liu NY, Hsu FC, Ko SS (2006) Arabidopsis Hsa32, a novel heat shock protein, is essential for acquired thermotolerance during long recovery after acclimation. Plant Physiol 140:1297–1305

Christensen B, Nielsen J (2000) Metabolic network analysis. A powerful tool in metabolic engineering. Adv Biochem Eng Biotechnol 66:209–231

Collins NC, Tardieu F, Tuberosa R (2008) Quantitative trait loci and crop performance under abiotic stress: where do we stand? Plant Physiol 147:469–486

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 101:15243–15248

Craigon DJ, James N, Okyere J, Higgins J, Jotham J, May S (2004) NASCArrays: a repository for microarray data generated by NASC's transcriptomics service. Nucleic Acids Res 32:D575–D577

Cramer GR, Ergul A, Grimplet J, Tillett RL, Tattersall EAR, Bohlman MC, Vincent D, Sonderegger J, Evans J, Osborne C, Quilici D, Schlauch KA, Schooley DA, Cushman JC (2007) Water and salinity stress in grapevines: early and late changes in transcript and metabolite profiles. Funct Integr Genomics 7:111–134

Cramer GR, Urano K, Delrot S, Pezzotti M, Shinozaki K (2011) Effects of abiotic stress on plants: a systems biology perspective. BMC Plant Biol 11(1):63

de la Fuente A, Brazhnik P, Mendes P (2002) Linking the genes: inferring quantitative gene networks from microarray data. Trends Genet: TIG 18:395–398

de Vienne D, Leonardi A, Damerval C, Zivy M (1999) Genetics of proteome variation for QTL characterization: application to drought-stress responses in maize. J Exp Bot 50:303–309

Deeb F, van der Weele CM, Wolniak SM (2010) Spermidine is a morphogenetic determinant for cell fate specification in the male gametophyte of the water fern marsilea vestita. Plant Cell 22:3678–3691

Des Rosiers C, Lloyd S, Comte B, Chatham JC (2004) A critical perspective of the use of C-13-isotopomer analysis by GCMS and NMR as applied to cardiac metabolism. Metab Eng 6:44–58

Deyholos MK (2010) Making the most of drought and salinity transcriptomics. Plant Cell Environ 33:648–654

Dieuaide-Noubhani M, Raffard G, Canioni P, Pradet A, Raymond P (1995) Quantification of compartmented metabolic fluxes in maize root tips using isotope distribution from 13C- or 14C-labeled glucose. J Biol Chem 270:13147–13159

Dixon RA, Strack D (2003) Phytochemistry meets genome analysis, and beyond. Phytochemistry 62:815–816

Dolezel J, Kubalakova M, Paux E, Bartos J, Feuillet C (2007) Chromosome-based genomics in the cereals. Chromosome Res: Int J Mol, Supramol Evol Aspects Chromosome Biol 15:51–66

Dooki AD, Mayer-Posner FJ, Askari H, Zaiee AA, Salekdeh GH (2006) Proteomic responses of rice young panicles to salinity. Proteomics 6:6498–6507

Druka A, Franckowiak J, Lundqvist U, Bonar N, Alexander J, Houston K, Radovic S, Shahinnia F, Vendramin V, Morgante M, Stein N, Waugh R (2011) Genetic dissection of barley morphology and development. Plant Physiol 155:617–627

Druka A, Muehlbauer G, Druka I, Caldo R, Baumann U, Rostoks N, Schreiber A, Wise R, Close T, Kleinhofs A, Graner A, Schulman A, Langridge P, Sato K, Hayes P, McNicol J, Marshall D, Waugh R (2006) An atlas of gene expression from seed to seed through barley development. Funct Integr Genomics 6:202–211

Eastmond PJ, van Dijken AJH, Spielman M, Kerr A, Tissier AF, Dickinson HG, Jones JDG, Smeekens SC, Graham IA (2002) Trehalose-6-phosphate synthase 1, which catalyses the first step in trehalose synthesis, is essential for Arabidopsis embryo maturation. Plant J 29:225–235

Edwards D, Batley J (2010) Plant genome sequencing: applications for crop improvement. Plant Biotechnol J 8:2–9

Edwards S, Nguyen BT, Do B, Roberts JKM (1998) Contribution of malic enzyme, pyruvate kinase, phosphoenolpyruvate carboxylase, and the krebs cycle to respiration and biosynthesis and to intracellular pH regulation during hypoxia in maize root tips observed by nuclear magnetic resonance imaging and gas chromatography-mass spectrometry. Plant Physiol 116:1073–1081

Ergen NZ, Budak H (2009) Sequencing over 13,000 expressed sequence tags from six subtractive cDNA libraries of wild and modern wheats following slow drought stress. Plant Cell Environ 32:220–236

Ergen NZ, Thimmapuram J, Bohnert HJ, Budak H (2009) Transcriptome pathways unique to dehydration tolerant relatives of modern wheat. Funct Integr Genomics 9:377–396

Espinoza C, Degenkolbe T, Caldana C, Zuther E, Leisse A, Willmitzer L, Hincha DK, Hannah MA (2010) Interaction with diurnal and circadian regulation results in dynamic metabolic and transcriptional changes during cold acclimation in Arabidopsis. PloS one 5(11):e14101

Feist AM, Herrgard MJ, Thiele I, Reed JL, Palsson BO (2009) Reconstruction of biochemical networks in microorganisms. Nat Rev Microbiol 7:129–143

Feuillet C, Keller B (2002) Comparative genomics in the grass family: molecular characterization of grass genome structure and evolution. Ann Bot 89:3–10

Fiehn O (2002) Metabolomics: the link between genotypes and phenotypes. Plant Mol Biol 48:155–171

Filiz E, Ozdemir BS, Budak F, Vogel JP, Tuna M, Budak H (2009) Molecular, morphological, and cytological analysis of diverse Brachypodium distachyon inbred lines. Genome 52:876–890 National Research Council Canada, Conseil National de Recherches Canada

Finkel E (2009) Imaging With 'phenomics', plant scientists hope to shift breeding into overdrive. Science 325:380–381

Finnie C, Melchior S, Roepstorff P, Svensson B (2002) Proteome analysis of grain filling and seed maturation in barley. Plant Physiol 129:1308–1319

Fleury D, Jefferies S, Kuchel H, Langridge P (2010) Genetic and genomic tools to improve drought tolerance in wheat. J Exp Bot 61:3211–3222

Frick O, Wittmann C (2005) Characterization of the metabolic shift between oxidative and fermentative growth in Saccharomyces cerevisiae by comparative C-13 flux analysis. Microb Cell Fact 4

Fujita M, Mizukado S, Fujita Y, Ichikawa T, Nakazawa M, Seki M, Matsui M, Yamaguchi-Shinozaki K, Shinozaki K (2007) Identification of stress-tolerance-related transcription-factor genes via mini-scale Full-length cDNA Over-eXpressor (FOX) gene hunting system. Biochem Bioph Res Co 364:250–257

Furbank RT, Tester M (2011) Phenomics: technologies to relieve the phenotyping bottleneck. Trends Plant Sci 16:635–644

Gagneul D, Ainouche A, Duhaze C, Lugan R, Larher FR, Bouchereau A (2007) A reassessment of the function of the so-called compatible solutes in the halophytic Plumbaginaceae Limonium latifolium. Plant Physiol 144:1598–1611

Garvin DF, McKenzie N, Vogel JP, Mockler TC, Blankenheim ZJ, Wright J, Cheema JJS, Dicks J, Huo NX, Hayden DM, Gu Y, Tobias C, Chang JH, Chu A, Trick M, Michael TP, Bevan MW, Snape JW (2010) An SSR-based genetic linkage map of the model grass Brachypodium distachyon. Genome 53:1–13 National Research Council Canada, Conseil National de Recherches Canada

Geu-Flores F, Nielsen MT, Nafisi M, Moldrup ME, Olsen CE, Motawia MS, Halkier BA (2009) Glucosinolate engineering identifies a gamma-glutamyl peptidase. Nat Chem Biol 5:575–577

Gibon Y, Usadel B, Blaesing OE, Kamlage B, Hoehne M, Trethewey R, Stitt M (2006) Integration of metabolite with transcript and enzyme activity profiling during diurnal cycles in Arabidopsis rosettes. Genome Biol 7:R76

Glawischnig E, Gierl A, Tomas A, Bacher A, Eisenreich W (2001) Retrobiosynthetic nuclear magnetic resonance analysis of amino acid biosynthesis and intermediary metabolism. Metabolic flux in developing maize kernels. Plant Physiol 125:1178–1186

Glawischnig E, Gierl A, Tomas A, Bacher A, Eisenreich W (2002) Starch biosynthesis and intermediary metabolism in maize kernels. Quantitative analysis of metabolite flux by nuclear magnetic resonance. Plant Physiol 130:1717–1727

Goda H, Sasaki E, Akiyama K, Maruyama-Nakashita A, Nakabayashi K, Li WQ, Ogawa M, Yamauchi Y, Preston J, Aoki K, Kiba T, Takatsuto S, Fujioka S, Asami T, Nakano T, Kato H, Mizuno T, Sakakibara H, Yamaguchi S, Nambara E, Kamiya Y, Takahashi H, Hirai MY, Sakurai T, Shinozaki K, Saito K, Yoshida S, Shimada Y (2008) The AtGenExpress hormone and chemical treatment data set: experimental design, data evaluation, model data analysis and data access. Plant J 55:526–542

Goff SA, Ricke D, Lan TH, Presting G, Wang RL, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, Hadley D, Hutchinson D, Martin C, Katagiri F, Lange BM, Moughamer T, Xia Y, Budworth P, Zhong JP, Miguel T, Paszkowski U, Zhang SP, Colbert M, Sun WL, Chen LL, Cooper B, Park S, Wood TC, Mao L, Quail P, Wing R, Dean R, Yu YS, Zharkikh A, Shen R, Sahasrabudhe S, Thomas A, Cannings R, Gutin A, Pruss D, Reid J, Tavtigian S, Mitchell J, Eldredge G, Scholl T, Miller RM, Bhatnagar S, Adey N, Rubano T, Tusneem N, Robinson R, Feldhaus J, Macalma T, Oliphant A, Briggs S (2002) A draft sequence of the rice genome (Oryza sativa L. ssp japonica). Science 296:92–100

Gomez-Porras JL, Riano-Pachon DM, Dreyer I, Mayer JE, Mueller-Roeber B (2007) Genome-wide analysis of ABA-responsive elements ABRE and CE3 reveals divergent patterns in Arabidopsis and rice. BMC Genomics 8(1):260

Gong QQ, Li PH, Ma SS, Rupassara SI, Bohnert HJ (2005) Salinity stress adaptation competence in the extremophile Thellungiella halophila in comparison with its relative Arabidopsis thaliana. Plant J 44:826–839

Grafahrend-Belau E, Schreiber F, Koschutzki D, Junker BH (2009) Flux balance analysis of barley seeds: a computational approach to study systemic properties of central metabolism. Plant Physiol 149:585–598

Gu YQ, Ma YQ, Huo NX, Vogel JP, You FM, Lazo GR, Nelson WM, Soderlund C, Dvorak J, Anderson OD, Luo MC (2009) A BAC-based physical map of Brachypodium distachyon and its comparative analysis with rice and wheat. BMC Genomics 10(1):496

Gupta AK, Kaur N (2005) Sugar signalling and gene expression in relation to carbohydrate metabolism under abiotic stresses in plants. J Biosciences 30:761–776

Gupta PK, Mir RR, Mohan A, Kumar J (2008) Wheat genomics: present status and future prospects. Int J Plant Genomics 2008:896451

Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat Biotechnol 17:994–999

Haberer G, Young S, Bharti AK, Gundlach H, Raymond C, Fuks G, Butler E, Wing RA, Rounsley S, Birren B, Nusbaum C, Mayer KF, Messing J (2005) Structure and architecture of the maize genome. Plant Physiol 139:1612–1624

Hache H, Lehrach H, Herwig R (2009) Reverse engineering of gene regulatory networks: a comparative study. EURASIP J Bioinf Syst Biol 2009:8

Hadiarto T, Tran LS (2011) Progress studies of drought-responsive genes in rice. Plant Cell Rep 30:297–310

Hagel JM, Facchini PJ (2008) Plant metabolomics: analytical platforms and integration with functional genomics. Phytochem Rev 7:479–497

Hajheidari M, Eivazi A, Buchanan BB, Wong JH, Majidi I, Salekdeh GH (2007) Proteomics uncovers a role for redox in drought tolerance in wheat. J Proteome Res 6:1451–1460

Hall RD (2006) Plant metabolomics: from holistic hope, to hype, to hot topic. New Phytol 169:453–468

Hamada K, Hongo K, Suwabe K, Shimizu A, Nagayama T, Abe R, Kikuchi S, Yamamoto N, Fujii T, Yokoyama K, Tsuchida H, Sano K, Mochizuki T, Oki N, Horiuchi Y, Fujita M, Watanabe M, Matsuoka M, Kurata N, Yano K (2011) OryzaExpress: an integrated database of gene expression networks and omics annotations in rice. Plant Cell Physiol 52:220–229

Hansen BG, Kliebenstein DJ, Halkier BA (2007) Identification of a flavin-monooxygenase as the S-oxygenating enzyme in aliphatic glucosinolate biosynthesis in Arabidopsis. Plant J 50:902–910

Hanzawa Y, Takahashi T, Michael AJ, Burtin D, Long D, Pineiro M, Coupland G, Komeda Y (2000) ACAULIS5 an Arabidopsis gene required for stem elongation, encodes a spermine synthase. EMBO J 19:4248–4256

Harrigan GG, Stork LG, Riordan SG, Ridley WP, Macisaac S, Halls SC, Orth R, Rau D, Smith RG, Wen L, Brown WE, Riley R, Sun D, Modiano S, Pester T, Lund A, Nelson D (2007) Metabolite analyses of grain from maize hybrids grown in the United States under drought and watered conditions during the 2002 field season. J Agric Food Chem 55:6169–6176

Hasegawa PM, Bressan RA, Zhu JK, Bohnert HJ (2000) Plant cellular and molecular responses to high salinity. Ann Rev Plant Physiol Plant Mol Biol 51:463–499

Hayes PM, Castro A, Marquez-Cedillo L, Corey A, Henson C, Jones B, Kling J, Mather D, Matus I, Rossi C, Sato K (2003) Genetic diversity for quantitatively inherited agronomic and malting quality traits. In: Von Bothmer R, Knupfeer H, van Hintum T, Sato K (eds) Diversity barley. Elsevier Science Publishers, Amsterdam

Helmy M, Tomita M, Ishihama Y (2011) OryzaPG-DB: rice proteome database based on shotgun proteogenomics. BMC Plant Biol 11:63

Hirai MY, Sugiyama K, Sawada Y, Tohge T, Obayashi T, Suzuki A, Araki R, Sakurai N, Suzuki H, Aoki K, Goda H, Nishizawa OI, Shibata D, Saito K (2007) Omics-based identification of Arabidopsis Myb transcription factors regulating aliphatic glucosinolate biosynthesis. Proc Natl Acad Sci USA 104:6478–6483

Hirai MY, Yano M, Goodenowe DB, Kanaya S, Kimura T, Awazuhara M, Arita M, Fujiwara T, Saito K (2004) Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in Arabidopsis thaliana. Proc Natl Acad Sci USA 101:10205–10210

Horan K, Jang C, Bailey-Serres J, Mittler R, Shelton C, Harper JF, Zhu JK, Cushman JC, Gollery M, Girke T (2008) Annotating genes of known and unknown function by large-scale coexpression analysis. Plant Physiol 147:41–57

Hu WH, Hu GC, Han B (2009) Genome-wide survey and expression profiling of heat shock proteins and heat shock factors revealed overlapped and stress specific response under abiotic stresses in rice. Plant Sci 176:583–590

Huo CM, Zhao BC, Ge RC, Shen YZ, Huang ZJ (2004) Proteomic analysis of the salt tolerance mutant of wheat under salt stress. Acta Genetica Sinica 31:1408–1414 Yi chuan xue bao

Huo NX, Gu YQ, Lazo GR, Vogel JP, Coleman-Derr D, Luo MC, Thilmony R, Garvin DF, Anderson OD (2006) Construction and characterization of two BAC libraries from Brachypodium distachyon, a new model for grass genomics. Genome 49(9):1099–1108 National Research Council Canada, Conseil National de Recherches Canada

Huo NX, Lazo GR, Vogel JP, You FM, Ma YQ, Hayde DM, Coleman-Derr D, Hill TA, Dvorak J, Anderson OD, Luo MC, Gu YQ (2008) The nuclear genome of Brachypodium distachyon: analysis of BAC end sequences. Funct Integr Genomics 8:135–147

Imai A, Matsuyama T, Hanzawa Y, Akiyama T, Tamaoki M, Saji H, Shirano Y, Kato T, Hayashi H, Shibata D, Tabata S, Komeda Y, Takahashi T (2004) Spermidine synthase genes are essential for survival of Arabidopsis. Plant Physiol 135:1565–1573

Ingram J, Bartels D (1996) The molecular basis of dehydration tolerance in plants. Ann Rev Plant Physiol Plant Mol Biol 47:377–403

Intergovernmental Panel on Climate Change (IPCC) (2007) In: Pachauri RK, Reisinger A (eds) Climate change 2007 synthesis report. IPCC, Geneva

Izawa T, Mihara M, Suzuki Y, Gupta M, Itoh H, Nagano AJ, Motoyama R, Sawada Y, Yano M, Hirai MY, Makino A, Nagamura Y (2011) Os-GIGANTEA confers robust diurnal rhythms on the global transcriptome of rice in the field. Plant Cell 23:1741–1755

Janz D, Behnke K, Schnitzler JP, Kanawati B, Schmitt-Kopplin P, Polle A (2010) Pathway analysis of the transcriptome and metabolome of salt sensitive and tolerant poplar species reveals evolutionary adaption of stress tolerance mechanisms. BMC Plant Biol 10:150

Jeanneau M, Gerentes D, Foueillassar X, Zivy M, Vidal J, Toppan A, Perez P (2002) Improvement of drought tolerance in maize: towards the functional validation of the Zm-Asr1 gene and increase of water use efficiency by over-expressing C4-PEPC. Biochimie 84:1127–1135

Jordan KW, Nordenstam J, Lauwers GY, Rothenberger DA, Alavi K, Garwood M, Cheng LL (2009) Metabolomic characterization of human rectal adenocarcinoma with intact tissue magnetic resonance spectroscopy. Dis Colon Rectum 52:520–525

Kaplan F, Kopka J, Haskell DW, Zhao W, Schiller KC, Gatzke N, Sung DY, Guy CL (2004) Exploring the temperature-stress metabolome of Arabidopsis. Plant Physiol 136:4159–4168

Kaplan F, Kopka J, Sung DY, Zhao W, Popp M, Porat R, Guy CL (2007) Transcript and metabolite profiling during cold acclimation of Arabidopsis reveals an intricate relationship of cold-regulated gene expression with modifications in metabolite content. Plant J 50:967–981

Kell DB, Brown M, Davey HM, Dunn WB, Spasic I, Oliver SG (2005) Metabolic footprinting and systems biology: the medium is the message. Nat Rev Microbiol 3:557–565

Kempa S, Krasensky J, Dal Santo S, Kopka J, Jonak C (2008) A central role of abscisic acid in stress-regulated carbohydrate metabolism. PLoS ONE 3:e3935

Kempa S, Rozhon W, Samaj J, Erban A, Baluska F, Becker T, Haselmayer J, Schleiff E, Kopka J, Hirt H, Jonak C (2007) A plastid-localized glycogen synthase kinase 3 modulates stress tolerance and carbohydrate metabolism. Plant J 49:1076–1090

Kholodenko BN, Kiyatkin A, Bruggeman FJ, Sontag E, Westerhoff HV, Hoek JB (2002) Untangling the wires: a strategy to trace functional interactions in signaling and gene networks. Proc Natl Acad Sci USA 99:12841–12846

Kiba T, Kudo T, Kojima M, Sakakibara H (2011) Hormonal control of nitrogen acquisition: roles of auxin, abscisic acid, and cytokinin. J Exp Bot 62:1399–1409

Kiefer P, Heinzle E, Zelder O, Wittmann C (2004) Comparative metabolic flux analysis of lysine-producing Corynebacterium glutamicum cultured on glucose or fructose. Appl Environ Microbiol 70:229–239

Kim ST, Cho KS, Yu S, Kim SG, Hong JC, Han CD, Bae DW, Nam MH, Kang KY (2003) Proteomic analysis of differentially expressed proteins induced by rice blast fungus and elicitor in suspension-cultured rice cells. Proteomics 3:2368–2378

Kimura M, Yamamoto YY, Seki M, Sakurai T, Sato M, Abe T, Yoshida S, Manabe K, Shinozaki K, Matsui M (2003) Identification of Arabidopsis genes regulated by high light-stress using cDNA microarray. Photochem Photobiol 77:226–233

Kishor PBK, Sangam S, Amrutha RN, Laxmi PS, Naidu KR, Rao KRSS, Rao S, Reddy KJ, Theriappan P, Sreenivasulu N (2005) Regulation of proline biosynthesis, degradation, uptake and transport in higher plants: Its implications in plant growth and abiotic stress tolerance. Curr Sci India 88:424–438

Kitano H (2000) Perspectives on systems biology. New Generat Comput 18:199–216

Kogel KH, Voll LM, Schafer P, Jansen C, Wu YC, Langen G, Imani J, Hofmann J, Schmiedl A, Sonnewald S, von Wettstein D, Cook RJ, Sonnewald U (2010) Transcriptome and metabolome profiling of field-grown transgenic barley lack induced differences but show cultivar-specific variances. Proc Natl Acad Sci USA 107:6198–6203

Kohli A, Narciso JO, Miro B, Raorane M (2012) Root proteases: reinforced links between nitrogen uptake and mobilization and drought tolerance. Physiol Plant 145:165–179

Komatsu S, Yano H (2006) Update and challenges on proteomics in rice. Proteomics 6:4057–4068

Kondou Y, Higuchi M, Takahashi S, Sakurai T, Ichikawa T, Kuroda H, Yoshizumi T, Tsumoto Y, Horii Y, Kawashima M, Hasegawa Y, Kuriyama T, Matsui K, Kusano M, Albinsky D, Takahashi H, Nakamura Y, Suzuki M, Sakakibara H, Kojima M, Akiyama K, Kurotani A, Seki M, Fujita M, Enju A, Yokotani N, Saitou T, Ashidate K, Fujimoto N, Ishikawa Y, Mori Y, Nanba R, Takata K, Uno K, Sugano S, Natsuki J, Dubouzet JG, Maeda S, Ohtake M,

Mori M, Oda K, Takatsuji H, Hirochika H, Matsui M (2009) Systematic approaches to using the FOX hunting system to identify useful rice genes. Plant J 57:883–894

Koornneef M, AlonsoBlanco C, Peeters AJM (1997) Genetic approaches in plant physiology. New Phytol 137:1–8

Kremling A, Fischer S, Gadkar K, Doyle FJ, Sauter T, Bullinger E, Allgower F, Gilles ED (2004) A benchmark for methods in reverse engineering and model discrimination: Problem formulation and solutions. Genome Res 14:1773–1785

Kreps JA, Wu Y, Chang HS, Zhu T, Wang X, Harper JF (2002) Transcriptome changes for Arabidopsis in response to salt, osmotic, and cold stress. Plant Physiol 130:2129–2141

Krishnan A, Guiderdoni E, An G, Hsing YIC, Han CD, Lee MC, Yu SM, Upadhyaya N, Ramachandran S, Zhang QF, Sundaresan V, Hirochika H, Leung H, Pereira A (2009) Mutant resources in rice for functional genomics of the grasses. Plant Physiol 149:165–170

Kruger NJ, Le Lay P, Ratcliffe RG (2007) Vacuolar compartmentation complicates the steady-state analysis of glucose metabolism and forces reappraisal of sucrose cycling in plants. Phytochemistry 68:2189–2196

Kumar CS, Wing RA, Sundaresan V (2005) Efficient insertional mutagenesis in rice using the maize En/Spm elements. Plant J 44:879–892

Kuromori T, Takahashi S, Kondou Y, Shinozaki K, Matsui M (2009) phenome analysis in plant species using loss-of-function and gain-of-function mutants. Plant Cell Physiol 50:1215–1231

Larkindale J, Vierling E (2008) Core genome responses involved in acclimation to high temperature. Plant Physiol 146:748–761

Lee TH, Kim YK, Pham TT, Song SI, Kim JK, Kang KY, An G, Jung KH, Galbraith DW, Kim M, Yoon UH, Nahm BH (2009) RiceArrayNet: a database for correlating gene expression from transcriptome profiling, and its application to the analysis of coexpressed genes in rice. Plant Physiol 151:16–33

Li Y, Shrestha B, Vertes A (2008) Atmospheric pressure infrared MALDI imaging mass spectrometry for plant metabolomics. Anal Chem 80:407–420

Lim CJ, Yang KA, Hong JK, Choi JS, Yun DJ, Hong JC, Chung WS, Lee SY, Cho MJ, Lim CO (2006) Gene expression profiles during heat acclimation in Arabidopsis thaliana suspension-culture cells. J Plant Res 119:373–383

Lugan R, Niogret MF, Leport L, Guegan JP, Larher FR, Savoure A, Kopka J, Bouchereau A (2010) Metabolome and water homeostasis analysis of Thellungiella salsuginea suggests that dehydration tolerance is a key response to osmotic stress in this halophyte. Plant J 64:215–229

Majoul T, Bancel E, Triboi E, Ben Hamida J, Branlard G (2003) Proteomic analysis of the effect of heat stress on hexaploid wheat grain: characterization of heat-responsive proteins from total endosperm. Proteomics 3:175–183

Majoul T, Bancel E, Triboi E, Ben Hamida J, Branlard G (2004) Proteomic analysis of the effect of heat stress on hexaploid wheat grain: characterization of heat-responsive proteins from non-prolamins fraction. Proteomics 4:505–513

Marbach D, Prill RJ, Schaffter T, Mattiussi C, Floreano D, Stolovitzky G (2010) Revealing strengths and weaknesses of methods for gene network inference. Proc Natl Acad Sci USA 107:6286–6291

Masoudi-Nejad A, Goto S, Jauregui R, Ito M, Kawashima S, Moriya Y, Endo TR, Kanehisa M (2007) EGENES: transcriptome-based plant database of genes with metabolic pathway information and expressed sequence tag indices in KEGG. Plant Physiol 144:857–866

Matsuda F, Morino K, Ano R, Kuzawa M, Wakasa K, Miyagawa H (2005) Metabolic flux analysis of the phenylpropanoid pathway in elicitor-treated potato tuber tissue. Plant Cell Physiol 46:454–466

Matsuda F, Shinbo Y, Oikawa A, Hirai MY, Fiehn O, Kanaya S, Saito K (2009) Assessment of metabolome annotation quality: a method for evaluating the false discovery rate of elemental composition searches. PloS One 4(10):e7490

Mattioli R, Falasca G, Sabatini S, Altamura MM, Costantino P, Trovato M (2009) The proline biosynthetic genes P5CS1 and P5CS2 play overlapping roles in Arabidopsis flower transition but not in embryo development. Physiol Plant 137:72–85

Mattioli R, Marchese D, D'Angeli S, Altamura MM, Costantino P, Trovato M (2008) Modulation of intracellular proline levels affects flowering time and inflorescence architecture in Arabidopsis. Plant Mol Biol 66:277–288

Mayer KF, Martis M, Hedley PE, Simkova H, Liu H, Morris JA, Steuernagel B, Taudien S, Roessner S, Gundlach H, Kubalakova M, Suchankova P, Murat F, Felder M, Nussbaumer T, Graner A, Salse J, Endo T, Sakai H, Tanaka T, Itoh T, Sato K, Platzer M, Matsumoto T, Scholz U, Dolezel J, Waugh R, Stein N (2011) Unlocking the barley genome by chromosomal and comparative genomics. Plant Cell 23:1249–1263

Mechin V, Balliau T, Chateau-Joubert S, Davanture M, Langella O, Negroni L, Prioul JL, Thevenot C, Zivy M, Damerval C (2004) A two-dimensional proteome map of maize endosperm. Phytochemistry 65:1609–1618

Mir RR, Zaman-Allah M, Sreenivasulu N, Trethowan R, Varshney RK (2012) Integrated genomics, physiology and breeding approaches for improving drought tolerance in crops. Theor Appl Genet 125:625–645

Mittler R (2006) Abiotic stress, the field environment and stress combination. Trends Plant Sci 11:15–19

Miyao A, Iwasaki Y, Kitano H, Itoh J, Maekawa M, Murata K, Yatou O, Nagato Y, Hirochika H (2007) A large-scale collection of phenotypic data describing an insertional mutant population to facilitate functional analysis of rice genes. Plant Mol Biol 63:625–635

Mochida K, Uehara-Yamaguchi Y, Yoshida T, Sakurai T, Shinozaki K (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

Mohammadi M, Kav NN, Deyholos MK (2007) Transcriptional profiling of hexaploid wheat (Triticum aestivum L.) roots identifies novel, dehydration-responsive genes. Plant Cell Environ 30:630–645

Moles CG, Mendes P, Banga JR (2003) Parameter estimation in biochemical pathways: a comparison of global optimization methods. Genome Res 13:2467–2474

Montero-Barrientos M, Hermosa R, Cardoza RE, Gutierrez S, Nicolas C, Monte E (2010) Transgenic expression of the Trichoderma harzianum hsp70 gene increases Arabidopsis resistance to heat and other abiotic stresses. J Plant Physiol 167:659–665

Mullarkey M, Jones P (2000) Isolation and analysis of thermotolerant mutants of wheat. J Exp Bot 51:139–146

Mutwil M, Klie S, Tohge T, Giorgi FM, Wilkins O, Campbell MM, Fernie AR, Usadel B, Nikoloski Z, Persson S (2011) PlaNet: combined sequence and expression comparisons across plant networks derived from seven species. Plant Cell 23:895–910

Nakagami H, Sugiyama N, Ishihama Y, Shirasu K (2012) Shotguns in the front line: phosphoproteomics in plants. Plant Cell Physiol 53:118–124

Nakagami H, Sugiyama N, Mochida K, Daudi A, Yoshida Y, Toyoda T, Tomita M, Ishihama Y, Shirasu K (2010) Large-scale comparative phosphoproteomics identifies conserved phosphorylation sites in plants. Plant Physiol 153:1161–1174

Needham CJ, Manfield IW, Bulpitt AJ, Gilmartin PM, Westhead DR (2009) From gene expression to gene regulatory networks in Arabidopsis thaliana. BMC Syst Biol 3(1):85

Nishiyama R, Watanabe Y, Fujita Y, Le DT, Kojima M, Werner T, Vankova R, Yamaguchi-Shinozaki K, Shinozaki K, Kakimoto T, Sakakibara H, Schmulling T, Tran LS (2011) Analysis of cytokinin mutants and regulation of cytokinin metabolic genes reveals important regulatory roles of cytokinins in drought, salt and abscisic acid responses, and abscisic acid biosynthesis. Plant Cell 23:2169–2183

Obayashi T, Kinoshita K, Nakai K, Shibaoka M, Hayashi S, Saeki M, Shibata D, Saito K, Ohta H (2007) ATTED-II: a database of co-expressed genes and cis elements for identifying co-regulated gene groups in Arabidopsis. Nucleic Acids Res 35:D863–D869

Obayashi T, Nishida K, Kasahara K, Kinoshita K (2011) ATTED-II updates: condition-specific gene coexpression to extend coexpression analyses and applications to a broad range of flowering plants. Plant Cell Physiol 52:213–219

Okazaki Y, Shimojima M, Sawada Y, Toyooka K, Narisawa T, Mochida K, Tanaka H, Matsuda F, Hirai A, Hirai MY, Ohta H, Saito K (2009) A chloroplastic UDP-glucose pyrophosphorylase from arabidopsis is the committed enzyme for the first step of sulfolipid biosynthesis. Plant Cell 21:892–909

Oksman-Caldentey KM, Saito K (2005) Integrating genomics and metabolomics for engineering plant metabolic pathways. Curr Opin Biotechnol 16:174–179

Oliver SG, Winson MK, Kell DB, Baganz F (1998) Systematic functional analysis of the yeast genome. Trends Biotechnol 16:373–378

O'Malley RC, Ecker JR (2010) Linking genotype to phenotype using the Arabidopsis unimutant collection. Plant J 61:928–940

Oono Y, Seki M, Nanjo T, Narusaka M, Fujita M, Satoh R, Satou M, Sakurai T, Ishida J, Akiyama K, Iida K, Maruyama K, Satoh S, Yamaguchi-Shinozaki K, Shinozaki K (2003) Monitoring expression profiles of Arabidopsis gene expression during rehydration process after dehydration using ca 7000 full-length cDNA microarray. Plant J: Cell Mol Biol 34:868–887

Osuna D, Usadel B, Morcuende R, Gibon Y, Blasing OE, Hohne M, Gunter M, Kamlage B, Trethewey R, Scheible WR, Stitt M (2007) Temporal responses of transcripts, enzyme activities and metabolites after adding sucrose to carbon-deprived Arabidopsis seedlings. Plant J 49:463–491

Palanivelu R, Brass L, Edlund AF, Preuss D (2003) Pollen tube growth and guidance is regulated by POP2, an Arabidopsis gene that controls GABA levels. Cell 114:47–59

Pasam RK, Sharma R, Malosetti M, van Eeuwijk FA, Haseneyer G, Kilian B, Graner A (2012) Genome-wide association studies for agronomical traits in a world wide spring barley collection. BMC Plant Biol 12:16

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, Mehboobur R, Ware D, Westhoff P, Mayer KF, Messing J, Rokhsar DS (2009) The Sorghum bicolor genome and the diversification of grasses. Nature 457:551–556

Paux E, Sourdille P, Salse J, Saintenac C, Choulet F, Leroy P, Korol A, Michalak M, Kianian S, Spielmeyer W, Lagudah E, Somers D, Kilian A, Alaux M, Vautrin S, Berges H, Eversole K, Appels R, Safar J, Simkova H, Dolezel J, Bernard M, Feuillet C (2008) A physical map of the 1-gigabase bread wheat chromosome 3B. Science 322:101–104

Peleg Z, Blumwald E (2011) Hormone balance and abiotic stress tolerance in crop plants. Curr Opin Plant Biol 14:290–295

Peleg Z, Reguera M, Tumimbang E, Walia H, Blumwald E (2011) Cytokinin-mediated source/sink modifications improve drought tolerance and increase grain yield in rice under water-stress. Plant Biotechnol J 9:747–758

Pennisi E (2008) Plant genetics: the blue revolution, drop by drop, gene by gene. Science 320:171–173

Persson S, Wei HR, Milne J, Page GP, Somerville CR (2005) Identification of genes required for cellulose synthesis by regression analysis of public microarray data sets. Proc Natl Acad Sci USA 102:8633–8638

Pinheiro C, Chaves MM (2011) Photosynthesis and drought: can we make metabolic connections from available data? J Exp Bot 62:869–882

Qureshi MI, Qadir S, Zolla L (2007) Proteomics-based dissection of stress-responsive pathways in plants. J Plant Physiol 164:1239–1260

Rabbani MA, Maruyama K, Abe H, Khan MA, Katsura K, Ito Y, Yoshiwara K, Seki M, Shinozaki K, Yamaguchi-Shinozaki K (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

Rajendran K, Tester M, Roy SJ (2009) Quantifying the three main components of salinity tolerance in cereals. Plant Cell Environ 32:237–249

Ratcliffe RG, Shachar-Hill Y (2006) Measuring multiple fluxes through plant metabolic networks. Plant J 45:490–511

Reynolds TL, Nemeth MA, Glenn KC, Ridley WP, Astwood JD (2005) Natural variability of metabolites in maize grain: differences due to genetic background. J Agric Food Chem 53:10061–10067

Riccardi F, Gazeau P, de Vienne D, Zivy M (1998) Protein changes in response to progressive water deficit in maize. Quantitative variation and polypeptide identification. Plant Physiol 117:1253–1263

Rizhsky L, Liang HJ, Shuman J, Shulaev V, Davletova S, Mittler R (2004) When defense pathways collide. The response of Arabidopsis to a combination of drought and heat stress. Plant Physiol 134:1683–1696

Rohlig RM, Eder J, Engel KH (2009) Metabolite profiling of maize grain: differentiation due to genetics and environment. Metabolomics 5:459–477

Rolletschek H, Melkus G, Grafahrend-Belau E, Fuchs J, Heinzel N, Schreiber F, Jakob PM, Borisjuk L (2011) Combined noninvasive imaging and modeling approaches reveal metabolic compartmentation in the barley endosperm. Plant Cell 23:3041–3054

Roscher A, Kruger NJ, Ratcliffe RG (2000) Strategies for metabolic flux analysis in plants using isotope labelling. J Biotechnol 77:81–102

Rossel JB, Wilson IW, Pogson BJ (2002) Global changes in gene expression in response to high light in Arabidopsis. Plant Physiol 130:1109–1120

Rostoks N, Mudie S, Cardle L, Russell J, Ramsay L, Booth A, Svensson JT, Wanamaker SI, Walia H, Rodriguez EM, Hedley PE, Liu H, Morris J, Close TJ, Marshall DF, Waugh R (2005) Genome-wide SNP discovery and linkage analysis in barley based on genes responsive to abiotic stress. Mol Genet Genomics 274:515–527

Royal Society (2009) Reaping the benefits: science and the sustainable intensification of global agriculture. London (The Royal Society; Policy Document 11/09)

Safar J, Bartos J, Janda J, Bellec A, Kubalakova M, Valarik M, Pateyron S, Weiserova J, Tuskova R, Cihalikova J, Vrana J, Simkova H, Faivre-Rampant P, Sourdille P, Caboche M, Bernard M, Dolezel J, Chalhoub B (2004) Dissecting large and complex genomes: flow sorting and BAC cloning of individual chromosomes from bread wheat. Plant J: Cell Mol Biol 39:960–968

Safar J, Simkova H, Kubalakova M, Cihalikova J, Suchankova P, Bartos J, Dolezel J (2010) Development of chromosome-specific BAC resources for genomics of bread wheat. Cytogenet Genome Res 129:211–223

Saisho D, Takeda K (2011) Barley: emergence as a new research material of crop science. Plant Cell Physiol 52:724–727

Saito K, Hirai MY, Yonekura-Sakakibara K (2008) Decoding genes with coexpression networks and metabolomics: 'majority report by precogs'. Trends Plant Sci 13:36–43

Sakurai T, Kondou Y, Akiyama K, Kurotani A, Higuchi M, Ichikawa T, Kuroda H, Kusano M, Mori M, Saitou T, Sakakibara H, Sugano S, Suzuki M, Takahashi H, Takahashi S, Takatsuji H, Yokotani N, Yoshizumi T, Saito K, Shinozaki K, Oda K, Hirochika H, Matsui M (2011) RiceFOX: a database of Arabidopsis mutant lines overexpressing rice full-length cDNA that contains a wide range of trait information to facilitate analysis of gene function. Plant Cell Physiol 52:265–273

Salekdeh GH, Siopongco J, Wade LJ, Ghareyazie B, Bennett J (2002) Proteomic analysis of rice leaves during drought stress and recovery. Proteomics 2:1131–1145

Samach A, Onouchi H, Gold SE, Ditta GS, Schwarz-Sommer Z, Yanofsky MF, Coupland G (2000) Distinct roles of CONSTANS target genes in reproductive development of Arabidopsis. Science 288:1613–1616

Sanchez DH, Siahpoosh MR, Roessner U, Udvardi M, Kopka J (2008) Plant metabolomics reveals conserved and divergent metabolic responses to salinity. Physiol Plant 132:209–219

Sangam S, Jayasree D, Reddy KJ, Chari PVB, Sreenivasulu N, Kavi Kishor PB (2005) Salt tolerance in plants-transgenic approaches. J Plant Biotechnol. 7:1–15

SanMiguel P, Gaut BS, Tikhonov A, Nakajima Y, Bennetzen JL (1998) The paleontology of intergene retrotransposons of maize. Nat Genet 20:43–45

Sato S, Arita M, Soga T, Nishioka T, Tomita M (2008) Time-resolved metabolomics reveals metabolic modulation in rice foliage. BMC Syst Biol 2(1):51

Satoh-Nagasawa N, Nagasawa N, Malcomber S, Sakai H, Jackson D (2006) A trehalose metabolic enzyme controls inflorescence architecture in maize. Nature 441:227–230

Sauer U, Lasko DR, Fiaux J, Hochuli M, Glaser R, Szyperski T, Wuthrich K, Bailey JE (1999) Metabolic flux ratio analysis of genetic and environmental modulations of Escherichia coli central carbon metabolism. J Bacteriol 181:6679–6688

Sawada Y, Akiyama K, Sakata A, Kuwahara A, Otsuki H, Sakurai T, Saito K, Hirai MY (2009) Widely targeted metabolomics based on large-scale MS/MS data for elucidating metabolite accumulation patterns in plants. Plant Cell Physiol 50:37–47

Schaeffer ML, Harper LC, Gardiner JM, Andorf CM, Campbell DA, Cannon EKS, Sen TZ, Lawrence CJ (2011) MaizeGDB: curation and outreach go hand-in-hand. Database-Oxford

Schauer N, Fernie AR (2006) Plant metabolomics: towards biological function and mechanism. Trends Plant Sci 11:508–516

Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, Scholkopf B, Weigel D, Lohmann JU (2005) A gene expression map of Arabidopsis thaliana development. Nat Genet 37:501–506

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, Minx P, Reily AD, Courtney L, Kruchowski SS, Tomlinson C, Strong C, Delehaunty K, Fronick C, Courtney B, Rock SM, Belter E, Du F, Kim K, Abbott RM, Cotton M, Levy A, Marchetto P, Ochoa K, Jackson SM, Gillam B, Chen W, Yan L, Higginbotham J, Cardenas M, Waligorski J, Applebaum E, Phelps L, Falcone J, Kanchi K, Thane T, Scimone A, Thane N, Henke J, Wang T, Ruppert J, Shah N, Rotter K, Hodges J, Ingenthron E, Cordes M, Kohlberg S, Sgro J, Delgado B, Mead K, Chinwalla A, Leonard S, Crouse K, Collura K, Kudrna D, Currie J, He R, Angelova A, Rajasekar S, Mueller T, Lomeli R, Scara G, Ko A, Delaney K, Wissotski M, Lopez G, Campos D, Braidotti M, Ashley E, Golser W, Kim H, Lee S, Lin J, Dujmic Z, Kim W, Talag J, Zuccolo A, Fan C, Sebastian A, Kramer M, Spiegel L, Nascimento L, Zutavern T, Miller B, Ambroise C, Muller S, Spooner W, Narechania A, Ren L, Wei S, Kumari S, Faga B, Levy MJ, McMahan L, Van Buren P, Vaughn MW, Ying K, Yeh CT, Emrich SJ, Jia Y, Kalyanaraman A, Hsia AP, Barbazuk WB, Baucom RS, Brutnell TP, Carpita NC, Chaparro C, Chia JM, Deragon JM, Estill JC, Fu Y, Jeddeloh JA, Han Y, Lee H, Li P, Lisch DR, Liu S, Liu Z, Nagel DH, McCann MC, SanMiguel P, Myers AM, Nettleton D, Nguyen J, Penning BW, Ponnala L, Schneider KL, Schwartz DC, Sharma A, Soderlund C, Springer NM, Sun Q, Wang H, Waterman M, Westerman R, Wolfgruber TK, Yang L, Yu Y, Zhang L, Zhou S, Zhu Q, Bennetzen JL, Dawe RK, Jiang J, Jiang N, Presting GG, Wessler SR, Aluru S, Martienssen RA, Clifton SW, McCombie WR, Wing RA, Wilson RK (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112–1115

Schulte D, Ariyadasa R, Shi B, Fleury D, Saski C, Atkins M, de Jong P, Wu CC, Graner A, Langridge P, Stein N (2011) BAC library resources for map-based cloning and physical map construction in barley (Hordeum vulgare L.). BMC Genomics 12(1):247

Schwab W (2003) Metabolome diversity: too few genes, too many metabolites? Phytochemistry 62:837–849

Seebauer JR, Moose SP, Fabbri BJ, Crossland LD, Below FE (2004) Amino acid metabolism in maize earshoots. Implications for assimilate preconditioning and nitrogen signaling. Plant Physiol 136:4326–4334

Seiler C, Harshavardhan VT, Rajesh K, Reddy PS, Strickert M, Rolletschek H, Scholz U, Wobus U, Sreenivasulu N (2011) ABA biosynthesis and degradation contributing to ABA homeostasis during barley seed development under control and terminal drought-stress conditions. J Exp Bot 62:2615–2632

Seki M, Ishida J, Narusaka M, Fujita M, Nanjo T, Umezawa T, Kamiya A, Nakajima M, Enju A, Sakurai T, Satou M, Akiyama K, Yamaguchi-Shinozaki K, Carninci P, Kawai J, Hayashizaki Y, Shinozaki K (2002a) Monitoring the expression pattern of around 7,000 Arabidopsis genes under ABA treatments using a full-length cDNA microarray. Funct Integr Genomics 2:282–291

Seki M, Narusaka M, Abe H, Kasuga M, Yamaguchi-Shinozaki K, Carninci P, Hayashizaki Y, Shinozaki K (2001) Monitoring the expression pattern of 1,300 Arabidopsis genes under drought and cold stresses by using a full-length cDNA microarray. Plant Cell 13:61–72

Seki M, Narusaka M, Ishida J, Nanjo T, Fujita M, Oono Y, Kamiya A, Nakajima M, Enju A, Sakurai T, Satou M, Akiyama K, Taji T, Yamaguchi-Shinozaki K, Carninci P, Kawai J, Hayashizaki Y, Shinozaki K (2002b) Monitoring the expression profiles of 7,000 Arabidopsis genes under drought, cold and high-salinity stresses using a full-length cDNA microarray. Plant J: Cell Mol Biol 31:279–292

Shingaki-Wells RN, Huang SB, Taylor NL, Carroll AJ, Zhou WX, Millar AH (2011) Differential molecular responses of rice and wheat coleoptiles to anoxia reveal novel metabolic adaptations in amino acid metabolism for tissue tolerance. Plant Physiol 156:1706–1724

Shulaev V, Cortes D, Miller G, Mittler R (2008) Metabolomics for plant stress response. Physiol Plant 132:199–208

Smit B, Ludlow L, Brklacich M (1988) Implications of a global climatic warming for agriculture: a review and appraisal. J Environ Qual 17:519–527

Sreenivasulu N, Altschmied L, Radchuk V, Gubatz S, Wobus U, Weschke W (2004a) Transcript profiles and deduced changes of metabolic pathways in maternal and filial tissues of developing barley grains. Plant J: Cell Mol Biol 37:539–553

Sreenivasulu N, Miranda M, Prakash HS, Wobus U, Weschke W (2004b) Transcriptome changes in foxtail millet genotypes at high salinity: Identification and characterization of a PHGPX gene specifically up-regulated by NaCl in a salt-tolerant line. J Plant Physiol 161:467–477

Sreenivasulu N, Radchuk V, Strickert M, Miersch O, Weschke W, Wobus U (2006) Gene expression patterns reveal tissue-specific signaling networks controlling programmed cell death and ABA- regulated maturation in developing barley seeds. Plant J: Cell Mol Biol 47:310–327

Sreenivasulu N, Schnurbusch T (2012) A genetic playground for enhancing grain number in cereals. Trends Plant Sci 17:91–101

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

Sreenivasulu N, Graner A, Wobus U (2008a) Barley genomics: an overview. Int J Plant Genomics 2008:486258

Sreenivasulu N, Usadel B, Winter A, Radchuk V, Scholz U, Stein N, Weschke W, Strickert M, Close TJ, Stitt M, Graner A, Wobus U (2008b) Barley grain maturation and germination: metabolic pathway and regulatory network commonalities and differences highlighted by new MapMan/PageMan profiling tools. Plant Physiol 146:1738–1758

Sreenivasulu N, Sunkar R, Wobus U, Strickert M (2010) Array platforms and bioinformatics tools for the analysis of plant transcriptome in response to abiotic stress. Methods Mol Biol 639:71–93

Sreenivasulu N, Harshavardhan VT, Govind G, Seiler C, Kohli A (2012) Contrapuntal role of ABA: does it mediate stress tolerance or plant growth retardation under long-term drought stress? Gene 125:625–645

Styczynski MP, Stephanopoulos G (2005) Overview of computational methods for the inference of gene regulatory networks. Comput Chem Eng 29:519–534

Swarbreck D, Wilks C, Lamesch P, Berardini TZ, Garcia-Hernandez M, Foerster H, Li D, Meyer T, Muller R, Ploetz L, Radenbaugh A, Singh S, Swing V, Tissier C, Zhang P, Huala E (2008) The Arabidopsis information resource (TAIR): gene structure and function annotation. Nucleic Acid Res 36:D1009–D1014

Sweetlove LJ, Fell D, Fernie AR (2008) Getting to grips with the plant metabolic network. Biochem J 409:27–41

Szekely G, Abraham E, Cseplo A, Rigo G, Zsigmond L, Csiszar J, Ayaydin F, Strizhov N, Jasik J, Schmelzer E, Koncz C, Szabados L (2008) Duplicated P5CS genes of Arabidopsis play distinct roles in stress regulation and developmental control of proline biosynthesis. Plant J: Cell Mol Biol 53:11–28

Takahashi S, Seki M, Ishida J, Satou M, Sakurai T, Narusaka M, Kamiya A, Nakajima M, Enju A, Akiyama K, Yamaguchi-Shinozaki K, Shinozaki K (2004) Monitoring the expression profiles of genes induced by hyperosmotic, high salinity, and oxidative stress and abscisic acid treatment in Arabidopsis cell culture using a full-length cDNA microarray. Plant Mol Biol 56:29–55

Tanaka T, Antonio BA, Kikuchi S, Matsumoto T, Nagamura Y, Numa H, Sakai H, Wu J, Itoh T, Sasaki T, Aono R, Fujii Y, Habara T, Harada E, Kanno M, Kawahara Y, Kawashima H, Kubooka H, Matsuya A, Nakaoka H, Saichi N, Sanbonmatsu R, Sato Y, Shinso Y, Suzuki M, Takeda JI, Tanino M, Todokoro F, Yamaguchi K, Yamamoto N, Yamasaki C, Imanishi T, Okido T, Tada M, Ikeo K, Tateno Y, Gojobori T, Lin YC, Wei FJ, Hsing YI, Zhao Q, Han B, Kramer MR, McCombie RW, Lonsdale D, O'Donovan CC, Whitfield EJ, Apweiler R, Koyanagi KO, Khurana JP, Raghuvanshi S, Singh NK, Tyagi AK, Haberer G, Fujisawa M, Hosokawa S, Ito Y, Ikawa H, Shibata M, Yamamoto M, Bruskiewich RM, Hoen DR, Bureau TE, Namiki N, Ohyanagi H, Sakai Y, Nobushima S, Sakata K, Barrero RA, Sato Y, Souvorov A, Smith-White B, Tatusova T, An S, An G, OOta S, Fuks G, Messing J, Christie KR, Lieberherr D, Kim H, Zuccolo A, Wing RA, Nobuta K, Green PJ, Lu C, Meyers BC, Chaparro C, Piegu B, Panaud O, Echeverria M (2008) The rice annotation project database (RAP-DB): 2008 update. Nucleic Acid Res 36:D1028–D1033

Thimm O, Essigmann B, Kloska S, Altmann T, Buckhout TJ (2001) Response of Arabidopsis to iron deficiency stress as revealed by microarray analysis. Plant Physiol 127:1030–1043

Thole V, Worland B, Wright J, Bevan MW, Vain P (2010) Distribution and characterization of more than 1000 T-DNA tags in the genome of Brachypodium distachyon community standard line Bd21. Plant Biotechnol J 8:734–747

Umezawa T, Fujita M, Fujita Y, Yamaguchi-Shinozaki K, Shinozaki K (2006) Engineering drought tolerance in plants: discovering and tailoring genes to unlock the future. Curr Opin Biotech 17:113–122

Urano K, Maruyama K, Ogata Y, Morishita Y, Takeda M, Sakurai N, Suzuki H, Saito K, Shibata D, Kobayashi M, Yamaguchi-Shinozaki K, Shinozaki K (2009) Characterization of the ABA-regulated global responses to dehydration in Arabidopsis by metabolomics. Plant J: Cell Mol Biol 57:1065–1078

Usadel B, Blasing OE, Gibon Y, Poree F, Hohne M, Gunter M, Trethewey R, Kamlage B, Poorter H, Stitt M (2008) Multilevel genomic analysis of the response of transcripts, enzyme activities and metabolites in Arabidopsis rosettes to a progressive decrease of temperature in the non-freezing range. Plant Cell Environ 31:518–547

Vain P, Worland B, Thole V, McKenzie N, Alves SC, Opanowicz M, Fish LJ, Bevan MW, Snape JW (2008) Agrobacterium-mediated transformation of the temperate grass Brachypodium distachyon (genotype Bd21) for T-DNA insertional mutagenesis. Plant Biotechnol J 6:236–245

Van Baarlen P, Van Esse HP, Siezen RJ, Thomma BPHJ (2008) Challenges in plant cellular pathway reconstruction based on gene expression profiling. Trends Plant Sci 13:44–50

van Dijken AJH, Schluepmann H, Smeekens SCM (2004) Arabidopsis trehalose-6-phosphate synthase 1 is essential for normal vegetative growth and transition to flowering. Plant Physiol 135:969–977

Vensel WH, Tanaka CK, Cai N, Wong JH, Buchanan BB, Hurkman WJ (2005) Developmental changes in the metabolic protein profiles of wheat endosperm. Proteomics 5:1594–1611

Verpoorte R, Memelink J (2002) Engineering secondary metabolite production in plants. Curr Opin Biotech 13:181–187

Vij S, Tyagi AK (2007) Emerging trends in the functional genomics of the abiotic stress response in crop plants. Plant Biotechnol J 5:361–380

Vinocur B, Altman A (2005) Recent advances in engineering plant tolerance to abiotic stress: achievements and limitations. Curr Opin Biotech 16:123–132

Vlad MO, Arkin A, Ross J (2004) Response experiments for nonlinear systems with application to reaction kinetics and genetics. Proc Natl Acad Sci USA 101:7223–7228

Vogel J, Hill T (2008) High-efficiency Agrobacterium-mediated transformation of Brachypodium distachyon inbred line Bd21-3. Plant Cell Rep 27:471–478

Vogel JP, Tuna M, Budak H, Huo N, Gu YQ, Steinwand MA (2009) Development of SSR markers and analysis of diversity in Turkish populations of Brachypodium distachyon. BMC Plant Biol 9:88

Walia H, Wilson C, Condamine P, Ismail AM, Xu J, Cui X, Close TJ (2007) Array-based genotyping and expression analysis of barley cv. Maythorpe and Golden Promise. BMC genomics 8:87

Walia H, Wilson C, Wahid A, Condamine P, Cui X, Close TJ (2006) Expression analysis of barley (Hordeum vulgare L.) during salinity stress. Funct Integr Genomics 6:143–156

Wan JX, Griffiths R, Ying JF, McCourt P, Huang YF (2009) Development of drought-tolerant canola (Brassica napus L.) through genetic modulation of ABA-mediated stomatal responses. Crop Sci 49:1539–1554

Wang H, Schauer N, Usadel B, Frasse P, Zouine M, Hernould M, Latche A, Pech JC, Fernie AR, Bouzayen M (2009) Regulatory features underlying pollination-dependent and -independent tomato fruit set revealed by transcript and primary metabolite profiling. Plant Cell 21:1428–1452

Wang W, Vinocur B, Altman A (2003) Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. Planta 218:1–14

Wanjugi H, Coleman-Derr D, Huo N, Kianian SF, Luo MC, Wu J, Anderson O, Gu YQ (2009) Rapid development of PCR-based genome-specific repetitive DNA junction markers in wheat. Genome 52:576–587 National Research Council Canada, Conseil National de Recherches Canada

Wasinger VC, Cordwell SJ, Cerpa-Poljak A, Yan JX, Gooley AA, Wilkins MR, Duncan MW, Harris R, Williams KL, Humphery-Smith I (1995) Progress with gene-product mapping of the Mollicutes: Mycoplasma genitalium. Electrophoresis 16:1090–1094

Wei F, Zhang J, Zhou S, He R, Schaeffer M, Collura K, Kudrna D, Faga BP, Wissotski M, Golser W, Rock SM, Graves TA, Fulton RS, Coe E, Schnable PS, Schwartz DC, Ware D, Clifton SW, Wilson RK, Wing RA (2009) The physical and genetic framework of the maize B73 genome. PLoS Genet 5:e1000715

Weigel D, Ahn JH, Blazquez MA, Borevitz JO, Christensen SK, Fankhauser C, Ferrandiz C, Kardailsky I, Malancharuvil EJ, Neff MM, Nguyen JT, Sato S, Wang ZY, Xia YJ, Dixon RA, Harrison MJ, Lamb CJ, Yanofsky MF, Chory J (2000) Activation tagging in Arabidopsis. Plant Physiol 122:1003–1013

Westerhoff HV, Palsson BO (2004) The evolution of molecular biology into systems biology. Nat Biotechnol 22:1249–1252

Wicker T, Mayer KFX, Gundlach H, Martis M, Steuernagel B, Scholz U, Simkova H, Kubalakova M, Choulet F, Taudien S, Platzer M, Feuillet C, Fahima T, Budak H, Dolezel J, Keller B, Stein N (2011) Frequent gene movement and pseudogene evolution is common to the large and complex genomes of wheat, barley, and their relatives. Plant Cell 23:1706–1718

Wiechert W, Mollney M, Petersen S, de Graaf AA (2001) A universal framework for 13C metabolic flux analysis. Metab Eng 3:265–283

Williams TC, Poolman MG, Howden AJ, Schwarzlander M, Fell DA, Ratcliffe RG, Sweetlove LJ (2010) A genome-scale metabolic model accurately predicts fluxes in central carbon metabolism under stress conditions. Plant Physiol 154:311–323

Wittmann C, Kiefer P, Zelder O (2004) Metabolic fluxes in Corynebacterium glutamicum during lysine production with sucrose as carbon source. Appl Environ Microbiol 70:7277–7287

Worch S, Rajesh K, Harshavardhan VT, Pietsch C, Korzun V, Kuntze L, Borner A, Wobus U, Roder MS, Sreenivasulu N (2011) Haplotyping, linkage mapping and expression analysis of barley genes regulated by terminal drought stress influencing seed quality. BMC Plant Biol 11:1

Xu J, Tian J, Belanger FC, Huang B (2007) Identification and characterization of an expansin gene AsEXP1 associated with heat tolerance in C3 Agrostis grass species. J Exp Bot 58:3789–3796

Xue GP, McIntyre CL, Chapman S, Bower NI, Way H, Reverter A, Clarke B, Shorter R (2006) Differential gene expression of wheat progeny with contrasting levels of transpiration efficiency. Plant Mol Biol 61:863–881

Yamaguchi T, Nakayama K, Hayashi T, Yazaki J, Kishimoto N, Kikuchi S, Koike S (2004) cDNA microarray analysis of rice anther genes under chilling stress at the microsporogenesis stage revealed two genes with DNA transposon Castaway in the 5'-flanking region. Biosci Biotech Bioch 68:1315–1323

Yamakawa H, Hakata M (2010) Atlas of rice grain filling-related metabolism under high temperature: joint analysis of metabolome and transcriptome demonstrated inhibition of starch accumulation and induction of amino acid accumulation. Plant Cell Physiol 51:795–809

Yang S, Vanderbeld B, Wan J, Huang Y (2010) Narrowing down the targets: towards successful genetic engineering of drought-tolerant crops. Mol Plant 3:469–490

Youens-Clark K, Buckler E, Casstevens T, Chen C, DeClerck G, Derwent P, Dharmawardhana P, Jaiswal P, Kersey P, Karthikeyan AS, Lu J, McCouch SR, Ren LY, Spooner W, Stein JC, Thomason J, Wei S, Ware D (2011) Gramene database in 2010: updates and extensions. Nucleic Acids Res 39:D1085–D1094

Yu J, Hu SN, Wang J, Wong GKS, Li SG, Liu B, Deng YJ, Dai L, Zhou Y, Zhang XQ, Cao ML, Liu J, Sun JD, Tang JB, Chen YJ, Huang XB, Lin W, Ye C, Tong W, Cong LJ, Geng JN, Han YJ, Li L, Li W, Hu GQ, Huang XG, Li WJ, Li J, Liu ZW, Li L, Liu JP, Qi QH, Liu JS, Li L, Li T, Wang XG, Lu H, Wu TT, Zhu M, Ni PX, Han H, Dong W, Ren XY, Feng XL, Cui P, Li XR, Wang H, Xu X, Zhai WX, Xu Z, Zhang JS, He SJ, Zhang JG, Xu JC, Zhang KL, Zheng XW, Dong JH, Zeng WY, Tao L, Ye J, Tan J, Ren XD, Chen XW, He J, Liu DF, Tian W, Tian CG, Xia HG, Bao QY, Li G, Gao H, Cao T, Wang J, Zhao WM, Li P, Chen W, Wang XD, Zhang Y, Hu JF, Wang J, Liu S, Yang J, Zhang GY, Xiong YQ, Li ZJ, Mao L, Zhou CS, Zhu Z, Chen RS, Hao BL, Zheng WM, Chen SY, Guo W, Li GJ, Liu SQ, Tao M, Wang J, Zhu LH, Yuan LP, Yang HM (2002) A draft sequence of the rice genome (Oryza sativa L. ssp indica). Science 296:79–92

Zamboni N, Sauer U (2004) Model-independent fluxome profiling from 2H and 13C experiments for metabolic variant discrimination. Genome Biol 5:R99

Zeller G, Henz SR, Widmer CK, Sachsenberg T, Ratsch G, Weigel D, Laubinger S (2009) Stress-induced changes in the Arabidopsis thaliana transcriptome analyzed using whole-genome tiling arrays. Plant J: Cell Mol Biol 58:1068–1082

Zhang HX, Lian CL, Shen ZG (2009) Proteomic identification of small, copper-responsive proteins in germinating embryos of Oryza sativa. Ann Bot 103:923–930

Zhang YX, Wu RH, Qin GJ, Chen ZL, Gu HY, Qu LJ (2011) Over-expression of WOX1 leads to defects in meristem development and polyamine homeostasis in Arabidopsis. J Integr Plant Biol 53:493–506

Zhou S, Wei F, Nguyen J, Bechner M, Potamousis K, Goldstein S, Pape L, Mehan MR, Churas C, Pasternak S, Forrest DK, Wise R, Ware D, Wing RA, Waterman MS, Livny M, Schwartz DC (2009) A single molecule scaffold for the maize genome. PLoS Genet 5:e1000711

Zhu H, Bilgin M, Snyder M (2003) Proteomics. Annu Rev Biochem 72:783–812

Chapter 9
Functional Genomics of Seed Development in Cereals

Ming Li, Sergiy Lopato, Nataliya Kovalchuk and Peter Langridge

9.1 Introduction

Seeds are the product of sexual reproduction in flowering plants. The seeds of cereals are the main source of staple food, animal feed and the raw material of food and fiber-based industries worldwide (Olsen 2001). More recently, cereal seeds have been used as a source of starch for the production of biofuels, although this use has become controversial (Fischer et al. 2009). New strategies for raising grain production have become a high international priority to help feed a growing world population in a scenario where resources are limiting and climate variability is increasing (Tester and Langridge 2010). Abiotic and biotic stresses such as drought, frost/cold, salt, micronutrient-deficiency, heavy metal toxicity and damage caused by microbes and pests can lead to dramatic yield loss and have a great impact on seed quality.

Grain yield is determined by a range of factors related to overall plant development. These include the ability of the plant to access water and nutrients through the roots, the development of leaves and photosynthetic tissues for carbon fixation and storage and the processes of carbon and nutrient relocation during grain filling. However, many processes associated with fertilization and grain development are also critical in determining the final size, shape and composition of grains, which have an impact on yield. The fertilization process is not only important in establishing grain number but is also a prime target for modifying the reproductive strategy, for example in apomixis. Modification of these pathways and processes requires a detailed understanding of the molecular events starting from seed initiation up to seed maturity.

M. Li · S. Lopato · N. Kovalchuk · P. Langridge (✉)
Australian Centre for Plant Functional Genomics, University of Adelaide,
Waite Campus, Urrbrae, South Australia
e-mail: peter.langridge@acpfg.com.au

P. K. Gupta and R. K. Varshney (eds.), *Cereal Genomics II*,
DOI: 10.1007/978-94-007-6401-9_9,

9.2 Structure of Cereal Seeds

Due to a closely fused seed coat and fruit coat, cereal seed usually refers to the caryopsis, which consists of an embryo, a large endosperm and a mass of maternal tissues. Endosperm is the tissue of the main nutritional value, since it comprises over 80 % of the cereal grain. Mature endosperm consists of five types of cells, known as aleurone (AL), sub-aleurone, starchy endosperm (SE), embryo surrounding region (ESR) and endosperm transfer cells (ETC). Maternal tissues such as fruit and seed coat (testa) enclose the embryo and endosperm. The embryo contains two main parts; embryo axis and scutellum. The scutellum is responsible for the transport of nutrients to the developing embryo axis and later, during seed germination, it provides the route for sugar transport from the endosperm to the germinating embryo (Aoki et al. 2006). The nutritional value of the embryo and endosperm for human diets are different: the embryo is rich in lipids and enzymes while the endosperm is the storage site for starch and proteins. The aleurone layer is rich in soluble protein (about 50 %) and is also a source of enzymes, lipids and vitamins. Cell walls in the starchy endosperm are thinner compared with other cells in the seeds. Cells of starchy endosperm are packed with starch graduals embedded in a protein matrix. Figure 9.1 shows different grain tissues in sections of mature barley caryopses at harvest.

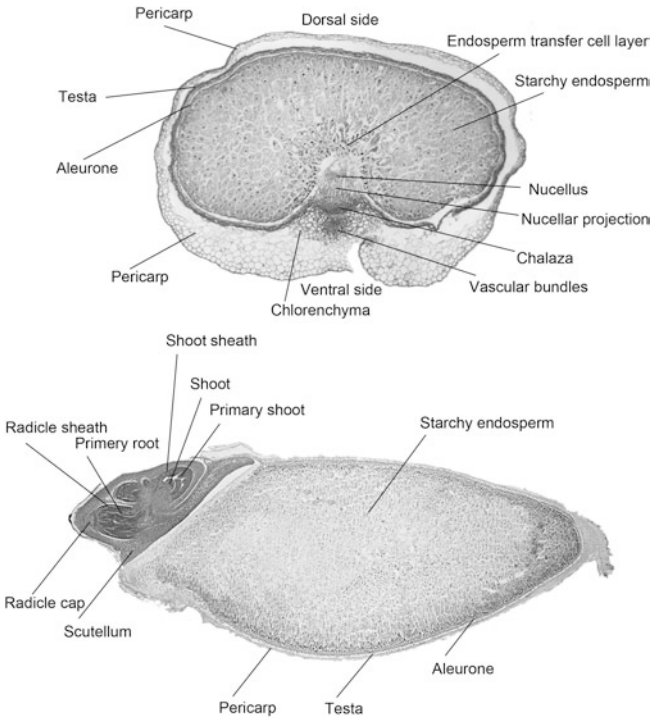

Fig. 9.1 Transverse and longitudinal sections of barley grain with the different grain tissues labelled

9.3 Methods in Cereal Seed Genomics

9.3.1 Micro-dissection

Traditional sampling methods used in cereal functional genomics often lead to a mixture of information because of the heterogeneous cell and tissue types present in the samples. Microdissection allows either isolation of specific cell or tissue types or a significant enrichment for particular cells or tissues of interest (reviewed in Brandt 2005; Nelson et al. 2006; Galbraith and Birnbaum 2006). Two methods are used in laser microdissection (LM): laser capture and laser cutting. In laser capture, cells of interest are captured from tissue sections onto a transfer film with the help of an infrared laser. In laser cutting, target cells are cut from tissue sections with the help of a UV laser and collected into a tube either by gravity or under the high pressure of a laser beam (Ohtsu et al. 2007).

LM is usually combined with high throughput technologies, including metabolomic, proteomic and transcript profiling (Ohtsu et al. 2007; Moco et al. 2009; Thiel et al. 2011). Transcript profiling and gene identification can be achieved by using LM in combination with sequencing. For example, transcriptome analysis of barley genes expressed in the nucellar projection and endosperm transfer cells during early endosperm development were successfully identified by using a 12 K microarray on tissue collected using the LM laser pressure catapulted method (Thiel et al. 2008).

In some cases, cells can be collected without any use of expensive LM instruments. For example, the presence of the large, liquid multinucleate syncytium during early endosperm development in wheat allowed collection of this fraction with a thin pipette tip. Sufficient RNA was sampled to allow generation of a yeast expression cDNA library. This library was used for the isolation of a number of grain specific genes (Kovalchuk et al. 2009, 2012a, b).

9.3.2 Yeast One–Hybrid and Two-Hybrid Screening

The yeast one-hybrid (Y1H) system provides a genetic assay for the identification of genes encoding proteins that bind to DNA elements of interest in vivo and has proved a useful tool to explore gene regulatory networks (Hens et al. 2012). Several different Y1H vector systems have been developed for the isolation of transcription factors and other DNA-binding proteins using characterized or predicted *cis*-elements or segments of gene promoters as baits (Deplanke et al. 2004; Chen et al. 2008; Klein and Dietz 2010; Reece-Hoyes et al. 2011). One example of the application of this method to isolate transcription factors potentially involved in regulating grain development resulted in the identification of more than 50 cDNA clones from cDNA libraries prepared from wheat, barley and maize spikes, grains and grain fractions (Pyvovarenko and Lopato 2011). The Y1H system

provides an opportunity to isolate genes encoding transcription factors and other DNA-binding proteins from species, which have complex unsequenced genomes and also have poor or limited cDNA resources. Provided the cDNA libraries are carefully prepared, the method allows isolation of long cDNAs belonging to genes with low abundance (Kovalchuk et al. 2012a).

Yeast two-hybrid (Y2H) screening has been used for the identification of novel partners in protein–protein interactions and confirmation of protein interactions predicted or identified using other methods. Novel members of protein complexes that are involved in the regulation of signal transduction pathways and control of grain developmental stages were identified using the Y2H approach. A wheat pre-mRNA splicing factor, *Ta*RSZ38, has been shown to be expressed in the embryo and also in the mitotically active part of syncytial and cellularizing wheat endosperm. Using *Ta*RSZ38 as bait in the Y2H screening, several cDNAs belonging to genes encoding known plant splicing factors were identified and some of these proteins were subsequently selected as baits for further Y2H screens. As a result, a large number of novel proteins involved in the pre-mRNA processing and mRNA transport have been reported (Lopato et al. 2006). The Y2H screen was also used to identify interacting partners of the ETC-specific lipid transfer protein TaPR60 (Kovalchuk et al. 2009) and grain-expressed NF-Y transcription factors (Lopato et al. unpublished).

Several interesting modifications of the Y2H system have been described. One of them, the nuclear transportation trap (NTT) system was used to isolate a number of nuclear proteins from rice NTT cDNA libraries (Moriguchi et al. 2005). Another efficient yeast-based system was developed for the isolation of plant cDNAs for genes encoding transcription factors and proteins with transcription activation functions (co-activators) in the developing rice embryo (Ye et al. 2004).

9.3.3 In vitro Fertilization (IVF)

Development of micromanipulation techniques, such as microdissection, facilitated selection and isolation, of single egg and sperm cells (reviewed in Weterings and Russell 2004). Fusion of defined gametes in vitro can be achieved by electrical or chemical methods, for example using calcium and polyethyleneglycol (Kranz and Lörz 1994; Kranz et al. 1998). IVF has provided new possibilities in the study of zygotes, primary endosperm cells, embryos and endosperm development after fertilization. It also enabled the analysis of events occurring immediately following fertilization. To date, complete IVF resulting in successful embryogenesis has been achieved only for maize and rice (for a review see Kranz and Scholten 2008).

By using molecular techniques and the IVF system, it is possible to perform the following studies: (1) identification of genes from isolated reproductive cells, (2) analysis of gene expression on a single-cell basis, (3) molecular analysis of differentially expressed genes with a targeted approach, (4) protein expression analysis, and (5) analysis of epigenetic modification (Kranz and Scholten 2008).

9.3.4 Somatic Embryogenesis

Highly differentiated somatic cells under optimum conditions remain competent to undergo embryogenesis; defined as totipotency. Non-zygotic or somatic embryos, commonly known as embryoids can be formed in tissue culture. The induced embryoids are structurally polarized and their developmental behavior is similar to zygotic embryos. Somatic embryogenesis was initially discovered in carrot and later proved to be a model system for analysis of hormonal regulation of cell competence and identification of proteins and molecules affecting cell fate (Raghavan 2006). Successful callus induction followed by embryogenesis was observed for a number of *Poaceae* species. However, due to slow growth of the embryo and rapid loss of embryogenesis potency, the application of the system was limited. In contrast, in some other monocotyledonous plants, such as orchard grass, embryogenesis can be easily achieved directly from the leaf mesophyll cells.

Several factors can affect somatic embryogenesis, including genotype, explant, tissue culture medium, growth regulators, source of carbohydrates and culture conditions such as temperature, light intensity and cycle. Somatic embryogenesis has a potential for large-scale plant propagation. Somatic embryos obtained in vitro are often free of viruses and pathogens; they can be a potential source for the production of synthetic seeds (Aquea et al. 2008; Thobunluepop et al. 2009).

9.3.5 Embryo Rescue and Endosperm Cultures

Major barriers to the formation of normal seeds can result from inter-specific or inter-generic crosses due to post-zygotic endosperm failure following double fertilization (Brink and Cooper 1947). These wide crosses have been important for cereal breeders since they allow the introgression of novel alleles for disease resistance, abiotic stress tolerance or grain quality from wild relatives. However, due to genomic incompatibility of the parents, double fertilization and endosperm formation is often prevented and these crosses frequently lead to seed abortion (reviewed in Haslam and Yeung 2011). Embryo rescue allows zygotic or immature embryos obtained from such crosses to grow and develop further into full-term embryos and to overcome seed dormancy (Raghavan 2003).

Embryo rescue has been successfully used for producing intergeneric hybrids involving a number of agriculturally important crops, particularly those belonging to Triticeae. The intergeneric combinations used for this purpose included the following: *Hordeum* × *Secale*, *Hordeum* × *Triticum*, *Hordeum* × *Agropyrum*, *Triticum* × *Aegilops* and *Triticum* × *Secale*.

Embryo culture has also been widely used in plant transformation since this tissue is usually highly receptive to transformation using both *Agrobacterium tumefaciens* and biolistic bombardment, thus making it possible to obtain transgenic plants (Tingay et al. 1997).

The first endosperm suspension culture was obtained for ryegrass and maize (Ashton and Polya 1978; Shannon 1982). These cultures were used for studies of hormonal regulation of cell division and expansion, and for the identification of genes controlling protein biosynthesis. In maize, effects of nitrogen-rich nutrients on the accumulation of seed storage proteins was also carried out using endosperm cell cultures derived from explants taken from developing seeds of several *opaque2* mutants harvested at 10 day after pollination. All mutants except Mo17o2R showed increased fresh and dry weight in response to increased nitrogen. The different response of Mo17o2R resulted from the absence of a transcription factor involved in the regulation of transcription of α-zein polypetides (Locatelli et al. 2001).

9.4 Development of Component Grain Tissues

9.4.1 Embryo Development

The life cycle of diploid angiosperm alternates between the diploid sporophytic and haploid gametophytic stages. The sporophytic phase usually lasts longer and supports the development of flowers where male and female gametophytes are produced. The male gametophyte or pollen grain develops in the anther and the female gametophyte, the embryo sac develops in the ovule (Raghavan 2006). In many angiosperm species, pollen grains usually comprise two cells, generative and vegetative cells. A polygonum type embryo sac consists of eight nuclei in seven cells defined as an egg, two synergids, three antipodals and one diploid central cell embedded in maternal ovule tissues (Drews and Yadegari 2002).

For the development of seed, several events take place sequentially starting from pollination to fertilization. Mature pollen grains hydrate and germinate after landing on stigmas, producing pollen tubes, which grow through the style and eventually reach the ovary. Upon arrival in the ovary, two sperm nuclei are released from the pollen tube into the embryo sac and migrate to the egg cell before fusion to the polar nuclei. During fertilization, one sperm nucleus fuses to the egg to form a diploid zygote, which develops into the embryo. The second sperm nucleus fertilizes the homo-diploid central cell which gives rise to a triploid endosperm. After fertilization, ovules develop into seeds containing the fertilized embryo and endosperm (Lopes and Larkins 1993; Reiser and Fischer 1993; Russell 1993; van-Went and Willemse 1984). The tissues derived from the sporophyte contribute to the remainder of the seed, such as the integuments, which develop into the seed coat and the ovary that becomes the bulk of the fruit (Koltunow 1993; Tucker et al. 2003; van-Went and Willemse 1984).

In angiosperms, the fertilized egg develops into a diploid zygote, which is the progenitor of the embryo. Cell division, expansion, differentiation and maturation are involved in the process of transforming a zygote into a mature embryo.

The polarity of the embryo is established with the first mitotic division. Cell divisions of the zygote occur synchronically with or later than those in the primary endosperm nucleus. Zygotes in some species may undergo a short "rest" period, which differs in length from a few hours to several days varying between species. In rice, the primary endosperm nucleus starts dividing about 3 h after fertilization (HAF), while zygotic nuclear division occurs at 6 HAF (Hu 1982).

The first four rounds of mitotic divisions following fertilization appear to be conserved in both monocots and eudicots. The apical-basal pattern known as embryo proper-suspensor is formed at the octant embryo stage, but embryo development in monocots differs from that in eudicots beyond this stage (Natesh and Rau 1984; Raghavan 2006). From embryo ontogeny to the mature seed, grasses (*Poaceae*), share some common features with other monocots and eudicots. However, the apical cell in monocots undergoes irregular divisions to generate the embryo proper, which shows radial symmetry while the basal cell produces a club-shaped suspensor complex at the same time (Raghavan 2006). In wheat, embryos initiate development of a scutellum and embryo axis at about 6 days after pollination (Hu 1982). The anterior part of the embryo axis gives rise to the coleoptile and shoot apex, while the posterior face forms the root apex. The mesocotyl lies between the nodes of the coleoptile and the scutellum (Raghavan 2006). Leaf primordia and other structures can be easily distinguished at about 40-45 DAP in the mature embryo (Hu 1982).

9.4.2 Endosperm Development

Three types of endosperm development, defined as nuclear, cellular and helobial, have been observed in angiosperms on the basis of whether there are walls formed with or without nuclear division (Vijayaraghavan and Prabhakar 1984). In the nuclear type, the primary endosperm cell undergoes mitotic divisions followed by repeated divisions of the daughter nuclei in the absence of wall formation. A distinct feature of cellular endosperm development is that the initial and subsequent divisions of the nuclei are accompanied by the formation of cell-plates. Hence, the endosperm remains in cellular form throughout development. Helobial type endosperm development is an intermediate type, in which the primary endosperm nucleus divides once, resulting in two cells differing in size and occupying a larger micropylar chamber and a smaller chalazal chamber. The chalazal chamber usually remains uninucleate or multinucleate resulting from a limited series of mitoses, whereas in the micropylar chamber, free-nuclear divisions take place before cytokinesis (Brink and Cooper 1947; Hu 1982; Vijayaraghavan and Prabhakar 1984).

In the nuclear endosperm development, several stages that are considered to be landmarks include syncytium, cellularization, differentiation and maturation (reviewed in Olsen 2001, 2004; Sreenivasulu et al. 2010). Many monocots, such as wheat, rice, barley and maize, and some eudicots like *Arabidopsis*, soybean

and cotton have nuclear endosperm and share the features of early endosperm development in the syncytial and cellular stages (Brown et al. 1999, 1994, 1996b; Klemsdal et al. 1991; Lid et al. 2004; Linnestad et al. 1998). The fate of endosperm cells differs dramatically after cellularization between some monocotyledonous and eudicotyledonous species, as demonstrated in cereals and *Arabidopsis*. Usually, the endosperm is persistent and turns into the storage site for cereals as the seed matures. By contrast, the endosperm of *Arabidopsis* is consumed by the developing embryo and degrades to a single cell layer in the mature seed, and in soybean it is absorbed entirely by the two distinct cotyledons (Linnestad et al. 1998).

Syncytial and cellularization phases of endosperm development are conserved among all groups of angiosperms (reviewed in Olsen 2001, 2004). The nuclear migration and cellularization happens in a radial manner initiated from the outmost layer of peripheral nuclei and gradually waves inwards to the center of the central cell. Microtubules are involved in endosperm ontogeny (Brown et al. 1994, 1997; Webb and Gunning 1991). In cereals, during the syncytial stage, successive and rapid mitotic divisions turn the triploid primary endosperm nucleus into a multinuclear peripheral lining of the cytoplasmic domain surrounding the large central vacuole (reviewed in Olsen 2004). A two-day mitotic interval in barley and, presumably in other cereals has been found for the period preceding cellularization (Linnestad et al. 1998). During this pause, the daughter nuclei that have already migrated to the distal cytoplasm are re-organized in the nuclear cytoplasmic domains (NCDs) by the radial-microtubule systems (RMS) set out from the surface of each nucleus at telophase. The anticlinal walls are deposited in the interzones restrained by the neighboring NCDs. No mitosis or phragmoplast is involved in this free-growing anticlinal wall (Brown et al. 1994, 1996b). The polarization of the coenocytic endosperm is marked with the elongation of NCDs along the axes perpendicular to the central cell wall, and intrude into the central vacuole. The unidirectional anticlinal walls continue growing, guided by adventitious phragmoplasts formed at the interfaces of microtubule systems emanating from the adjacent NCDs. As a consequence, the central cytoplasm is compartmentalized into a tube-like alveoli structure with the open-end pointing toward the central vacuole. Periclinal divisions followed by cytokinesis of the open-ended alveoli results in a peripheral cell layer along the integument and this growth moves inwards to the center of the enlarged embryo sac. The renewed anticlinal wall formation occurscentripetally to complete cellularization of the central cell (Brown et al. 1994, 1996a) (Fig. 9.2).

Comparison of aniline blue and immunofluorescence images show that all three types of walls, known as the free-growing anticlinal walls, anticlinal walls and periclinal walls, are deposited during the cellularization process and are rich in callose ($1 \rightarrow 3$ β-glucan) (Brown et al. 1994, 1997). Cell wall deposition in the multinucleate cytoplasm is wave-like, moving bidirectionally from the small ventral region near the nucellar projection towards the large dorsal area (Brown et al. 1996a; Mares et al. 1975) (Fig. 9.3).

In cereals, the free-nuclear divisions in the absence of cytokinesis last for a few days after pollination depending on growth conditions. The first division of the primary endosperm nucleus can be seen as early as 5 h after pollination

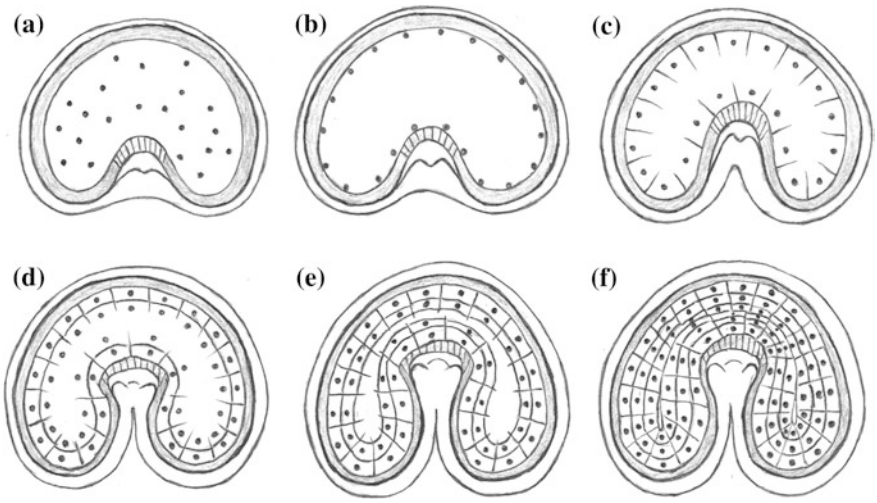

Fig. 9.2 Schematic representation of wheat endosperm cellularization process. **a** Free mitotic nuclear divisions in the coenocyte. **b** Nuclear migration to peripheral part of the multinucleate coenocyte. **c** Formation of the first anticlinal cell walls and nuclear divisions directed to the centre of multinucleate endosperm. **d** Formation of periclinal cell walls and growth of the second layer of anticlinal cell walls. **e** Periclinal divisions followed by cytokinesis of the open-ended alveoli resulting in one more peripheral cell layer. Further growth moves inwards to the center of partially cellularized endosperm until the end of cellularization (**f**)

(HAP) (Bennett et al. 1973). Wall formation starts from 3–4 days after pollination (DAP) and cellularization is complete for the entire endosperm at 6–8 DAP. In wheat, rice and maize, cellularization ends at 4–5 DAP, but ends 6–8 DAP in barley (Brown et al. 1994, 1996a; Lid et al. 2004; Mares et al. 1975, 1977; Olsen 2001). The outermost layer of peripheral cells at the dorsal surface and extending into the nucellar projection is rich in cytoplasm during the period of the first anticlinal wall deposition (Brown et al. 1996a, b; Mares et al. 1977). At 5 DAP, in wheat, these cells show signs of differentiation to form the Aleurone Layer (AL) at the end of cellularization (Mares et al. 1977). The ingrowths of anticlinal and peripheral walls perform centripetally from outer layers towards the center of the embryo sac and result in the Central Starchy Endosperm (CSE). Differentiation of the Transfer Cell Layer (TL) or Basal Endosperm Transfer Layer (BETL), and Embryo Surrounding Region (ESR) continue in the cellular endosperm until 21 DAP, when the grains reach their maximum size (Bosnes et al. 1992).

By using high-pressure-frozen/freezing substitution and electron microscopy, a detailed description of cell plate formation and the unique structure of the mini-phragmoplasts in the cytoplasmic endosperm have been observed (Otegui and Staehelin 2000). Besides callose, a new type of cell plate, which lacks fructose residues in backbones of xyloglucans synthesized in the Golgi body of the endosperm cytoplasm, was found to be a major component of the endosperm cell walls (Otegui et al. 2001; Otegui and Staehelin 2000).

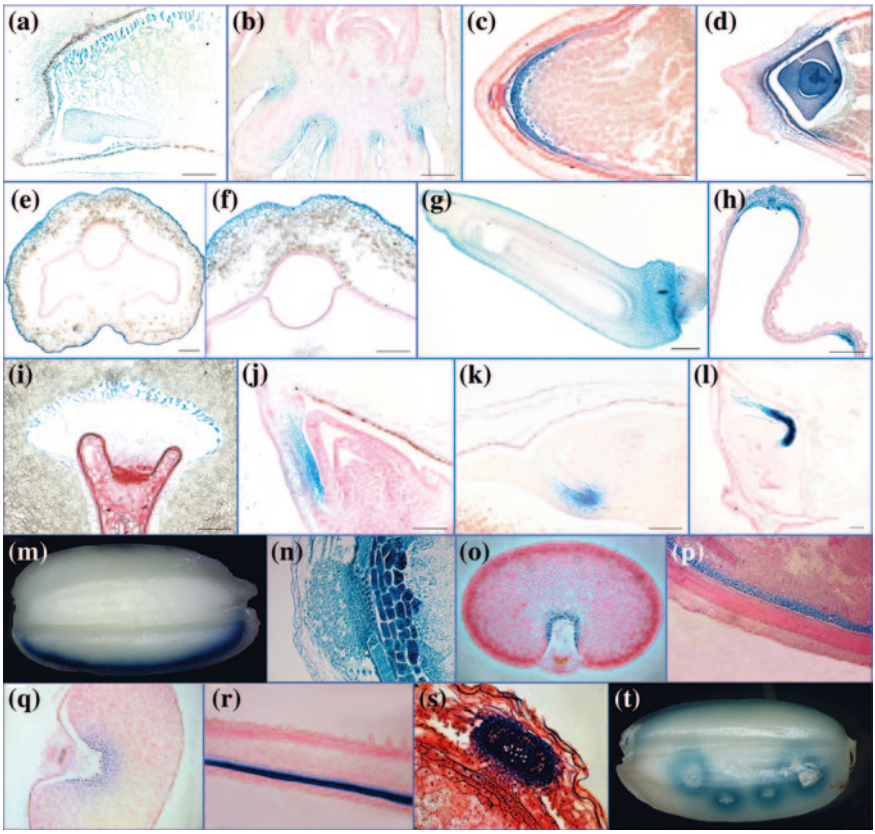

Fig. 9.3 Activity of grain specific promoters in transgenic wheat, barley and rice. **a**–**d** Activity of the *TdPR61* promoter in: **a** embryo and embryo-surrounding endosperm of wheat grain at 9 DAP, **b** radicals of the embryo axis of barley at 20 DAP, **c** rice ETC layers at 8 DAP, **d** rice embryo at 10 DAP. **e**–**h** Activity of the wheat defensin (*PRPI*) promoters in transgenic wheat and barley: **e**, **f** the *TdPRPI-10* promoter in epidermal layers of wheat grain at 3 DAP, **g** the *TdPRPI-11* promoter in the epidermal layers of rice grain at 2 DAP, **h** the *TdPRPI-11* promoter in the vascular tissues of rice lemma. **i** Activity of the *TdPR60* promoter in ETC of transgenic wheat at 31 DAP. **j**–**l**: **j** Activity of the *TdGL7* promoter in the main scutellar vascular bundle of transgenic wheat grain at 23 DAP, **k** barley grain at 28 DAP, **l** rice grain at 25 DAP. **m**–**o** Activity of the *OsPR60*2 promoter: **m**, **n** in ETC and vascular bundle of transgenic rice grain at 9 DAP, **o** in ETC layer of barley grain at 30 DAP. Activity of the *OsPR9a* promoter **p**, **q**: **p** in ETC of rice grain at12 DAP, **q** barley grain at 20 DAP. **r**, **s** Activity of the *OsPRPI* promoter: **r** in the vascular bundle of the rice lemma, **s** main vascular bundle of transgenic rice grain at 12 DAP. **t** Induction of the same promoter by mechanical wounding in epidermal layers of transgenic rice grain

Digital models of developing barley (*Hordeum vulgare*) grains were reconstructed from serial sections to visualize the complex three-dimensional (3D) grain anatomy. They provided detailed spatial descriptions of developing grains at anthesis, at the syncytial stage of endosperm development and at the onset of starch accumulation, visualizing and quantifying 18 tissues or tissue complexes (Gubatz et al. 2007).

The mature endosperm of cereals is persistent and contains at least four tissue types, the AL, BETL/TL, CSE and ESR (Klemsdal et al. 1991). Cell fate specification has been better understood through the analyses of endosperm development mutants in maize. However, little is known about endosperm cell specification in wheat, barley, rice and other small grain cereals. Moreover, the studies in maize have mainly focused on late stages of grain filling. The early stages, crucial for endosperm cell identity and development, have not been well studied. The four types of cells in the cereal endosperm will be discussed in the following sections.

Aleurone Layer

The cells of aleurone layer (AL) are believed to function in conditioning desiccation toward the end of seed maturation. During seed germination, upon receiving a gibberellic acid (GA) signal from the embryo, aleurone cells become active in producing hydrolytic enzymes including glucanases and proteinases for mobilization of the starch and storage proteins in the starchy endosperm (Fincher 1989; Lopes and Larkins 1993). This thin but uniformly arranged layer(s) of cells is morphologically distinct from the irregular-shaped starchy endosperm. There is one aleurone cell layer in wheat and maize, while three layers are present in barley and one to six layers in rice grains (reviewed in Olsen 2004; Becraft and Yi 2011).

The fate of the aleurone cell is most likely fixed after the first periclinal division in the barley endosperm alveoli at 5 DAP (Brown et al. 1994). Developing aleurone cells can be first distinguished at 8 DAP in barley, 10 DAP in wheat and 10–14 DAP in maize depending on the genotype (Becraft and Asuncion-Crabb 2000; Bosnes et al. 1992; Morrison et al. 1975, 1978).

It has been shown that specification of aleurone cells depends on positional signals set off from the embryo and also on cell signaling. Most gene products preferentially produced in the aleurone cells are also detected in the embryo (Aalen et al. 1994; Becraft and Asuncion-Crabb 2000; Opsahl-Sorteberg et al. 2004; Sreenivasulu et al. 2008). Characterization of *AtDEK1*, *ZmDEK1*, *ZmCR4*, and *ZmSAL1* (*SUPERNUMERARY ALEURONE LAYER* 1) suggests a current model for the specification of aleurone cell fate in maize and presumably also for epidermal cell differentiation in other plants (Tian et al. 2007). In this model, it has been proposed that DEK1 senses the positional cues in the outer plasma membrane of the cells located at the endosperm surface and possibly interacts with a membrane-bound transcription factor (Kim et al. 2006; Tian et al. 2007). The *dek1* mutants have pleiotropic phenotypes including a lack of aleurone cells, aborted embryos, carotenoid deficiency, and a soft, floury endosperm that is deficient in zeins. Recently, the *thick aleurone1* (*thk1*) mutant was described, which defines a novel negative function in the regulation of aleurone differentiation (Yi et al. 2011). The *thk1* mutants possess multiple layers of aleurone cells as well as aborted embryos. Double *thk1*/*dek1* mutants restored the ability of endosperm to accumulate carotenoids and zeins and to differentiate aleurone. Therefore, the *thk1* mutation

identified a negative regulator that functions downstream of *dek1* in the signaling system that controls aleurone specification and other aspects of endosperm development (Yi et al. 2011).

The gene *ZmCR4* encodes a receptor-like kinase involved in cell-to-cell signaling, suggesting that CR4 mediates lateral signaling of aleurone cell identity in the symplastic sub-domain of the specialized plasmodesmata. This proposed function is based on its high similarity to the mammalian Tumor Necrosis Factor Receptor (TNFR) (Becraft et al. 2001; Tian et al. 2007). SAL1 is located in the plasma membrane and is thought to maintain the balanced concentrations of DEK1 and CR4 by internalization and degradation through the SAL-positive endosomes (Shen et al. 2003; Tian et al. 2007). However, the expression of these three genes in diverse maize tissues implies that some of the cues necessary for aleurone cell specification are present throughout development. By studying the maize mutant, described as *extra cell layers 1* (*xcl1*), which has multiple epidermal cell layers including aleurone cells, Kessler and co-workers (2002) concluded that positional information is adopted by the early dividing cells without differentiation signals, while the late dividing cells acquire differentiation signals via lineage information. Another mutated gene *Hvdes5* whose product affects secondary wall formation in aleurone cells is also thought to be involved in the signaling process in the above model (Olsen et al. 2008). A better understanding of the mutated *des5* gene will provide more hints on the control of aleurone cell fate specification.

So far, many genes are found to be predominately expressed in aleurone cells, such as *LTP1*, *LTP2*, *Chi26*, *Chi33*, *B22E*, *ole-1*, *ole-2, pZE40* and *per-1* in barley, *Glu-B-1* in rice and *C1* in Maize (Becraft and Asuncion-Crabb 2000; Klemsdal et al. 1991; Leah et al. 1994; Lid et al. 2004; Madrid 1991; Opsahl-Sorteberg et al. 2004; Skriver et al. 1992; Smith et al. 1992; Stacy et al. 1999). More research needs to be done to focus on the biochemical pathways and cell-to-cell signaling in cereals to further optimize this model for aleurone cell identity specification.

Basal Endosperm Transfer Layer

In a transverse section from the middle of a developing caryopsis of a small grain cereal such as wheat, barley or rice, a few layers of cells over the nucellar projection, which join the aleurone layer on the ventral side, can be easily distinguished due to their thickened cell wall and dense cytoplasm. These are transfer cells, which constitute the so-called modified groove aleurone layer or crease aleurone layer (Drea et al. 2005; Klemsdal et al. 1991; Stacy et al. 1999). These cells help in transferring nutrients such as sucrose, monosaccharides and amino acids, from maternal vascular tissue into the developing seed for starch and protein biosyntheses and accumulation. Studies on sugar transport in developing seed of legumes suggest that the transfer-cell mediated uptake mechanism is an energy-coupled process. The wall in-growth during endosperm cellularization creates a gradient of enlarged plasma membrane surface area to promote solute transport (Serna et al. 2001). Transfer cells are thought to be involved in sugar transport from

the maternal tissue to the endosperm and in turn, sugar supplies directly affect transfer cell formation. Associated molecular markers from barley indicate that the fate of transfer cells is specified at the endosperm coenocyte stage (Becraft 2001; Lid et al. 2004). *ENDOSPERM1* (*END1*) transcripts are detected in the transfer-cell area of the syncytial barley endosperm over the nucellar projection in the developing barley grains at 6 DAP (Doan et al. 1996). At 10 DAP, when cellularization is complete in barley, transfer cells and the adjacent starchy endosperm cells show a strong *END1* signal (Doan et al. 1996). Abundant *END1* transcripts were also detected in the transfer cells of developing wheat caryopsis at 9 DAP (Drea et al. 2005). Promoter activation of *END1*-like genes from durum wheat and rice (*TdPR60* and *OsPR602*) in ETC of transgenic wheat and rice plants was observed at 7–9 DAP (Li et al. 2008; Kovalchuk et al. 2009). Although the function of *END1*-like genes remains unknown, the expression of these genes does give some hints to the time of fate specification of the transfer cells. Four genes designated as *BETL1-4* (*Basal Endosperm Transfer Layer1-4*) isolated from maize provide molecular evidence of the function of transfer layers. This group of genes encoded small, cysteine-rich proteins with putative signal peptides. They were predominantly expressed in transfer cells during a short period in early- to mid- grain development (Hueros et al. 1995; Hueros et al. 1999). Sequence similarity of these proteins with defensins and proteinase inhibitors suggests that they may have anti-pathogen activities (Hueros et al. 1995; Hueros et al. 1999; Serna et al. 2001). Basal layer antifungal protein1 (Bap1) belongs to another group of potential anti-fungal proteins identified from maize. Using in situ hybridization, mRNA for *Bap1* and *Bap3* was detected exclusively in the developing endosperm at 10–18 DAP. In an immunolocalization experiment, *Bap2* (synonym *Betl2*) was detected in the intracellular matrix of the in-growth transfer cells in the endosperm and the adjacent placento-chalazal cells of the pedicel (Serna et al. 2001). Heterologous expression of *Bap2* peptide in *E. coli* and an in vitro test for fungistatic activity provided evidence for its possible role in antifungal defense during grain development. Unlike BETL1, which is tightly bound to the cell wall, BAP2 is secreted to a high level and is released by transfer cells (Serna et al. 2001).

Mutants of the maize gene *Miniature1* (*mni1*) show an anatomical lesion in the pedicel region and reduced size of the kernel. *Miniature1* encodes a cell wall invertase (INCW2) that is localized in the basal endosperm and pedicel (Cheng et al. 1996; Miller and Chourey 1992). A recent study of the barley cell wall-bound invertase genes revealed a similar expression pattern. *HvCWINV1*and *HvCWINV2* were preferentially expressed in the maternal-basal endosperm boundary from 3 DAP to 6 DAP just before cellularization (Weschke et al. 2003). Interestingly, a fructosyltransferase gene *HvSF6FT1* was also expressed in the inner cell layers of maternal pericarp above the dorsal cells at 3 DAP. The expression of *HvSF6FT1* then moved to the ventral pericarp and to the transfer cells at 4 DAP and is exclusively detected in the transfer layers at 6 DAP (Weschke et al. 2003). These findings indicate that cell wall invertases participate in the process of establishing a sucrose concentration gradient between maternal symplast

and endosperm apoplast by hydrolysis of sucrose to hexose (fructose and glucose) for subsequent import into the endosperm via SF6FT1 in the transfer cells (Cheng et al. 1996; Miller and Chourey 1992; Weschke et al. 2003).

Another mutation *Reduced grain filling* (*rgf1*) causes grain weight loss of up to 70 % in maize (Maitz et al. 2000). Morphologically similar to the *mni1* mutant, *rgf1* is expressed in the endosperm transfer cell layer of *rgf1* mutant. The *Rgf1* gene product causes a decrease in the level of expression of *BETL1* and *BETL2* in transfer cells and maternal placentochalaza. In vitro culture experiments show that high sugar concentrations suppress placentochalaza formation. Starch accumulation (not synthesis) is reduced in *rgf1* kernels. Therefore, *Rgf1* may be involved in sugar sensing or transport in endosperm transfer cells (Maitz et al. 2000).

Transcription factors are also involved in mobilizing photosynthates into developing seed. *ZmMRP-1* was the first transfer cell-specific transcription factor identified in cereals (Gómez et al. 2002). This single-copy gene encodes a nuclear protein with a MYB-related DNA binding (DB) domain and a nuclear localization signal. ZmMRP-1 transcript is detected in the cytoplasmic region of the basal endosperm coenocytes, which gives rise to transfer cells in maize, as early as 3 DAP. High-level expression is restricted to the transfer cell layers from 3 DAP until 16 DAP and peaks at 11 DAP (Gómez et al. 2002). Strikingly, ectopic expression of *ZmMRP-1* under the control of the ubiquitin promoter in *BETL-1*:GUS transgenic maize plants, revealed that ZmMRP-1 is able to transactivate the transfer cell-specific gene expression. However, expression of *ZmMRP-1* occurred prior to the *BETL* gene transactivation. Co-transformation in tobacco protoplasts showed a possible in vitro interaction between ZmMRP-1 and the promoter of *BETL* genes (Gómez et al. 2002). In addition to this finding, the promoter of *maternally expressed gene1* (*Meg1*) was also found to be activated by ZmMRP-1 in a separate study (Costa et al. 2004; Gutiérrez-Marcos et al. 2004) when the transcriptional GUS fusion construct of *MEG1* and a transcriptional 35S:*MRP1* construct were co-transformed into tobacco protoplasts (Gutiérrez-Marcos et al. 2004). It was shown that *MEG1* was expressed in the basal transfer cells from 10 to 20 DAP, peaking at 10–12 DAP (Costa et al. 2004; Gutiérrez-Marcos et al. 2004). Using MEG1-specific antibody in an immunolocalization experiment, MEG1 was found to localize to the cell wall ingrowths of the endosperm transfer layers (Gutiérrez-Marcos et al. 2004).

ZmTCRR-1 and *ZmTCRR-2* encode members of the type-A response regulators of the two-component system (TCS), which is responsible for the phosphotransfer-based signal transduction (Muñiz et al. 2006, 2010). The genes were found to be expressed exclusively in the endosperm transfer-cell layer 8–14 days after pollination, when transfer-cell differentiation is most active. The promoter of *ZmTCRR-1* was strongly transactivated in heterologous systems by the transfer cell-specific transcription factor ZmMRP-1. It was suggested that the possible role of TCCR proteins is to integrate external signals with seed developmental processes (Muñiz et al. 2006, 2010).

Recently the new rice gene, *AL1*, expressed in the ETC-containing region was isolated. The gene encodes a putative anthranilate N-hydroxycinnamoyl/benzoylt

ransferase and is expressed in the dorsal aleurone layer adjacent to the main vascular bundle. In rice, transfer cells are differentiated in this region (Kuwano et al. 2011).

Taken together, the molecular markers identified from maize, wheat, barley and rice help determine when transfer cells identity is defined. A popular hypothesis for transfer cell identity is that two to three layers of cells in the basal endosperm region give rise to transfer cells at the syncytial stage and their subsequent development is subject to elevated signals derived from the adjacent maternal pedicel tissues (reviewed in Olsen 2004). This proposed mechanism differs from the earlier proposed mechanism of aleurone cell fate specification (Becraft and Asuncion-Crabb 2000; Morrison et al. 1975, 1978). Characterization of maize *globby-1* (*glo-1*) mutants in which cell fate acquisition and differentiation of endosperm tissues are affected, suggests that transfer cell identity is defined irreversibly within a narrow window of syncytium development (Costa et al. 2003).

Central Starchy Endosperm

CSE forms the main body of a cereal seed (caryopsis) and represents the major storage site for carbohydrate and protein reserves (reviewed in Becraft 2001; Olsen 2004). Maize endosperm consists of nearly 90 % starch and about 10 % protein (Gibbon and Larkins 2005). The CSE is the factory for carbohydrate and protein biosyntheses and accumulation. In most angiosperm species, starch is the main form of storage carbohydrate and comprises two α-glucan polymers, amylose and amylopectin (reviewed in Lopes and Larkins 1993). After transport into the endosperm, sucrose is converted to glucose-1-phosphate which in turn is incorporated into ADP-glucose for starch synthesis and storage (reviewed in Lopes and Larkins 1993).

Recently, a comprehensive temporal and spatial picture of the deposition and modification of cell wall polysaccharides during barley grain development, from endosperm cellularization at 3 DAP through differentiation to the mature grain at 38 DAP was obtained by using immunolabeling, combined with chemical analyses and transcript profiling (Wilson et al. 2012).

Storage proteins which mainly store nitrogen and sulfur for germinating seed are also synthesized and deposited in the CSE. Almost 70 % of the proteins in the maize endosperm are zeins composed of prolamin proteins (Gibbon and Larkins 2005). Prolamins consist of large amounts of the non-charged amino acids proline and glutamine, resulting in insolubility of these proteins in aqueous solvents. Gliadins and glutenins are two basic types of prolamin protein found in wheat. Other proteins believed to participate in defense, namely α-amylase inhibitors, protease inhibitors, ribosome-inactivating proteins (RIPs), lectins and thionins are also accumulated in CSE (reviewed in Lopes and Larkins 1993).

Although starchy endosperm plays an important role in accumulation of storage products for later use by the developing embryo and germinating seed, the genetic regulators of metabolic pathways, in particular during seed development, are poorly understood. *Opaque 2* (*O2*) is considered to be the first direct

genetic regulator identified for zein synthesis in maize (reviewed in Becraft 2001). Mutations in maize *opaque* genes alter zein synthesis, resulting in reduced number, size and abnormal morphology of protein bodies. *O2* encodes a basic domain/leucine-zipper (bZIP) transcription factor (Schmidt et al. 1987), which predominantly regulates the expression of genes encoding 22-kD α-zeins (Damerval and Guilloux 1998; Habben et al. 1993). *Floury2* (*Fl2*) is another mutant characterized at the molecular level. It carries a point mutation altering the signal peptide of a 22-kD α-zein protein, which leads to a soft texture of the starchy endosperm due to the failure of correct processing of the signal peptide during protein biosynthesis on the rough endoplasmic reticulum (Coleman et al. 1997; Gillikin et al. 1997). Quality protein maize (QPM) was created by selecting genetic modifiers that convert the starchy endosperm of an *o2* mutant to a hard, vitreous phenotype. Genetic analysis identified multiple, unlinked *O2* modifiers (Opm) of unknown origin. Two independently developed QPM lines were used to map several major *Opm* QTLs to chromosomes 1, 7 and 9. A microarray hybridization performed with RNA obtained from true breeding *o2* progeny with vitreous and opaque kernel phenotypes identified a small group of differentially expressed genes, some of which were mapped at or near the *Opm* QTLs. Several of identified genes are associated with ethylene and ABA signaling that might suggest a potential linkage of *o2* endosperm modification with programmed cell death (Holding et al. 2008).

Zein protein synthesis in a number of maize mutant lines, namely W64A *opaque 2*, *DeB30* (*Defective endosperm B30*) and *floury2*, can be reduced by 45–65 % compared of the wild type (Hunter et al. 2002; Scanlon and Meyers 1998). Most *dek* (*defective kernel*) mutants were identified from maize mutated either chemically or by Mu transposition (Olsen 2004). Among these genes, *Emp2* (*Empty pericarp2*) and *Dsc1* (*Discolored 1*) have been isolated (Balasubramanian and Schneitz 2000, 2002; Scanlon and Meyers 1998). The *Dsc1* transcripts accumulate specifically in the developing kernel at 5–7 DAP (Scanlon and Meyers 1998). EMP2 is a putative protein similar to Heat-shock protein1. The *emp2* mutant showed embryo lethality due to abortion of the kernel during early embryogenesis (Balasubramanian and Schneitz 2002). Although studies of *Emp2* and *Dsc1* indicate that they may play important roles in kernel development, the functions of these genes remain unknown at the molecular level.

Two maize mutants, *zmsmu2-1* and *zmsmu2-3* are a result from insertion of a Mutator (Mu) transposable element in the first exon of a gene homologous to the nematode gene, *smu-2*, which is involved in RNA splicing. Besides other phenotypical changes, the mutants had reduced levels of zein storage proteins in a starchy endosperm. The gene encoding ZmSMU2 is expressed in the endosperm, embryo, and shoot apex, and is required for efficient ribosomal RNA processing, ribosome biogenesis, and protein synthesis in developing endosperm. Apart the pleiotropic nature of the mutations, a possible role for ZmSMU2 in the development of maize endosperm has been proposed (Chung et al. 2007). A series of maize seed mutants demonstrate increases in lysine content and decreased amounts of zeins in starchy endosperm. One of them is *opaque7* (*o7*) is a classic maize starchy endosperm mutant was characterized and *O7* gene was cloned by map-based cloning and

gene function was confirmed by functional complementation in transgenic plants and using an RNAi approach. The *O7* gene encodes an acyl-activating enzyme with high similarity to AAE3. Analysis of metabolites suggested that the *O7* gene might affect amino acid biosynthesis by modifying levels of α-ketoglutaric acid and oxaloacetic acid. The cloning of *O7* revealed a novel regulatory mechanism for storage protein synthesis, which may be an effective target for the genetic manipulation of storage protein contents in cereal grains (Wang et al. 2011).

The determinant of CSE cell identity is unclear. Olsen and co-workers proposed that two types of cells are the sources of starchy endosperm (Lid et al. 2004). One source is the periclinal divided inner layer of the daughter cells of aleurone cells (also known as subaleurone cells). The other source is the randomly-oriented cells growing inward into the central vacuole (Lid et al. 2004). Diverged from transfer cells, the CSE and AL may share the same determinant but result from different positional signals set off by the maternal tissues (Becraft and Asuncion-Crabb 2000). The maize *dek1* and *crinkly4* mutants both lack aleurone cells. However, CSE cells are differentiated in place of the aleurone (Becraft and Asuncion-Crabb 2000; Becraft et al. 1996; Lid et al. 2002). Mutants in which aleurone cell development is disrupted also show shrunken CSE and various defects in embryo development; these include *dek1*, *cr4*, *dal1* and *dal2* (Becraft et al. 2002, 1996; Klemsdal et al. 1991; Lid et al. 2002).

Current research on the genetic control of CSE development still focuses on genes encoding enzymes that are involved in the biosynthesis of starch and storage proteins. In addition, the study of genes that are co-expressed in the TL and CSE during endosperm cell differentiation will be of high priority for understanding grain development and the process of sugar and amino acid transport.

Embryo Surrounding Region

The ESR is a specialized endosperm tissue in cereals. Based on its morphological and positional features in the grain, this "shield" is believed to act as a physical barrier, nutritional path, and messenger between the endosperm and embryo (reviewed in Olsen 2004; Kawashima and Goldberg 2010). The ESR has been intensively studied but little is known about this specialized region in small grain cereals. In maize, *Esr1*, *Esr2*, *Esr3* were the first genes identified as being expressed in the ESR (Schel et al. 1984). Expression levels of these genes differ in the ESR from 5 to 9 DAP, and peak at 7 DAP (Bonello et al. 2000, 2002; Opsahl-Ferstad et al. 1997). The *Esr3* gene product belongs to a family of small hydrophilic proteins, which share a conserved motif with Clavata3 (Clv3), the ligand of the receptor-like kinase CLV1 (Bonello et al. 2002). This finding suggests a role for ESR in the signaling pathway between endosperm and embryo during embryogenesis. ESR1, 2 and 3 proteins were subcellularly localized to the cell walls of the ESR. *ZmAE1* (*Zea mays androgenic embryo1*) and *ZmAE3* were found to be preferentially expressed in the ESR between 5 and 20 DAP (Magnard et al. 2000). *ZmEBE-1* and *ZmEBE-2* represent genes of unknown function, with expression

in the central cell before fertilization, and in the resulting TL and the ESR up to 20 DAP with a peak at 7 DAP. Expression of *ZmEBE-1* is at least one tenth that of *ZmEBE-2* and is subject to a maternal effect (Magnard et al. 2003). A defensin gene, *ZmESR-6*, was reported to play a protective role for the ESR (Balandin et al. 2005). ZmESR-6 possesses a central domain and a C-terminal acidic domain, which is present in many characterized defensins and thionins. Its antifungal activity has been shown in vitro. Expression of this gene is localized to the ESR and the pedicel region from 11 to 16 DAP and is strongest in the placentochalazal region of the pedicel adjacent to the ESR (Balandin et al. 2005). The gene encoding another lipid transfer protein, *TdPR61*, was found to be expressed in the endosperm transfer cells, the embryo surrounding region, and in the scutellum and radicals of embryo. This pattern of expression suggests similarity in functions of these grain tissues, all of which have a role in nutrient/lipid and/or signal transfer from maternal tissues to the developing embryo (Kovalchuk et al. 2012b).

Like the transfer layer (TL), the ESR has similar functions for the protection of developing kernels in maize. Cell fate specification of the ESR is postulated to occur in the endosperm coenocyte around the embryo (reviewed in Olsen 2004). The TaGL9 transcription factor, which is produced in the syncitial endosperm around embryo as early as at 3 DAP, may be involved in the regulation of ESR formation (Kovalchuk et al. 2012a). However, cell identity of the specialized endosperm domains may be specified in the central cells prior to fertilization, based on the continuous expression pattern of the *ZmEBE* genes (Magnard et al. 2003). Further study on functions of *ZmEBE* genes may provide insights into the initiation of the TL and ESR.

9.4.3 Maternal Tissues

Although maternal tissues do not contribute significantly to the final grain, some components, such as the nucellus, serve as an important nutrient supply pathway for nurturing the developing embryo at early stages of seed development. Moreover, maternal tissues including nucellus, pericarp and testa form a barrier to protect developing seed from pathogen attack. These tissues can also represent a physical restraint for the growth of the embryo and endosperm.

The nucellus provides nutrients to the syncytial and cellularizing endosperm and degenerates through programed cell death (Domínguez et al. 2001). Proteolytic enzymes are thought to be involved in degradation of the nucellar projection cells (Morrison et al. 1978). These cells are part of nucellar tissues and morphologically highly similar to endosperm transfer cells in developing wheat grains. In cereal seed, there are no symplastic connections between the maternal tissues and the endosperm. In order to facilitate solute transport, both the nucellar projection and the adjacent endosperm epithelial cells in the wheat caryopsis differentiate into nucellar projection transfer cells (NPTC) and endosperm transfer cells (ETC). It was demonstrated that ETC and NPTC developed synchronously and have both

similarities and differences in their morphological features. Wall ingrowths of ETC and NPTC form initially in the first layer nearest to the endosperm cavity, and later in the inner layer further from the endosperm cavity. The mature ETC usually comprises three layers of cells and the mature NPTC comprise four layers. Wall ingrowths of ETC are flange type and wall ingrowths NPTC are reticulate. NPTC are not nutrient-storing cells; in contrast, the first layer of ETC shows aleurone cell features, and the second and third layers of ETC accumulate starch granules and protein bodies (Zheng and Wang 2011). There are clear differences in the regulation of transcription of the ETC-specific *END1*-like genes in wheat/barley and rice. Promoters of both *OsPR602* and *TdPR60* are activated strictly in ETC of transgenic wheat and barley, while in rice grain they are also active in NPTC and some other adjacent maternal tissues (Li et al. 2008; Kovalchuk et al. 2009).

Studies have shown that the degeneration of nucellus occurs early in grain development (Morrison et al. 1978). MADS box family member, MADS29, which is preferentially expressed in the nucellus and the nucellar projection of rice grain, was shown to be responsible for the programmed cell death of the nucellus and nucellar projection at early stages of grain development (Yin and Xue 2012). The seed coat of wheat grain is comprised of nucellar epidermis and testa, which is derived from the inner integument of the ovary and is thought to be the true seed coat. In milling technology, nucellar epidermis and testa form the intermediate layer of wheat bran and are regarded as important for the human diet and animal feed (Antoine et al. 2003). Red pigment in the testa of wheat grains is associated with tolerance to pre-harvest sprouting controlled by *R-1* genes. It is now clear that *Tamyb10* genes encode R2R3-type MYB domain proteins, which control the synthesis of the red pigment, proanthocyanidin. Molecular markers developed for *Tamyb10* genes are used for detecting *R-1* genes (Himi et al. 2011).

The maternal pericarp consists of tube, cross, hypodermal and epidermal cells (Bradbury et al. 1956). Early in grain development, the pericarp supplies nutrients to the young endosperm and embryo. Studies of cell wall composition of the pericarp suggest that cell walls of the pericarp function to protect the developing grains from pathogen invasion (Antoine et al. 2003). Another level of grain protection from pathogens might be provided by defensins, which represent small proteins with anti-fungal properties. After fertilization a group of PRPI defensins are strongly expressed in the outer cell layers of the pericarp and in the main vascular bundle of the grain. Expression of defensins was observed in epidermal layers of developing grain until maturation and drying of outer grain cell layers at about 18–20 DAP (Kovalchuk et al. 2010).

9.5 Modified Grain Composition and Structure

Several technologies are now available for the targeted modification of cereal grain composition to enrich content of certain proteins, starch and micronutrients such as iron, zinc and β-carotene. These include modification through

genetic engineering or through selection of genotypes with elevated level of micronutrients (Nestel et al. 2006; Welch and Graham 2004).

9.5.1 Micronutrients

Zinc and iron are micronutrients which are important for human health. In higher plants, nicotianamine (NA) is involved in the transport of metal cations such as Fe^{2+} and Fe^{3+}. Biosynthesis of NA is catalyzed by NA synthase enzymes in the trimerization of S-adenosylmethionine. It has been shown that over-expression of *OsNAS* genes in conjunction with ferritin can increase iron concentration in rice endosperm by 2–6 fold. Amongst the three *OsNAS* genes, *OsNAS2* has been shown to carry a particularly high potential to biofortify iron and zinc in the rice endosperm and therefore could provide a sustainable yet simple genetic solution to iron and zinc deficiency disorders that affect billions of people worldwide (Johnson et al. 2011).

Plants also contain anti-nutrients such as polyphenolics (tannins) or phytate (IP6), which impairs absorption of iron, zinc and calcium by gut. To reduce the phytate content in grains for food, knocking down key enzymes involved in phytate biosynthesis pathway or over-expression of phytase, which degrades phytate in the target grain tissues provide two approaches (White and Broadley 2005). Developing lines with low concentration of phytate through selection of low phytate varieties represents a non-transgenic approach.

9.5.2 Starch Composition

The ration of amylose to amylopectin also represents an important nutritional trait. High amylose content starch helps prevent several diseases such as colon cancer, diabetes, obesity, osteoporosis and cardiovascular diseases in human. It has been demonstrated that silencing genes encoding starch branching enzymes of class II (SBEIIa) using RNAi in durum wheat under the control of an endosperm-specific promoter resulted in an increase of amylose content (Sestili et al. 2010).

9.5.3 Oil Content

In maize, over-expression of *ZmLEC1* leads to increased oil content by up to 48 %. However, these lines also suffered from negative pleiotropic effects including reduced germination, seedling growth, and grain yield. Over-expression of *ZmWRI1* leads to increased oil content similar to that in over-expression of *ZmLEC1* but without the concomitant undesirable traits (Shen et al. 2010). Plant oil content was also improved by constitutive overexpression of genes encoding

plant non-specific lipid-transfer proteins Puroindoline a and b (PINA and PINB) under maize Ubiquitinin promoter in transgenic corn. Over-expression of *Pin* resulted in a siginificant increase in germ or embryo size without negatively impacting seed size. Germ yield increased 33.8 % while total seed oil content increased by 25.23 %. Seed oil content increases were primarily the result of increased germ size (Zhang et al. 2010).

9.5.4 Cell Walls and Dietary Fibre

Grain structure can also be modified by manipulating cell wall composition in the endosperm. It has been shown that over-expression of the barley *CslF6* gene under the control of an endosperm-specific promoter, results in increases of over 80 % in (1,3;1,4)-β-D-glucan content in grain of transgenic barley. Over-expression of *HvCslF* under the control of 35S promoter leads to an increase of (1,3;1,4)-β-D-glucan content in both leaf and grain. In a recent study, it was shown that there is scope and potential to alter cell wall biomass for animal feed and dietary fiber for human (Burton et al. 2011).

9.5.5 Seed Storage Proteins

Seed storage proteins (SSPs) are synthesized and deposited in storage organelles in the endosperm during seed maturation as a nitrogen source for germinating seedlings. Rice glutelin, globulin, and prolamin knockdown lines showed that reduction of one or a few SSP(s) was compensated for by increases in other SSPs at both the mRNA and protein levels. Reduction of glutelins or sulfur-rich 10-kD prolamin levels was preferentially compensated by sulfur-poor or other sulfur-rich prolamins, respectively, indicating that sulfur-containing amino acids are involved in regulating SSP composition. Furthermore, a reduction in the levels of 13-kD prolamin resulted in enhancement of the total lysine content by 56 % when compared with the wild type. This observation can be mainly accounted for by the increase in lysine-rich proteins. Hence, manipulation with expression of particular groups of SSPs can be used to improve the nutritional quality of grain (Kawakatsu et al. 2010).

9.5.6 Animal Proteins

Cereal grain can also be used for the production of animal proteins, e.g. collagen and gelatin-related proteins with predetermined composition and structure. Full-length collagen type I alpha1 (rCIa1) was successfully expressed in transgenic barley and corn seeds (Eskelin et al. 2009; Zhang et al. 2009).

9.5.7 Allergens

Another area of biotechnological application for improvement of seed quality is sequential deletion/reduction of grain-localized allergens and the development of hypo-allergenic transgenic lines. For example, rice seed proteins are known to be a causative antigen in some patients with food allergy, especially cereal allergy, with clinical symptoms such as eczema and dermatitis. The α-amylase/trypsin inhibitors (14–16 kDa), α-globulin (26 kDa) and β-glyoxalase I (33 kDa) are regarded as major potential allergens of rice seed, based on specific recognition by serum IgE from allergy patients. In order to suppress the production of these major allergens in rice grains, a mutant in the 'Koshihikari' background lacking the 26 kDa allergen (GbN-1) was used as a host for RNA silencing. A binary vector harboring two RNAi gene cassettes for suppression of 14–16 and 33 kDa allergens driven by the 13 and 10 kDa prolamin endosperm-specific promoters, respectively, was introduced into the GbN-1 genome by *Agrobacterium*-mediated transformation. In the most promising transgenic lines, the content of the three potential allergens was remarkably reduced to a very faint level without a change in seed phenotype (Wakasa et al. 2011).

9.5.8 Seed-Specific Promoters—Tools for Grain Biotechnology

Earlier in this section, we outlined several areas where genetic modification of grain composition offers opportunities for increasing the nutritional or processing quality. In most of these applications, grain-specific expression of the transgene is required. For this purpose, a series of cereal grain-specific promoters have been identified and characterized over the past decade. Using reporter genes in transgenic cereals, these promoters have been shown to induce spatial and temporal tissue- and/or cell-specific expression. The main group of grain-specific promoters comprise starchy endosperm-specific promoters; for example, alpha-gliadin promoter induced expression primarily in the cells of the starchy endosperm, subaleurone and aleurone layers from 11 days after anthesis (DAA) until grain maturity (Van Herpen et al. 2008). Promoters of 15 genes including 10 seed storage protein genes and five genes for enzymes involved in carbohydrate and nitrogen metabolism were also tested in transgenic rice. The promoters for the glutelins and the 13 and 16 kDa prolamins directed endosperm-specific expression, especially in the outer portion (peripheral region) of the endosperm, whilst the embryo globulin and 18 kDa oleosin promoters directed expression in the embryo and aleurone layer (Qu and Takaiwa 2004). A characterization of the promoter of a wheat high molecular weight glutenin subunit (HMW subunit) gene, *Glu-1D-b* using transgenic wheat plants revealed that expression of the reporter gene was initiated first in the central lobes of the starchy endosperm, and then spread throughout the endosperm tissue, while no expression was detected in the aleurone layer

(Lamacchia et al. 2001). This group of promoters can be used for improvement of endosperm quality and for expression of animal proteins important for technological and medical purposes.

Another important group of promoters comprise promoters specific for ETC (Fig. 9.2). Two promoters of LTP genes, *OsPR602* and *OsPR9a*, identified from rice have shown similar activity in the endosperm transfer cells during cell differentiation and seed maturation in transgenic rice and barley grains (Li et al. 2008) (Fig. 9.2m–q). The promoter of *TdPR60,* a gene homolologous to *OsPR602*, demonstrated transfer cell-specific activity in transgenic bread wheat and barley, although in transgenic rice it was less specific (Kovalchuk et al. 2009) (Fig. 9.2i). ETC-specific expression was also demonstrated for a putative anthranilate N-hydr oxycinnamoyl/benzoyltransferase (AL1) gene (Kuwano et al. 2011). ETC-specific promoters will be useful for enhancing uptake of nutrients from the maternal tissues and protecting seeds from pathogen attack.

The promoter of another member of the *TdPR60* clade, designated as *TdPR61* showed a more complex pattern of expression (Kovalchuk et al. 2012b). *TdPR61* encodes a non-specific lipid transfer protein involved in lipid transfer to apoplast. The activity of *TdPR61* promoter was detected predominantly in the embryo, embryo surrounding region and the endosperm transfer cells in transgenic wheat, barley and rice (Fig. 9.2a–d). The activity of a very strong embryo-specific promoter, 1 Cys-Prx, was analyzed in transgenic rice (Kim et al. 2011). The reporter gene expression driven by the 1 Cys-Prx promoter was strong in the embryo and aleurone layer. The activity of the 1 Cys-Prx promoter was higher than that of previously-identified embryo-specific promoters, and comparable to that of strong endosperm-specific promoters in rice (Kim et al. 2011). The promoter of *Early methionine* (*Em*) gene from wheat had similar spatial specificity of expression as 1 Cys-Prx when it was tested in transgenic barley (Furtado and Henry 2005). Several strong maize promoters preferably active in embryo were isolated and assessed by transgene expression. One of the most active was the *globulin-1* promoter. Several other highly active embryo-specific promoters were also identified and can be used for the simultaneous expression of multiple foreign proteins in embryo tissues (Streatfield et al. 2010).

The *PRPIs* are promoters of defensin genes and have been identified and isolated from rice and wheat. The activity of *PRPI* has been shown in the ovary shortly before anthesis in wheat and in the outer layer of the pericarp and vascular bundle of developing grains (Kovalchuk et al. 2010). The wheat *PRPI* promoters are also active during seed germination in transgenic rice plants. All PRPIs are strongly inducible by wounding in leaves, stems and grains from transgenic rice plants (Fig. 9.2e–h, r–t). PRPI promoters can be applied in cases when specific targeting and accumulation of proteins conferring resistance to pathogens in vulnerable tissues of developing and germinating grain is required.

The promoter of a HD-Zip IV transcription factor from durum wheat, designated as *TdGL9H1* was activated in embryo and embryo surrounding part of the syncytial endosperm in transgenic wheat and barley. However, after endosperm cellularization the promoter activity was observed exclusively in the scutellar

vascular bundle. Expression was still detectable several weeks after grain harvest, but vanished shortly after imbibition. (Kovalchuk et al. 2012a) (Fig. 9.2j–l). The *TdGL9H1* promoter could be a useful tool for engineering early seedling vigor and protecting the endosperm to embryo axis pathway from pathogens during grain desiccation and storage.

9.6 Summary and Outlook

Cereal grains provide a major source of food for humans and farm animals. Over the past few years, grains have also been widely used as a source of industrial starch and for biofuel production. The demand for cereals is predicted to grow substantially over the next few decades and this has triggered renewed research activity into a range of aspects of cereal production. Modifications in fertilization process and grain characteristics have become important targets for improvement. As a consequence, seed development, particularly endosperm initiation and development in cereals has been intensively studied in maize, barley, wheat and rice. However, our knowledge and understanding of the regulation of grain development is still incomplete and many important objectives, such as inducing apomixis, remain elusive. Progress in modifying grain composition is proving far more encouraging with a wide range of projects and targets and several novel grains now in commercial production.

References

Aalen R, Opsahl-Ferstad H, Linnestad C, Olsen O-A (1994) Transcripts encoding an oleosin and a dormancy-related protein are present in both the aleurone layer and the embryo of developing barley (*Hordeum vulgare* L.) seeds. Plant J 5:385–396

Antoine C, Peyron S, Mabille F, Lapierre C, Bouchet B, Abecassis J, Rouau X (2003) Individual contribution of grain outer layers and their cell wall structure to the mechanical properties of wheat bran. J Agr Food Chem 51:2026–2033

Aoki N, Scofield GN, Wang XD, Offler CE, Patrick JW, Furbank RT (2006) Pathway of sugar transport in germinating wheat seeds. Plant Physiol 141:1255–1263

Aquea F, Poupin MJ, Matus JT, Gebauer M, Medina C, Arce-Johnson P (2008) Synthetic seed production from somatic embryos of Pinus radiata. Biotechnol Lett 30:1847–1852

Ashton AR, Polya GM (1978) Cyclic adenosine 3':5'-monophosphate in axenic rye grass endosperm cell cultures. Plant Physiol 61:718–722

Balandin M, Royo J, Gomez E, Muniz LM, Molina A, Hueros G (2005) A protective role for the embryo surrounding region of the maize endosperm, as evidenced by the characterisation of *ZmESR-6*, a defensin gene specifically expressed in this region. Plant Mol Biol 58:269–282

Balasubramanian S, Schneitz K (2000) NOZZLE regulates proximal-distal pattern formation, cell proliferation and early sporogenesis during ovule development in *Arabidopsis thaliana*. Development 127:4227–4238

Balasubramanian S, Schneitz K (2002) NOZZLE links proximal-distal and adaxial-abaxial pattern formation during ovule development in *Arabidopsis thaliana*. Development 129:4291–4300

Becraft P (2001) Cell fate specification in the cereal endosperm. Semin Cell Dev Biol 12:387–394

Becraft P, Asuncion-Crabb Y (2000) Positional cues specify and maintain aleurone cell fate in maize endosperm development. Development 127:4039–4048

Becraft P, Kang S-H, Suh S-G (2001) The maize CRINKLY4 receptor kinase controls a cell-autonomous differentiation response. Plant Physiol 127:486–496

Becraft P, Li K, Dey N, Asuncion-Crabb Y (2002) The maize dek1 gene functions in embryonic pattern formation and cell fate specification. Development 129:5217–5225

Becraft P, Stinard P, McCarty D (1996) CRINKLY4: A TNFR-like receptor kinase involved in maize epidermal differentiation. Science 273:1406–1409

Becraft PW, Yi G (2011) Regulation of aleurone development in cereal grains. J Exp Bot 62:1669–1675

Bennett MD, Rao MK, Smith JB, Bayliss MW (1973) Cell developmetn in the anther, the ovule, and the young seed of *Triticum aestivum* L. var. Chinese Spring. Phil Trans Royal S B: Biol Sci 266:49–81

Bonello J-F, Opsahl-Ferstad H-G, Perez P, Dumas C, Rogowsky PM (2000) Esr genes show different levels of expression in the same region of maize endosperm. Gene 246:219–227

Bonello J-F, Sevilla-Lecoq S, Berne A, Risueno M-C, Dumas C, Rogowsky PM (2002) Esr proteins are secreted by the cells of the embryo surrounding region. J Exp Bot 53:1559–1568

Bosnes M, Weideman F, Olsen O-A (1992) Endosperm differentiation in barley wild-type and sex mutants. Plant J 2:661–674

Bradbury D, MacMasters MM, Cull IM (1956) Structure of the mature wheat kernel II. Microscopic structure of pericarp, seed coat, and other coverings of endosperm and germ of hard red winter wheat. Cereal Chem 33:342–360

Brandt SP (2005) Microgenomics: gene expression analysis at the tissue-specific and single-cell levels. J Exp Bot 56:495–505

Brink RA, Cooper DC (1947) The endosperm in seed development. Bot Rev 13:423–541

Brown RC, Lemmon BE, Nguyen H, Olsen O-A (1999) Development of endosperm in *Arabidopsis thaliana*. Sex Plant Reprod 12:32–42

Brown RC, Lemmon BE, Olsen O-A (1994) Endosperm development in barley: microtubule involvement in the morphogenetic pathway. Plant Cell 6:1241–1252

Brown RC, Lemmon BE, Olsen O-A (1996a) Development of the endosperm in rice (*Oryza sativa* L.): cellularization. J Plant Res 109:301–313

Brown RC, Lemmon BE, Olsen O-A (1996b) Polarization predicts the pattern of cellularization in cereal endosperm. Protoplasma 192:168–177

Brown RC, Lemmon BE, Stone BA, Olsen O-A (1997) Cell wall (1 → 3)- and (1 → 3, 1 → 4)-β-glucans during early grain development in rice (*Oryza sativa* L.). Planta 202:414–426

Burton RA, Collins HM, Kibble NAJ, Smith JA, Shirley NJ, Jobling SA, Henderson M, Singh RR, Pettolino F, Wilson SM, Bird AR, Topping DL, Bacic A, Fincher GB (2011) Over-expression of specific HvCslF cellulose synthase-like genes in transgenic barley increases the levels of cell wall (1,3;1,4)-β-d-glucans and alters their fine structure. Plant Biotechnol J 9:117–135

Chen G, DenBoer L, Shin J (2008) Design of a single plasmid-based modified yeast one-hybrid system for investigation of in vivo protein–protein and protein-DNA interactions. Biotechniques 45:295–304

Cheng WH, Taliercio EW, Chourey PS (1996) The *miniature1* seed locus of maize encodes a cell wall invertase required for normal development of endosperm and maternal cells in the pedicel. Plant Cell 8:971–983

Chung T, Kim CS, Nguyen HN, Meeley RB, Larkins BA (2007) The maize *zmsmu2* gene encodes a putative RNA-splicing factor that affects protein synthesis and RNA processing during endosperm development. Plant Physiol 144:821–835

Coleman CE, Clore AM, Ranch JP, Higgins R, Lopes MA, Larkins BA (1997) Expression of a mutant alpha-zein creates the *floury2* phenotype in transgenic maize. Proc Natl Acad Sci USA 94:7094–7097

Costa LM, Gutierrez-Marcos JF, Brutnell TP, Greenland AJ, Dickinson HG (2003) The *globby1-1* (*glo1-1*) mutation disrupts nuclear and cell division in the developing maize seed causing alterations in endosperm cell fate and tissue differentiation. Development 130:5009–5017

Costa LM, Gutierrez-Marcos JF, Dickinson HG (2004) More than a yolk: the short life and complex times of the plant endosperm. Trends Plant Sci 9:507

Damerval C, Guilloux ML (1998) Characterization of novel proteins affected by the *o2* mutation and expressed during maize endosperm development. Mol Gen Genet 257:354–361

Deplancke B, Dupuy D, Vidal M, Walhout AJ (2004) A gateway-compatible yeast one-hybrid system. Genome Res 14:2093–2101

Doan DNP, Linnestad C, Olsen O-A (1996) Isolation of molecular markers from the barley endosperm coenocyte and the surrounding nucellus cell layers. Plant Mol Biol 31:877–886

Domínguez F, Moreno J, Cejudo FJ (2001) The nucellus degenerates by a process of programmed cell death during the early stages of wheat grain development. Planta 213:352–360

Drea S, Leader DJ, Arnold BC, Shaw P, Dolan L, Doonan JH (2005) Systematic spatial analysis of gene expression during wheat caryopsis development. Plant Cell 17:2172–2185

Drews GN, Yadegari R (2002) Development and function of the angiospern female gametophyte. Ann Rev Genet 36:99–124

Eskelin K, Ritala A, Suntio T, Blumer S, Holkeri H, Wahlström EH, Baez J, Mäkinen K, Maria NA (2009) Production of a recombinant full-length collagen type I alpha-1 and of a 45-kDa collagen type I alpha-1 fragment in barley seeds. Plant Biotechnol J 7:657–672

Fincher GB (1989) Molecular and cellular biology associated with endosperm mobilization in germinating cereal grains. Ann Rev Plant Physiol Plant Mol Biol 40:305–346

Fischer G, Hizsnyik E, Prieler S, Shah M, Van Velthuizen H (2009) Biofuels and food security: implications of an accelerated biofuels production. Intl Inst Applied Systems Analysis, OFID Pamphlet Series 38

Furtado A, Henry RJ (2005) The wheat *Em* promoter drives reporter gene expression in embryo and aleurone tissue of transgenic barley and rice. Plant Biotechnol J 3:421–434

Galbraith DW, Birnbaum K (2006) Global studies of cell type-specific gene expression in plants. Annu Rev Plant Biol 57:451–475

Gibbon BC, Larkins BA (2005) Molecular genetic approaches to developing quality protein maize. Trends Genet 21:227

Gillikin JW, Zhang F, Coleman CE, Bass HW, Larkins BA, Boston RS (1997) A defective signal peptide tethers the floury-2 zein to the endoplasmic reticulum membrane. Plant Physiol 114:345–352

Gómez E, Royo J, Guo Y, Thompson R, Hueros G (2002) Establishment of cereal endosperm expression domains: identification and properties of a maize transfer cell-specific transcription factor, ZmMRP-1. Plant Cell 14:599–610

Gubatz S, Dercksen VJ, Brüss C, Weschke W, Wobus U (2007) Analysis of barley (*Hordeum vulgare*) grain development using three-dimensional digital models. Plant J 52:779–790

Gutiérrez-Marcos JF, Costa LM, Biderre-Petit C, Khbaya B, O'Sullivan DM, Wormald M, Perez P, Dickinson HG (2004) *Maternally expressed gene1* is a novel maize endosperm transfer cell-specific gene with a maternal parent-of-origin pattern of expression. Plant Cell 16:1288–1301

Habben JE, Kirleis AW, Larkins BA (1993) The origin of lysine-containing proteins in opaque-2 maize endosperm. Plant Mol Biol 23:825–838

Haslam TM, Yeung EC (2011) Zygotic embryo culture: an overview plant embryo culture. In: Thorpe TA, Yeung EC (eds). Humana Press, pp 3-15

Hens K, Feuz J-D, Deplancke B (2012) A High-throughput gateway-compatible yeast one-hybrid screen to detect Protein–DNA interactions. In: Deplancke B, Gheldof N (eds) Methods in molecular biology. Gene regulatory networks. Methods and protocols. Humana Press, vol 786, pp 335–355

Himi E, Maekawa M, Miura H, Noda K (2011) Development of PCR markers for *Tamyb10* related to *R-1* red grain color gene in wheat. Theor Appl Genet 122:1561–1576

Holding DR, Hunter BG, Chung T, Gibbon BC, Ford CF, Bharti AK, Messing J, Hamaker BR, Larkins BA (2008) Genetic analysis of *opaque2* modifier loci in quality protein maize. Theor Appl Genet 117:157–170

Hu S (1982) Embryology of angiosperms. People's Education Press, Beijing

Hueros G, Gomez E, Cheikh N, Edwards J, Weldon M, Salamini F, Thompson RD (1999) Identification of a promoter sequence from the *BETL1* gene cluster able to confer transfer-cell-specific expression in transgenic maize. Plant Physiol 121:1143–1152

Hueros G, Varotto S, Salamini F, Thompson RD (1995) Molecular characterization of BET1, a gene expressed in the endosperm transfer cells of maize. Plant Cell 7:747–757

Hunter B, Beatty M, Singletary G, Hamaker B, Dilkes B, Larkins B, Jung R (2002) Maize opaque endosperm mutations create extensive changes in patterns of gene expression. Plant Cell 14:2591–2612

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

Kawakatsu T, Hirose S, Yasuda H, Takaiwa F (2010) Reducing rice seed storage protein accumulation leads to changes in nutrient quality and storage organelle formation. Plant Physiol 154:1842–1854

Kawashima T, Goldberg RB (2010) The suspensor: not just suspending the embryo. Trends Plant Sci 15:23–30

Kessler S, Seiki S, Sinha N (2002) Xcl1 causes delayed oblique periclinal cell divisions in developing maize leaves, leading to cellular differentiation by lineage instead of position. Development 129:1859–1869

Kim JH, Jung IJ, Kim DY, Fanata WI, Son BH, Yoo JY, Harmoko R, Ko KS, Moon JC, Jang HH, Kim WY, Kim JY, Lim CO, Lee SY, Lee KO (2011) Proteomic identification of an embryo-specific 1Cys-Prx promoter and analysis of its activity in transgenic rice. Biochem Biophys Res Commun 408:78–83

Kim Y, Kim S, Park J, Park H, Lim M, Chua N, Park C (2006) A membrane-bound NAC transcription factor regulates cell division in *Arabidopsis*. Plant Cell 18:3132–3144

Klein P, Dietz KJ (2010) Identification of DNA-binding proteins and protein–protein interactions by yeast one-hybrid and yeast two-hybrid screen. Methods Mol Biol 639:171–192

Klemsdal S, Hughes W, Lonneborg A, Aalen R, Olsen O (1991) Primary structure of a novel barley gene differentially expressed in immature aleurone layers. Mol Gen Genet 228:9–16

Koltunow AM (1993) Apomixis: embryo sacs and embryos formed without meiosis or fertilization in ovules. Plant Cell 5:1425–1437

Kovalchuk N, Li M, Wittek F, Reid N, Singh R, Shirley N, Ismagul A, Eliby S, Johnson A, Milligan AS, Hrmova M, Langridge P, Lopato S (2010) Defensin promoters as potential tools for engineering disease resistance in cereal grains. Plant Biotechnol J 8:47–64

Kovalchuk N, Smith J, Bazanova N, Pyvovarenko T, Singh R, Shirley N, Ismagul A, Johnson A, Milligan AS, Hrmova M, Langridge P, Lopato S (2012a) Characterization of the wheat gene encoding a grain-specific lipid transfer protein TdPR61, and promoter activity in wheat, barley and rice. J Exp Bot 63:2025–2040

Kovalchuk N, Smith J, Pallotta M, Singh R, Ismagul A, Eliby S, Bazanova N, Milligan A, Hrmova M, Langridge P, Lopato S (2009) Characterization of the wheat endosperm transfer cell-specific protein TaPR60. Plant Mol Biol 71:81–98

Kovalchuk N, Wu W, Eini O, Bazanova N, Pallotta M, Shirley N, Singh R, Ismagul A, Eliby S, Johnson A, Langridge P, Lopato S (2012b) The scutellar vascular bundle–specific promoter of the wheat HD-Zip IV transcription factor shows similar spatial and temporal activity in transgenic wheat, barley and rice. Plant Biotechnol J 10:43–53

Kranz E, Lörz H (1994) In vitro fertilisation of maize by single egg and sperm cell protoplast fusion mediated by high calcium and high pH. Zygote 2:125–128

Kranz E, Scholten S (2008) *In vitro* fertilization: analysis of early post-fertilization development using cytological and molecular techniques. Sexual Plant Reprod 21:67–77

Kranz E, von Wiegen P, Quader H, Lorz H (1998) Endosperm development after fusion of isolated, single maize sperm and central cells in vitro. Plant Cell 10:511–524

Kuwano M, Masumura T, Yoshida KT (2011) A novel endosperm transfer cell-containing region-specific gene and its promoter in rice. Plant Mol Biol 76:47–56

Lamacchia C, Shewry PR, Di Fonzo N, Forsyth JL, Harris N, Lazzeri PA, Napier JA, Halford NG, Barcelo P (2001) Endosperm-specific activity of a storage protein gene promoter in transgenic wheat seed. J Exp Bot 52:243–250

Leah R, Skriver K, Knudsen S, Ruud-Hansen J, Raikhel NV, Mundy J (1994) Identification of an enhancer/silencer sequence directing the aleurone-specific expression of a barley chitinase gene. Plant J 6:579–589

Li M, Singh R, Bazanova N, Milligan AS, Shirley N, Langridge P, Lopato S (2008) Spatial and temporal expression of endosperm transfer cell-specific promoters in transgenic rice and barley. Plant Biotechnol J 6:465–476

Lid SE, Al RH, Krekling T, Meeley RB, Ranch J, Opsahl-Ferstad H-G, Olsen O-A (2004) The maize *disorganized aleurone layer* 1 and 2 (*dil1*, *dil2*) mutants lack control of the mitotic division plane in the aleurone layer of developing endosperm. Planta 218:370–378

Lid SE, Gruis D, Jung R, Lorentzen JA, Ananiev E, Chamberlin M, Niu X, Meeley R, Nichols S, Olsen O-A (2002) The *defective kernel 1* (*dek1*) gene required for aleurone cell development in the endosperm of maize grains encodes a membrane protein of the calpain gene superfamily. Proc Natl Acad Sci USA 99:5460–5465

Linnestad C, Doan DNP, Brown RC, Lemmon BE, Meyer DJ, Jung R, Olsen O-A (1998) Nucellain, a barley homolog of the dicot vacuolar-processing protease, is localized in nucellar cell walls. Plant Physiol 118:1169–1180

Locatelli F, Manzocchi LA, Viotti A, Genga A (2001) The nitrogen-induced recovery of α-zein gene expression in in vitro cultured *opaque2* maize endosperms depends on the genetic background. Physiol Plant 112:414–420

Lopato S, Borisjuk L, Milligan A, Shirley N, Bazanova N, Langridge P (2006) Systematic identification of factors involved in post-transcriptional processes in wheat grain. Plant Mol Biol 62:637–653

Lopes MA, Larkins BA (1993) Endosperm origin, development, and function. Plant Cell 5:1383–1399

Madrid SM (1991) The barley lipid transfer protein is targeted into the lumen of the endoplasmic reticulum. Plant Physol Biochem 29:695–704

Magnard J-L, Le Deunff E, Domenech J, Rogowsky PM, Testillano PS, Rougier M, Risueño MC, Vergne P, Dumas C (2000) Genes normally expressed in the endosperm are expressed at early stages of microspore embryogenesis in maize. Plant Mol Biol 44:559–574

Magnard J-L, Lehouque G, Massonneau A, Frangne N, Heckel T, Gutierrez-Marcos JF, Perez P, Dumas C, Rogowsky PM (2003) ZmEBE genes show a novel, continuous expression pattern in the central cell before fertilization and in specific domains of the resulting endosperm after fertilization. Plant Mol Biol 53:821–836

Maitz M, Santandrea G, Zhang Z, Lal S, Hannah LC, Salamini F, Thompson RD (2000) rgf1, a mutation reducing grain filling in maize through effects on basal endosperm and pedicel development. Plant J 23:29–42

Mares DJ, Norstog K, Stone BA (1975) Early stages in the development of wheat endosperm. I The change from free nuclear to cellular endosperm. Aust J Bot 23:311–326

Mares DJ, Stone BA, Jeffery C, Norstog K (1977) Early stages in the development of wheat endosperm. II Ultrastructural observations on cell wall formation. Aust J Bot 25:599–613

Miller ME, Chourey PS (1992) The Maize Invertase-Deficient miniature-1 seed mutation is associated with aberrant pedicel and endosperm development. Plant Cell 4:297–305

Moco S, Schneider B, Vervoort J (2009) Plant micrometabolomics: the analysis of endogenous metabolites present in a plant cell or tissue. J Proteome Res 8:1694–1703

Moriguchi K, Suzuki T, Ito Y, Yamazaki Y, Niwa Y, Kurata N (2005) Functional isolation of novel nuclear proteins showing a variety of subnuclear localizations. Plant Cell 17:389–403

Morrison IN, Kuo J, O'Brien TP (1975) Histochemistry and fine structure of developing wheat aleurone cells. Planta 123:105–116

Morrison IN, O'Brien TP, Kuo J (1978) Initital cellularization and differentiation of the aleurone cells in the ventral region of the developing wheat grain. Planta 140:19–30
Muñiz LM, Royo J, Gómez E, Barrero C, Bergareche D, Hueros G (2006) The maize transfer cell-specific type-A response regulator ZmTCRR-1 appears to be involved in intercellular signalling. Plant J 48:17–27
Muñiz LM, Royo J, Gómez E, Baudot G, Paul W, Hueros G (2010) Atypical response regulators expressed in the maize endosperm transfer cells link canonical two component systems and seed biology. BMC Plant Biol 10:84–100
Natesh S, Rau MA (1984) The Embryo. In: Johri BM (ed) Embryology of angiosperms. Springer, Berlin, pp 377–434
Nelson T, Tausta SL, Gandotra N, Liu T (2006) Laser microdissection of plant tissue: what you see is what you get. Annu Rev Plant Biol 57:181–201
Nestel P, Bouis HE, Meenakshi JV, Pfeiffer W (2006) Biofortification of staple food crops. J Nutr 136:1064–1067
Ohtsu K, Takahashi H, Schnable PS, Nakazono M (2007) Cell type-specific gene expression profiling in plants by using a combination of laser microdissection and high-throughput technologies. Plant Cell Physiol 48:3–7
Olsen LT, Divon HH, Kjetil Fosnes RA, Lid SE, Opsahl-Sorteberg H-G (2008) The defective seed5 (des5) mutant: effects on barley seed development and *HvDek1*, *HvCr4*, and *HvSal1* gene regulation. J Exp Bot 59:3753–3765
Olsen O-A (2001) Endosperm development: cellularization and cell fate specification. Ann Rev Plant Physiol Plant Mol Biol 52:233–267
Olsen O-A (2004) Nuclear endosperm development in cereals and *Arabidopsis thaliana*. Plant Cell 16:S214–S227
Opsahl-Ferstad H-G, Deunff EL, Dumas C, Rogowsky PM (1997) ZmEsr, a novel endosperm-specific gene expressed in a restricted region around the maize embryo. Plant J 12:235–246
Opsahl-Sorteberg H-G, Divon HH, Nielsen PS, Kalla R, Hammond-Kosack M, Shimamoto K, Kohli A (2004) Identification of a 49-bp fragment of the HvLTP2 promoter directing aleurone cell specific expression. Gene 341:49–58
Otegui MS, Mastronarde DN, Kang B-H, Bednarek SY, Staehelin LA (2001) Three-dimensional analysis of syncytial-type cell plates during endosperm cellularization visualized by high resolution electron tomography. Plant Cell 13:2033–2051
Otegui MS, Staehelin LA (2000) Syncytial-type cell plates: a novel kind of cell plate involved in endosperm cellularization of Arabidopsis. Plant Cell 12:933–947
Pyvovarenko T, Lopato S (2011) Isolation of plant transcription factors using a yeast one-hybrid system. Methods Mol Biol 754:45–66
le Qu Q, Takaiwa F (2004) Evaluation of tissue specificity and expression strength of rice seed component gene promoters in transgenic rice. Plant Biotechnol J 2:113–125
Raghavan V (2003) One hundred years of zygotic embryo culture investigations. In Vitro Cellular and Developmental Biology—Plant 39:437–442
Raghavan V (2006) Double fertilization: embryo and endosperm development in flowering plants. Springer, Berlin
Reece-Hoyes JS, Diallo A, Lajoie B, Kent A, Shrestha S, Kadreppa S, Pesyna C, Dekker J, Myers CL, Walhout AJ (2011) Enhanced yeast one-hybrid assays for high-throughput gene-centered regulatory network mapping. Nat Methods 8:1059–1064
Reiser L, Fischer R (1993) The ovule and the embryo sac. Plant Cell 5:1291–1301
Russell S (1993) The egg cell: development and role in fertilization and early embryogenesis. Plant Cell 5:1349–1359
Scanlon MJ, Meyers AM (1998) Phenotypic analysis and molecular cloning of discolored-1 (dsc1), a maize gene required for early kernel development. Plant Mol Biol 37:483–493
Schel JHN, Kieft H, Van Lammeren AAM (1984) Interactions between embryo and endosperm during early developmental stages of maize caryopses (*Zea mays*). Can J Bot 62:483–493

Schmidt RJ, Burr FA, Burr B (1987) Transposon tagging and molecular analysis of the maize regulatory locus *opaque-2*. Science 238:960–963

Serna A, Maitz M, O'Connell T, Santandrea G, Thevissen K, Tienens K, Hueros G, Faleri C, Cai G, Lottspeich F, Thompson RD (2001) Maize endosperm secretes a novel antifungal protein into adjacent maternal tissue. Plant J 25:687–698

Sestili F, Janni M, Doherty A, Botticella E, D'Ovidio R, Masci S, Jones HD, Lafiandra D (2010) Increasing the amylose content of durum wheat through silencing of the SBEIIa genes. BMC Plant Biol 10:144–155

Shannon J (1982) Maize endosperm cultures. In: Sheridan W (ed) Maize for Biological Research. Plant Molecular Biology Association, Charlottesville, pp 397–400

Shen B, Allen WB, Zheng P, Li C, Glassman K, Ranch J, Nubel D, Tarczynski MC (2010) Expression of *ZmLEC1* and *ZmWRI1* increases seed oil production in maize. Plant Physiol 153:980–987

Shen B, Li C, Min Z, Meeley RB, Tarczynski MC, Olsen O-A (2003) Sal1 determines the number of aleurone cell layers in maize endosperm and encodes a class E vacuolar sorting protein. Proc Natl Acad Sci USA 100:6552–6557

Skriver K, Leah R, Muller-Uri F, Olsen F, Mundy J (1992) Structure and expression of the barley lipid transfer protein gene Ltp1. Plant Mol Biol 18:585–589

Smith LM, Handley J, Li Y, Martin H, Donovan L, Bowles DJ (1992) Temporal and spatial regulation of a novel gene in barley embryos. Plant Mol Biol 20:255

Sreenivasulu N, Borisjuk L, Junker BH, Mock HP, Rolletschek H, Seiffert U, Weschke W, Wobus U (2010) Barley grain development toward an integrative view. Int Rev Cell Mol Biol 281:49–89

Sreenivasulu N, Usadel B, Winter A, Radchuk V, Scholz U, Stein N, Weschke W, Strickert M, Close TJ, Stitt M, Graner A, Wobus U (2008) Barley grain maturation and germination: metabolic pathway and regulatory network commonalities and differences highlighted by new MapMan/PageMan profiling tools. Plant Physiol 146:1738–1758

Stacy RAP, Nordeng TW, Culianez-Macia FA, Aalen RB (1999) The dormancy-related peroxiredoxin anti-oxidant, PER1, is localized to the nucleus of barley embryo and aleurone cells. Plant J 19:1–8

Streatfield SJ, Bray J, Love RT, Horn ME, Lane JR, Drees CF, Egelkrout EM, Howard JA (2010) Identification of maize embryo-preferred promoters suitable for high-level heterologous protein production. GM Crops 1:162–172

Tester M, Langridge P (2010) Breeding technologies to increase crop production in a changing world. Science 327:818–822

Thiel J, Weier D, Sreenivasulu N, Strickert M, Weichert N, Melzer M, Czauderna T, Wobus U, Weber H, Weschke W (2008) Different hormonal regulation of cellular differentiation and function in nucellar projection and endosperm transfer cells: a microdissection-based transcriptome study of young barley grains. Plant Physiol 148:1436–1452

Thiel J, Weier D, Weschke W (2011) Laser-capture microdissection of developing barley seeds and cDNA array analysis of selected tissues. Methods Mol Biol 755:461–475

Thobunluepop P, Pawelzik E, Vearasilp S (2009) Possibility of sweet corn synthetic seed production. Pak J Biol Sci 12:1085–1089

Tian Q, Olsen L, Sun B, Lid SE, Brown RC, Lemmon BE, Fosnes K, Gruis DF, Opsahl-Sorteberg H-G, Otegui MS, Olsen O-A (2007) Subcellular localization and functional domain studies of *Defective Kernerl* 1 in maize and *Arabidopsis thaliana* suggests a model for aleurone cell fate specification involving CRINKLY 4 and *Supernumerary Aleurone Layer* 1, Plant Cell 19:3127–3145

Tingay S, McElroy D, Kalla R, Fieg S, Wang M, Thornton S, Brettell R (1997) *Agrobacterium tumefaciens*-mediated barley transformation. Plant J 11:1369–1376

Tucker MR, Araujo A-CG, Paech NA, Hecht V, Schmidt EDL, Rossell J-B, de Vries SC, Koltunow AMG (2003) Sexual and apomictic reproduction in Hieracium subgenus Pilosella are closely interrelated developmental pathways. Plant Cell 15:1524–1537

Van Herpen TW, Riley M, Sparks C, Jones HD, Gritsch C, Dekking EH, Hamer RJ, Bosch D, Salentijn EM, Smulders MJ, Shewry PR, Gilissen LJ (2008) Detailed analysis of the expression of an alpha-gliadin promoter and the deposition of alpha-gliadin protein during wheat grain development. Ann Bot 102:331–342

van Went JL, Willemse MTM (1984) Fertilization. In: Johri BM (ed) Embryology of angiosperms. Springer, Berlin, pp 273–318

Vijayaraghavan MR, Prabhakar K (1984) The endosperm. In: Johri BM (ed) Embryology of angiosperms. Springer, Berlin, pp 319–338

Wakasa Y, Hirano K, Urisu A, Matsuda T, Takaiwa F (2011) Generation of transgenic rice lines with reduced contents of multiple potential allergens using a null mutant in combination with an RNA silencing method. Plant Cell Physiol 52:2190–2199

Wang G, Sun X, Wang G, Wang F, Gao Q, Sun X, Tang Y, Chang C, Lai J, Zhu L, Xu Z, Song R (2011) *Opaque7* encodes an acyl-activating enzyme-like protein that affects storage protein synthesis in maize endosperm. Genetics 189:1281–1295

Webb MC, Gunning BES (1991) The microtubular cytoskeleton during development of the zygote, proembryo and free-nuclear endosperm in *Arabidopsis thaliana* (L.) Heynh. Planta 184:187–195

Welch RM, Graham RD (2004) Breeding for micronutrients in staple food crops from a human nutrition perspective. J Exp Bot 55:353–364

Weschke W, Panitz R, Gubatz S, Wang Q, Radchuk R, Weber H, Wobus U (2003) The role of invertases and hexose transporters in controlling sugar ratios in maternal and filial tissues of barley caryopses during early development. Plant J 33:395–411

Weterings K, Russell SD (2004) Experimental analysis of the fertilization process. Plant Cell 16(Suppl):S107–S118

White PJ, Broadley MR (2005) Biofortifying crops with essential mineral elements. Trends Plant Sci 10:586–593

Wilson SM, Burton RA, Collins HM, Doblin MS, Pettolino FA, Shirley N, Fincher GB, Bacic A (2012) Pattern of deposition of cell wall polysaccharides and transcript abundance of related cell wall synthesis genes during differentiation in barley endosperm. Plant Physiol 159:655–670

Ye R, Yao QH, Xu ZH, Xue HW (2004) Development of an efficient method for the isolation of factors involved in gene transcription during rice embryo development. Plant J 38:348–357

Yi G, Lauter AM, Scott MP, Becraft PW (2011) The thick aleurone1 mutant defines a negative regulation of maize aleurone cell fate that functions downstream of defective kernel1. Plant Physiol 156:1826–1836

Yin LL, Xue HW (2012) The MADS29 transcription factor regulates the degradation of the nucellus and the nucellar projection during rice seed development. Plant Cell 24:1049–1065

Zhang C, Baez J, Pappu KM, Glatz CE (2009) Purification and characterization of a transgenic corn grain-derived recombinant collagen type I alpha 1. Biotechnol Prog 25:1660–1668

Zhang J, Martin JM, Beecher B, Lu C, Hannah LC, Wall ML, Altosaar I, Giroux MJ (2010) The ectopic expression of the wheat Puroindoline genes increase germ size and seed oil content in transgenic corn. Plant Mol Biol 74:353–365

Zheng Y, Wang Z (2011) Contrast observation and investigation of wheat endosperm transfer cells and nucellar projection transfer cells. Plant Cell Rep 30:1281–1288

Chapter 10
Genomics of Cereal-Based Functional Foods

Nidhi Rawat, Barbara Laddomada and Bikram S. Gill

10.1 Introduction

Functional foods have been defined by the Food and Nutrition Board (FNB) of the National Academy of Sciences, USA as "any modified food or food ingredient that may provide a health benefit beyond that of the traditional nutrients it contains". Japan was the first country to promote the concept of functional foods as Food for Specific Health Use (FOSHU) endorsed by the Japanese Ministry of Health (Arai 1996). With the increase in public awareness about nutrition and health, functional foods or "foods with a purpose" have gained increased popularity (Verbeke et al. 2009).

Fruits, nuts, berries and vegetables are the most widely known sources of bioactive compounds, whereas cereals, with an annual consumption of 332 kg/person (estimation for 2015, FAO Corporate Documentary Repository, http://www.fao.org/docrep/005/Y4252E/y4252e05.htm, accessed on May 16, 2012), have often been marginalized as functional foods. Recent findings about the health

N. Rawat · B. Laddomada · B. S. Gill (✉)
Wheat Genetic and Genomic Resources Center and Department of Plant Pathology, Throckmorton Plant Sciences Center, Kansas State University, Manhattan, KS 66506-5502, USA
e-mail: bsgill@ksu.edu

N. Rawat
e-mail: nidhirwt@ksu.edu

B. Laddomada
e-mail: barbara.laddomada@ispa.cnr.it

B. Laddomada
Istituto di Scienze delle Produzioni Alimentari, Consiglio Nazionale delle Ricerche (ISPA-CNR), Via Monteroni, 73100 Lecce, Italy

B. S. Gill
Faculty of Science, Genomics and Biotechnology Section, Department of Biological Sciences, King Abdulaziz University, Jeddah 21589, Saudi Arabia

P. K. Gupta and R. K. Varshney (eds.), *Cereal Genomics II*,
DOI: 10.1007/978-94-007-6401-9_10,

benefits of whole-grain cereals and cereal products (Behall et al. 2006; Fardet et al. 2008; He et al. 2010) have renewed interest in the potential of cereals as functional foods. Whole grain cereals have greater nutritional value than the refined or polished cereals, because the bran and germ portion have high fiber content and the majority of bioactive compounds (Champ 2008; Fardet et al. 2008; He et al. 2010). Cereal-based foods have functional food properties due to their carbohydrate constituents (β-glucans, arabinoxylans, inulin), and bioactive compounds such as phenolics (flavones, chalcones, alkylresorcinols, ferulic acid, anthocyanins), carotenoids (β-carotene, xanthophylls), and vitamin E. The physical location of functional components present in the various parts of grains of common cereals is summarized in Table 10.1.

Cereals can be a good source of both probiotic and prebiotic foods because of their diverse carbohydrate composition (Charalampoulos et al. 2002). Probiotic foods contain microorganisms that benefit the consumer's health by improving their intestinal microbial balance (Fuller 1989). Prebiotic food on the other hand, is not digested in the upper gastrointestinal tract but beneficially affects the host health by selectively stimulating the growth and/or activity of useful bacteria in the colon (Gibson and Roberfroid 1995). Table 10.2 summarizes the content of carbohydrate-based functional components in common cereals. Bioactive compounds present in cereals, like phenolic acids, flavonoids, carotenoids, and tocopherols have useful antioxidant properties. They help reduce oxidative stress in the cells and quench the damaging free-radicals, thereby protecting cells from ageing, degeneration, and carcinogenesis (Astorg 1997; He et al. 2010). In fact, bioactive compounds like tocotrienols even reduce the bad cholesterol levels in blood thereby playing protective role against cardiovascular diseases (Das et al. 2008).

The genome sequencing and gene annotation of cereals such as rice (http://rice.plantbiology.msu.edu/), maize (http://magi.plantgenomics.iastate.edu/), barley (International Barley Sequencing Consortium, 2012) and sorghum (Paterson et al. 2009), and the ongoing genome sequencing of wheat (http://www.wheatgenome.org/), have provided a wealth of information about the genes related to bioactive components. Genetic variation studies indicated high heritability for arabinoxylan fiber, carotenoids, and other bioactive compounds, but, significant genotype x environment interactions make it difficult to identify breeding lines with consistently high bioactive compounds across environments and years (Shewry et al. 2010). Here we summarize recent advances in the genomics of various functional food components of common cereals based on their carbohydrate components (Class I) and bioactive components (Class II).

10.2 Carbohydrate-Based Functional Food Components (Class I)

10.2.1 Beta Glucans

The cell walls of grasses are characterized by the presence of (1,3: 1,4)-β-D-glucans composed of unsubstituted, unbranched polysaccharide containing

Table 10.1 Tissue distribution of major functional food components in different cereal grains. Modified from Champ (2008)

Part of grain	Food component	Cereals	Functional properties	References
Pericarp and testa	Arabinoxylans	Wheat, barley, rice, rye	Increase fecal biomass, improves gut health and lipid metabolism	Glei et al. (2006); Neyrinck et al. (2011)
	Phenolic acids	All	Improve redox state, anticancer	Rhodes and Price (1997)
	Flavonoids	All	Antioxidant, anticancer, antiaging, hepatoprotective	Hertog et al. (1993); Middleton et al. (2000)
Aleurone	Minerals	All	Mg maintains cardiac health, fights muscle fatigue; Fe, Zn and Cu maintain proper blood circulation, growth, development and other bodily functions; Ca is essential for good bone health	Cox et al. (1991); Prasad (1998); Haas and Brownlie (2001); Institute of medicine, food and nutrition board (2001)
	Arabinoxylans	As described in pericarp and testa	Described above	As described in pericarp and testa
	Inulin	Wheat, barley, rye	Prebiotic effect, improves gut health, slows glycemic response	Schneeman (1999)
Endosperm	β-glucans	Oat, barley, rye	Reduce glycemic index, Prebiotic effect	Keegstra and Walton (2006); Pennisi (2009)
	Resistant starch	All	Slows glycemic response, prebiotic effect	Topping and Clifton (2001)
Germ	Lipids, unsaturated fatty acids	All	May decrease oxidative stress in patients with hypercholesteromia	Simopoulos (1991)
	Vitamins : E, B6, Folate, Carotenoids	All	Reduce oxidative stress, protect against macular degeneration due to ageing, anticancer	Kohlmeier and Hastings (1995); Astorg (1997); Sen et al. (2006)
	Minerals	All	As described in aleurone	As described in aleurone

Table 10.2 Carbohydrate based functional food components (percent dry grain weight) of common cereals

Cereal	β-glucan (%)	Arabinoxylans (%)	Amylose (%)	Fructan (%)	References
Wheat	0.4–1.4	5.8	23–28	0.9–1.8	Henry (1987); Shelton and Lee (2000); Fretzdorff and Welge (2003); Izydorczyk and Dexter (2008)
Rice	0.4	2.6	16–33	–	Shelton and Lee (2000); Dermibas (2005); Izydorczyk and Dexter (2008)
Maize	0.5	3.5–6.1	24	0.1	Caimi et al. (1996); Shelton and Lee (2000); Dermibas (2005); Izydorczyk and Dexter (2008)
Barley	2.5–11.3	3.50–6.05	22–26	0.6–1	Shelton and Lee (2000); Jones (2007); Izydorczyk and Dexter (2008)
Oat	2.2–7.8	2.7–3.5	16–27	–	Shelton and Lee (2000); Izydorczyk and Dexter (2008)
Rye	1.2–2	7.6–12	24–31	0.6–1	Shelton and Lee (2000); Jones (2007); Izydorczyk and Dexter (2008)
Millets	0.5	–	17	–	Shelton and Lee (2000); Dermibas (2005)

β-D-glucopyranosoyl monomers linked through C(O)3 and C(O)4 atoms with the (1,4)-linkage being more abundant (Burton and Fincher 2009; Burton et al. 2010). Generally, the degree of polymerization (DP) of (1,3: 1,4)-β-D-glucans may vary up to 1,000-fold or more in most grasses (Fincher 2009). Among the cereals, barley has the highest content of (1,3: 1,4)-β-d-glucan (2.5–11.3 %), followed by oat (2.2–7.8 %), rye (1.2–2 %) and wheat (0.4–1.4 %) (Izydorczyk and Dexter 2008). Beta-glucans have been reported to lower serum cholesterol, improve lipid metabolism, reduce glycemic index and even reduce the risk of colorectal cancer (Keegstra and Walton 2006; Pennisi 2009). Beta-glucans are excellent prebiotic components of functional foods because they selectively promote the growth of lactobacilli and bifidobacteria in vivo (Snart et al. 2006) and in vitro (Jaskari et al. 1993).

Genetic and environmental variation of β-glucan content in barley has been investigated by various workers (Stuart et al. 1988; Kenn et al. 1993; Fastnaught et al. 1996). Although the genetic variation exists for breeding high β-glucan barley lines, environmental variation strongly impacts the β-glucan content and hence breeding for consistently high β-glucan content (Shewry 2008). Manickavelu et al. (2011) mapped four quantitative trait loci (QTL) on chromosomes 3A, 1B, 5B and 6D in a wheat recombinant inbred population contributing up to 43 % of variation in β-glucan content.

The synthesis of (1,3: 1,4)-β-D-glucan is mediated by *cellulose synthase-like* (*Csl*) genes that share a superfamily with *cellulose synthase* (*CesA*) genes. The Csl proteins are predicted to be integral membrane proteins having a "DDDQXXRW" motif (Hazen et al. 2002). Thirty-seven *Csl* genes are known in rice which belong to six families, *CslA*, *CslC*, *CslD*, *CslE*, *CslF*, and *CslH*, having 10, 9, 4, 5, 8 and 2 genes, respectively.

Burton et al. (2006) used a comparative genomics approach to clone the *CslF* group of genes on rice chromosome 7 that correspond to a highly significant QTL on barley chromosome 2H affecting (1,3: 1,4)-β-D-glucan content in mature barley grain (Han et al. 1995). Burton et al. (2006) identified six genes (*OsCslF1*, *OsCslF2*, *OsCslF3*, *OsCslF4, OsCslF8* and *OsCslF9*) located on a 118 kb interval on chromosome 7 in rice. These genes when mobilized into *Arabidopsis* resulted in (1,3: 1,4)-β-D-glucan synthesis in cell walls, which is lacking in wild type plants. The other two genes of this family, *OsCslF6* and *OsCslF7*, are located on rice chromosomes 8 and 10, respectively (Burton et al. 2006).

Burton et al. (2008) identified and mapped seven genes of the *HvCslF* family in barley. Of these seven genes, *HvCslF3*, *HvCslF4*, *HvCslF8* and *HvCslF10* were located in the centromeric region of chromosome 2H; *HvCslF6* near the centromere on 7H; *HvCslF7* on 5H long arm and *HvCslF9* on 1H short arm near the centromere. Transcript profiles of the *HvCslF* family members showed individual patterns of abundance in different tissues, with the exception of *HvCslF6*, which showed consistently higher expression in many of the tissues examined (Burton et al. 2008). Later, Burton et al. (2010) reported that over-expression of barley *HvCslF6* under the control of endosperm specific oat globulin promoter resulted in more than 80 % increase in (1,3: 1,4)-β-D-glucan content in transgenic barley grains.

Nemeth et al. (2010) used microarray analysis to identify potential candidate genes involved in (1,3: 1,4)-β-D-glucan synthesis in wheat using cDNA isolated from whole caryopses and fractions enriched with starchy endosperm tissue, during various stages of development. They found that *TaCslF6*, an ortholog of barley gene *HvCslF6*, had high expression in wheat endosperm and, moreover, its down regulation by RNAi resulted in decreased (1,3: 1,4)-β-D-glucan content in the endosperm. In oat, Chawade et al. (2010) used Targeting induced Local Lesions in Genome (TILLING) to identify mutants in the *AsCslF6* gene that affected (1,3: 1,4)-β-D-glucan content. Comparative genomics, expression profiling, mutant selection and gene knockouts are providing better understanding of the enzymes and genes regulating cell wall synthesis and it will be possible to manipulate cell wall composition of cereal grains in the near future to meet the dietary and industrial requirements.

10.2.2 Arabinoxylans

Arabinoxylans are linear chain backbone consisting of β-D-xylopyranosyl (Xyl*p*) residues linked through (1 → 4) glycosidic linkages. Some of the Xyl*p* residues have α-L-arabinofuranosyl (Ara*f*) residues attached to them, leading to four structural elements in the molecules of arabinoxylans viz., monosubstituted Xyl*p* at O-2 or O-3, di-substituted Xyl*p* at O-3, and unsubsituted Xyl*p*. The relative ratios of these structural elements vary across species (Izydorczyk and Dexter 2008). Arabinoxylans (AX) improve gut health, by promoting growth of useful bifidobacteria (Glei et al. 2006; Neyrinck et al. 2011). Neyrinck et al. 2012 found that wheat-derived arabinoxylans increased satietogenic gut peptides and reduced metabolic endotoxemia in diet-induced obese mice. The genes for the assembly of arabinoxylans are not well characterized, although the genes of cellulose synthase-like (*Csl*) and Glycosyl trasferases (*GT*) families have been reported to play important roles in synthesis and feruloylation of arabinoxylans (Urahara et al. 2004; Mitchell et al. 2007). Mitchell et al. (2007) used a bioinformatics approach with differential expression of orthologous genes between *Arabidopsis* and rice, to identify genes involved in AX synthesis and feruloylation, assuming that AX synthesis genes will be expressed more in grasses than in dicots. Genes of families *GT43*, *GT47* and *GT61* and proteins containing the PF02458 domain, which are expressed at higher levels in grasses and are integral membrane proteins, were reported to be the candidates for AX synthesis (Mitchell et al. 2007). They reported that genes in GT43 family coded β, 1 → 4 xylan synthase, GT47 family encoded xylan α-1,2 or α-1,3 arabinosyl transferases and genes in GT61 family encoded feruloyl-AX-β-1,2 Xylosyl transferases. Oikawa et al. (2010), using plant protein family information-based predictor for endomembrane (PFANTOM) reported that GT43 and GT47 family genes play important role in xylan synthesis in rice. Bosch et al. (2011) while studying cell wall biogenesis in maize elongating and non-elongating internodes found maize ortholgoues of rice GT61, GT43

and GT47 to be the most promising candidates for xylan synthesis. More studies are needed to develop better understanding of cell wall biosynthesis in cereals to manipulate the levels of AX to be nutritionally beneficial (Bosch et al. 2011).

10.2.3 Resistant Starch

Starch is composed of two structural components, amylose and amylopectin. Amylose is a long, essentially linear, polymer of glucose monomers with α-1,4 linkages whereas amylopectin is more complex with α-1,6 branching in addition to the α-1,4 bonds. Generally reserve starches contain amylose and amylopectin in the ratio of about 1:3 (Rahman et al. 2007). High amylose content is associated with starch resistant to digestion by the amylolytic enzymes present in the upper digestive tract and acts as a substrate for fermentation by the microflora inhabiting the large intestine (Bird and Topping 2001; Ito et al. 1999). Short-chain fatty acids produced as a result have been reported to benefit gut health (Topping and Clifton 2001). Figure 10.1 shows the amylose biosynthesis pathway in cereals. Enzymes significantly affecting amylose content in cereals are Granule Bound Starch Synthase-I (GBSS-I), Starch Synthase (SS) I–IV and Starch Branching Enzymes (SBE) I-II.

GBSS-I (Wx) is essential for amylose biosynthesis in wheat, rice and maize, and the absence of GBSS-I leads to waxy endosperm with no amylose (Shannon and Garwood 1984; Kirubuchi-Otobe et al. 1997). Nakamura et al. (1995) combined the null alleles of *GBSS-I* homoeoloci of wheat to produce waxy or amylose-free wheat. Slade et al. (2005) used TILLING to identify mutations in all the three homoeoloci of *GBSS-I* in hexaploid and tetraploid wheat cultivars and combined all the mutant homoeoalleles to produce a waxy phenotype with a significantly reduced level of amylose content. Amylopectin branching or content does not appear to be affected by the absence of GBSS-I (Rahman et al. 2007). Itoh et al. (2003) showed overexpression of *Wx* to increase amylose content in rice, but more studies are needed to propose this as a method of choice.

Yamamori et al. (2000) demonstrated *SSIIa* to be more important in determining the structure of amylopectin. Mutation in this gene in wheat resulted in shorter chain starch molecules and about 35 % higher amylose content over wild type (Yamamori et al. 2000). The amount of resistant starch in the high amylose SSIIa mutant increased by more than 10-fold after autoclaving as compared to wild type wheat in the native state (Yamamori et al. 2006). In barley, mutation in *SSIIa* leads to even higher increase (65 %) in amylose content compared to the wild type (Morell et al. 2003). In maize, Zhang et al. (2004) demonstrated that an insertion in *SSIIa* leads to a sugary-2 mutation with a simultaneous increase of 26–40 % in amylose content. *SSIa* mutants did not affect amylose content in rice (Fujita et al. 2006). Mutants for the *SSIa* gene have not been reported in other cereals (Rahman et al. 2007).

OsSSIIIa may play an important role in generating long chains of starch molecules in rice (Ryoo et al. 2007). Null mutants of rice *SSIIIa*, generated by T-DNA insertions, had smaller and rounder starch granules that were loosely packed in

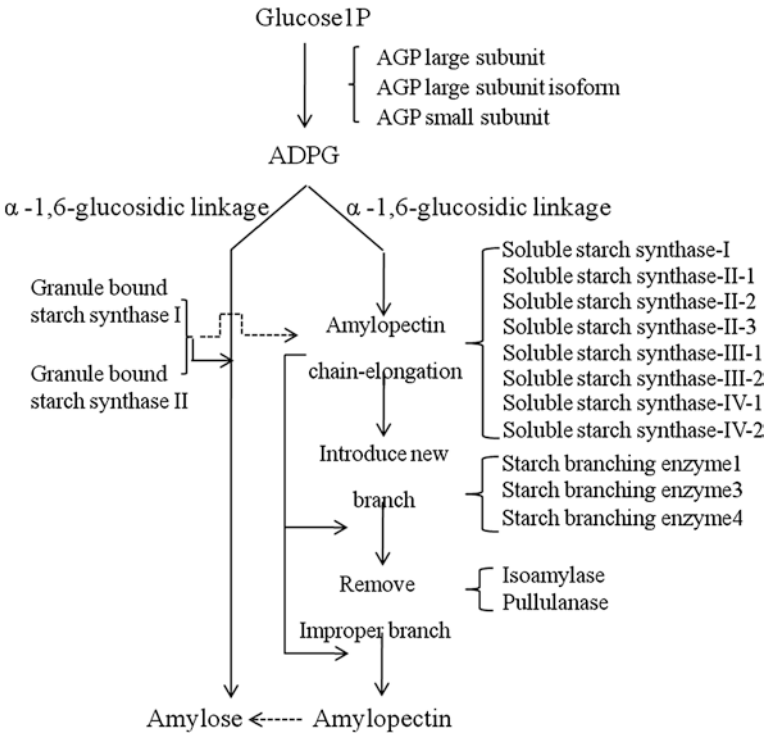

Fig. 10.1 Starch synthesis in cereals and the role of various enzymes in different steps of starch synthesis. Modified from Tian et al. (2009)

the endosperm. Hence, the *Oryza sativa SSIIIa* (*OsSSIIIa*) mutations were named *white-core floury endosperm* 5–1 (*flo5–1*) and *flo5–2* and had reduced content of long chains having DP 30 or above. The loss in *SSIII* of maize led to the dull-I phenotype (Gao et al. 2001) and only moderate increase in amylose content.

SBE induces 1,6 branching in starch and thus is important for amylopectin formation. SBE-I, SBE-IIa and SBE-IIb are the three isoforms of the enzyme. Loss of SBE activity leads to an increased level of amylose. In maize, loss of *SBEIIa* leads to nearly 80 % higher amylose content and has been commercially exploited as amylose extender to produce Hi-maize (Brown 2004). A similar mutation in rice led to an increase in amylose content of only 25–35 % (Nishi et al. 2001). In wheat, down-regulation of *SBEIIa* and *SBEIIb* by RNAi increased amylose content by 80 % (Regina et al. 2006). The high amylose wheat lines in rat feeding trials showed the benefits of resistant starch on gut health (Regina et al. 2006). Likewise, in durum wheat silencing of *SBEIIa* by RNAi led to an increase of amylose content up by 75 % (Sestili et al. 2010). However, a similar approach with *SBEIIb* did not increase amylose content. In barley, simultaneous down-regulation of both *SBEIIa and SBEIIb* by RNAi by more than 80 % produced a high amylose phenotype (>70 %) whereas a reduction in the expression of either

of these isoforms alone had only a minor impact on amylose content (Regina et al. 2010). Thus, increasing the expression of *GBSS* and decreasing *SSII* and *SBEII* activities have been successfully used to increase resistant starch in cereal grains.

10.2.4 Inulin

Inulin is a member of fructan group of polysaccharides having chains of β (2–1) linked fructose units (Degree of polymerization, DP: 2–60) attached to a sucrose molecule. It is highly water soluble alternative storage form of carbohydrate and occurs in the cell vacuoles of about 15 % of the species of the flowering plants (Hellwege et al. 2000). The most common dietary sources of inulin are wheat, onion, garlic, banana and leek. Because of the β-configuration of the anomeric C2 in its fructose monomers, inulin resists hydrolysis by the human small intestine digestive enzymes which specifically hydrolyze α-glycosidic bonds (Roberfroid 2007). In the colon, inulin supports growth of useful bacteria that are beneficial in preventing colon cancer (Reddy et al. 1997; Poulsen et al. 2002). As an ideal dietary fiber, inulin increases fecal biomass and regularizes bowel habits (Gibson et al. 1995; Kleessen et al. 1997). It is also known to enhance bioavailability of minerals in the diet (Abrams et al. 2007) and to improve body defense mechanisms (Guarner 2005).

The inulin biosynthesis model was first proposed by Edelman and Jefford (1968) in *Helianthus tuberosus*. Of the two enzymes, sucrose: sucrose 1-fructosyltransferase (1-SST) and fructan: fructan 1-fructosyltransferase (1-FFT), 1-SST catalyzes the synthesis of the trisaccharide 1-kestose from two molecules of sucrose. Subsequently, 1-FFT transfers fructosyl residues reversibly from one fructan to another, producing a mixture of fructans with variable chain lengths. Some modifications have been reported in this generalized model (Duchateau et al. 1995). In vitro synthesis of inulin using 1-SST from *H. tuberosus* (Lüscher et al. 1996) and 1-FFT from *Chicorium intybus* (Van den Ende and Laere 1996) yielded fructans with DP less than 25. In the modified model, enzyme 6-fructosyltransferase (6-FST) introduces new fructosyl units in the elongating fructan chain (Nagaraj et al. 2004). Furthermore, enzymes such as fructan exohydrolases (FEHs) can modify the structure of synthesized fructan by specific trimming of fructosyl chains.

Sprenger et al. (1995) were the first to clone a gene for a plant enzyme for fructan biosynthesis, *6-FST*, from barley. Transformation of *Nicotiana plumbaginifolia*, lacking fructans, with barley *6-FST* led to fructan production (Sprenger et al. 1995). Kawakami and Yoshida (2002) cloned *6-FST* and *1-SST* from wheat. Functional characterization was done in the methylotrophic yeast *Pichia pastoris*, which showed fructosyltransferase activity upon transformation. Kawakami and Yoshida (2005) cloned *1-FFT* gene from wheat and studied its function by overexpressing it in *P. pastoris*. Their results indicated that *1-FFT* is essential for biosynthesis of fructans accumulating in frost-tolerant wheat. Fructan accumulates in wheat stems during growth and anthesis, from where it is mobilized to grains by fructan 1-exohydrolase (1-FEH) activity during grain filling. Van den Ende et al. (2003) cloned two

isoforms of *1-FEH* in wheat and showed that they play important role in trimming fructans not only during grain filling but also during active fructan synthesis. Van Riet et al. (2006) cloned *fructan* 6-*exohydrolase* (6-*FEH*) from wheat and found that it plays an important role in the trimming of the fructans in conjunction with 1-*FEH*.

Huynh et al. (2008) mapped five QTL for fructan accumulation on wheat chromosomes 2B, 3B, 5A, 6D and 7A. The QTL on 6D and 7A contributed to the largest phenotypic variance of 17 and 27 %, respectively. Zhang et al. (2008) determined the intron–exon structure of 1-*FEH* genes in wheat, mapped them on chromosomes 6A, 6B and 6D and verified their postulated role in fructan accumulation in grains.

Long-chain inulin molecules are desirable for foodstuffs such as ice-cream, milkshakes, yogurt, cookies, cakes, pudding, breakfast cereal, and as a neutral base in cosmetic applications and pharmaceuticals. Jenkins et al. (2011) reported recently that long-chain inulin molecules (with DP>15) beneficially modulate microbial growth in the gut that yield healthy short chain fatty acids (SCFAs). The processes for accumulating long chain inulin molecules rather than crude mixtures of long and short chain inulin molecules in root extracts of artichoke have been developed (Hellwege et al. 2008). Manipulating the trimming enzymes of the inulin biosynthesis pathway (FEH) may be a feasible approach to accumulate long-chain inulin molecules, preferentially in the cereal grains. Bird et al. (2004a, b) reported a mutant (*M292*) in in a hull-less barley variety 'Himalaya' that lowered plasma cholesterol and enhanced short-chain fatty acids in the guts of rats and pigs. Clarke et al. (2008) reported that M292 had a mutation in *Starch synthase* (*SSIIa*) gene which, in addition to enhancing free sugars, β-glucans and arabinoxylans also increased inulin content by 42-fold compared to the wild type variety. The wild type variety 'Himalaya' had 0.1 mg/kg inulin in the grains, whereas the mutant M292 had 4.2 mg/kg grain inulin content (Clarke et al. 2008). More studies are needed to validate the role of *SSIIa* in increasing grain inulin content.

10.3 Bioactive Compounds (Class II)

10.3.1 Polyphenols

Polyphenols are compounds bearing one or more aromatic rings with one or more hydroxyl groups (Liu 2007). Though termed secondary metabolites, polyphenols play an essential role in protecting plants from UV radiation (Stalikas 2007), inhibiting pathogens (Abdel-Aal et al. 2001) and providing structural integrity to the cell wall (Klepacka and Fornal 2006). Cereals contain high levels of polyphenols that contribute in the prevention of degenerative diseases such as cancer and cardiovascular diseases (Liu 2007; He et al. 2010). The health effects of phenolic compounds depend on the amount consumed and on their bioavailability (Manach et al. 2004).

Cereals contain a variety of polyphenols including phenolic acids, flavonoids (flavonols, flavones, flavonones, isoflavones and anthocyanins),

proanthocyanidins, condensed tannins, catechins and lignans. The majority of phenolics in cereals are present in the bran fraction as insoluble and bound compounds in the form of ester and ether linkages with polysaccharides such as arabinoxylan and lignin in the cell wall (Liyana-Pathirana and Shahidi 2006; Fernandez-Orozco et al. 2010).

Genetic variation for polyphenol accumulation and composition has been documented among different cereals (Adom et al. 2003; Menga et al. 2010; Shewry et al. 2010). Significant correlations between the contents of bioactive components and environmental factors were found and even highly heritable components differed in amount over different years and sites (Fernandez-Orozco et al. 2010; Shewry et al. 2010). Bound phenolics, which comprise the greatest proportion of the total phenolics, resulted in the most heritable compounds compared to the free and conjugated forms (Fernandez-Orozco et al. 2010). Higher levels of total phenolics, ferulic acid and flavonoids were detected in Emmer wheat compared to Einkorn and bread wheat species (Li et al. 2008; Serpen et al. 2008), but further studies are needed on a larger sample of wheats with various ploidy levels.

The biosynthesis of phenolics is initiated by the shikimic acid pathway (Heldt 2005) which produces phenylalanine, the first substrate of the phenyl propanoid pathway and proceeds with the synthesis of different classes of compounds, including phenolic acids and flavonoids (Fig. 10.2). The pathway is known to be strongly affected by various stimuli including light, pathogens and wounding

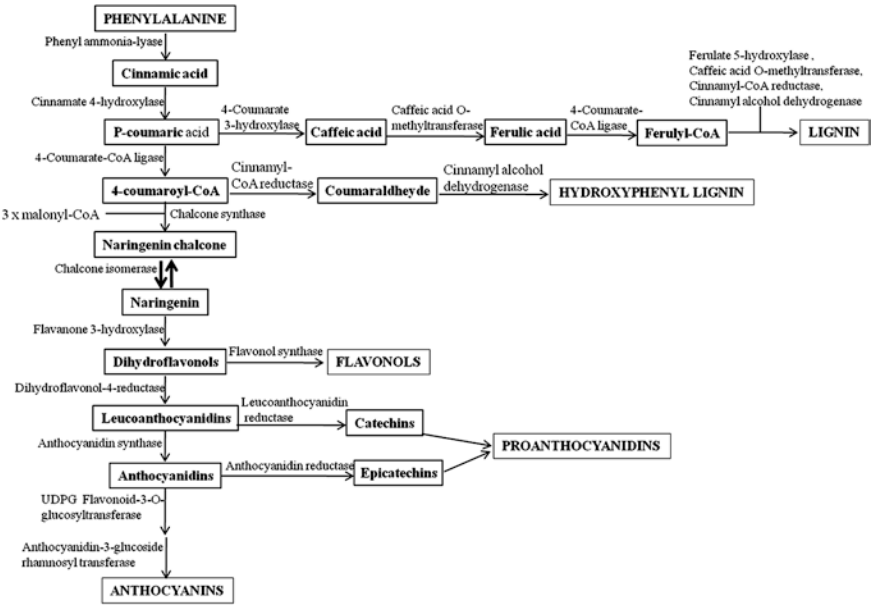

Fig. 10.2 Schematic representation of the general phenylpropanoid pathway in plants, leading to the synthesis of phenolic acids, lignin, different classes of flavonoids and proanthocyanidins. Modified from Deluc et al. (2006)

(Weaver and Herrmann 1997). Possible strategies to enhance the biosynthesis of specific phenolics include over-expression of structural genes involved in rate-limiting steps, and the manipulation of transcription factors that simultaneously activate several genes in one pathway (Grotewold 2008).

Phenolic Acids

Phenolic acids represent the most common form of phenolic compounds found in whole grains. Among these, the most abundant are derivatives of hydroxycinnamic acids (Sosulski et al. 1982). The biosynthesis of hydroxycinnamic acids begins with the deamination of phenylalanine to produce cinnamic acid by the enzyme phenylalanine ammonia-lyase (Fig. 10.2). Further enzymatic reactions include hydroxylation of the aromatic ring, methylation of selected phenolic hydroxyl groups, activation of the cinnamic acids to cinnamoyl-CoA esters, and reduction of these esters to cinnamaldehydes and cinnamyl alcohols.

In most plants, the enzyme phenylalanine ammonia-lyase (PAL) is encoded by a small gene family (Wanner et al. 1995; Zhu et al. 1995). In monocotyledons, genes involved in the synthesis of PAL were isolated from DNA libraries in rice (Minami and Tanaka 1993; Zhu et al. 1995) and wheat (Li and Liao 2003). Kervinen et al. (1998) isolated five different genes in barley encoding PAL from a root cDNA library that were highly similar to the wheat and rice *PAL* sequences. Similar approaches were used to clone other key genes involved in the biosynthesis of phenolic acids in maize (Collazo et al. 1992), wheat (Ma et al. 2002) and rice (Yang et al. 2005).

Only a few attempts have been made to specifically increase the content of phenolic acids in cereal crops. Dias and Grotewold (2003) reported higher content of ferulic, chlorogenic and other phenolic acids in cultured maize cells transformed by the transcription factor *ZmMyb-IF35*. Mao et al. (2007) studied secondary metabolism in maize lines transformed with the wheat oxalate oxidase (*OxO*) gene. In leaves of the *OxO* maize lines, the amount of phenolic acids significantly increased while synthesis of DIMBOA (2,4-dihydroxy-7-methoxy-1,4-benzoxazin-3-one), a naturally occurring hydroxamic acid insecticide was reduced. Ferulic acid exhibited the largest increase and accounted for 80.4 % of the total soluble phenolics. These results depend on a diversion in the shikimate pathway leading to production of phenolic and hydroxamic acids. More studies are needed to manipulate phenolic acid synthesis pathway in a nutritionally applicable way.

Flavonoids

Flavonoids represent a large family of low-molecular-weight phenolics involved in a wide range of functions (Dixon and Paiva 1995). In cereals, dozens of different flavonoids have been identified mostly conjugated to various sugar moieties (Dykes and Rooney 2007). Variation in flavonoid synthesis depends upon the

enzymatic function/activity of genes in either the core or side branches of the flavonoid pathway (Fig. 10.2). Multiple copies of genes and specific regulatory factors are responsible for the variation in flavonoids in different tissues and organs of plants (Dias and Grotewold 2003; Zhou et al. 2010).

The biosynthesis of flavonoids is initiated by the step catalysed by the enzyme chalcone synthase (CHS) which produces the aglycone flavonoid naringenin chalcone from malonyl-CoA and coumaroyl-CoA precursors (Heller and Forkmann 1994). In maize, CHS is encoded by a duplicated genetic locus (Wienand et al. 1986; Franken et al. 1991). In the majority of plants including cereals, chalcones are not the end-products. The pathway proceeds with several enzymatic steps to flavanones, dihydroflavonols and, finally, to the anthocyanins, the major water soluble pigments in flowers and fruits (Grotewold and Peterson 1994; Deboo et al. 1995). The synthesis of isoflavones, aurones, flavones, proanthocyanidins and flavonols is well documented in maize and more than 20 structural and regulatory genes have been identified (Mol et al. 1998; Grotewold 2006). However, little is known about the final transfer of anthocyanins into the vacuole (Marrs et al. 1995; Alfenito et al. 1998).

Most of the structural genes involved in the flavonoid pathway have been identified, characterized and mapped in wheat (Munkvold et al. 2004; Himi and Noda 2004, 2005; Himi et al. 2011). Khlestkina et al. (2008a) identified four distinct copies of *Flavanone 3-hydroxylase* (*F3H*) gene in bread wheat by PCR-based cloning. In barley, a cDNA library screened with a probe from *Antirrhinum majus* was used to clone the gene encoding flavanone-3-hydroxylase (Meldgaard 1992). Some of the genes involved in the synthesis of flavonoids in cereals have also been mapped. In wheat, *CHS* was found to map to chromosomes of homoeologous groups 1 and 2 (Li et al. 1999), *CHI* to homoeologous group 5 and 7D (Li et al. 1999), *F3H* and *DFR* to homoeologous groups 2 (Khlestkina et al. 2008b) and 3 (Himi and Noda 2004), respectively.

The regulation of flavonoid metabolism is achieved mainly through transcriptional regulation of genes involved in biosynthetic pathway (Martin et al. 2001; Davies and Schwinn 2003). A number of regulatory genes required for anthocyanin regulation have been identified, cloned, and characterized in several species. These transcription factors belong to two classes, MYB superfamily and basic-Helix-Loop-Helix (bHLH), and together with a WD40 protein, are thought to regulate the anthocyanin biosynthetic genes co-operatively (Koes et al. 2005).

Regulatory genes controlling the tissue specificity of structural genes were identified by mutant analysis in maize (Paz-Ares et al. 1986; Cone et al. 1993a, b; Pilu et al. 2003), *Arabidopsis* (Paz-Ares et al. 1987; Vom Endt et al. 2002), *Antirhinum* (snapdragon; Martin et al. 1991), *Petunia* (Quattrocchio et al. 1993), *Vitis vinifera* (grape; Deluc et al. 2008) and wheat (Himi et al. 2011). Two types of transcription factors grouped as the *R*/*B* family (basic helix–loop–helix, bHLH-type) and the *C1*/*Pl* family (Myb-type) were shown to upregulate the structural genes required for the production of anthocyanin (Consonni et al. 1993; Pilu et al. 2003). In addition, transcription factors *P*, *TT2*, *TT8* and *Del* also regulate part of the flavonoid biosynthesis (Martin et al 1991; Vom Endt et al. 2002).

The enzymes that direct the splitting of flavonoid synthesis pathway from the phenylpropanoid pathway are critical for the increased production of various flavonoids. Shin et al. (2006) obtained the novel synthesis of several classes of flavonoids in the endosperm of rice by expressing two maize regulatory genes (*C1* and *R-S*) using an endosperm-specific promoter. *C1*, when transferred to wheat induced anthocyanin pigmentation in otherwise non-pigmented wheat coleoptiles (Ahmed et al. 2003). In addition, the *R* and *Rc-1* genes were shown to upregulate key genes of the flavonoid pathway in wheat (Hartmann et al. 2005; Himi and Noda 2005; Himi et al. 2011).

10.3.2 Carotenoids

Carotenoids are pigments conferring the characteristic yellow to red color to fruits and flowers. Structurally, they are isoprenoid compounds having generally eight isoprene units and long polyene chains with 3–15 conjugated double bonds (Weedon and Moss 1995). More than 600 carotenoids have been identified in plants including α-carotene, β-carotene, lycopene, lutein, zeaxanthin, cryptoxanthin, citroxanthin and violaxanthin, etc., (Kahlon and Keagy 2003). The most famous member of the carotenoids is β-carotene, which is a precursor of vitamin A; its deficiency leads to xerophthalmia and also cataracts and macular degeneration with ageing. Carotenoids may also have protective effects in cardiovascular diseases and cancer (Kohlmeier and Hastings 1995; Astorg 1997).

Carotenoid synthesis starts in the plastids of higher plants by the action of IPP isomerase and GGPP synthase converting four molecules of isopentyl diphosphate (IPP), to geranyl geranyl diphosphate (GGPP) (Giuliano et al. 2008). Phytoene synthase subsequently condenses two molecules of GGPP to form 15-*cis*-phytoene, which is the first dedicated step in the carotenoid biosynthesis (Beyer et al. 1985). Figure 10.3 gives a schematic representation of the carotenoid biosynthesis pathway in plants.

Ye et al. (2000) produced golden rice with increased β-carotene content by introducing the *phytoene synthase* (*psy*) gene from daffodil together with a bacterial phytoene desaturase (*crtI*) gene from *Erwinia uredovora* placed under control of the endosperm-specific glutelin (Gt1) and the constitutive cauliflower mosaic virus (CaMV) 35*S* promoters, respectively. Paine et al. (2005) developed golden rice-2 with 23-fold higher total carotenoid accumulation by introducing the maize *psy* gene compared to the original golden rice (Ye et al. 2000). Giuliano et al. (2008) estimated that 100 % of the recommended dietary allowance (RDA) of vitamin A for children and 38 % for adults can be obtained with 60 g/day consumption of golden rice-2.

Wong et al. (2004) reported QTL mapping of β-carotene synthesis pathway genes in maize. The β-carotene biosynthetic pathway in maize was also studied using loss-of-function mutants (Buckner et al. 1990, 1996; Li et al. 2007; Zhu et al. 2008). A mutant of *phytoene synthase* (*y1*) of maize has white endosperm

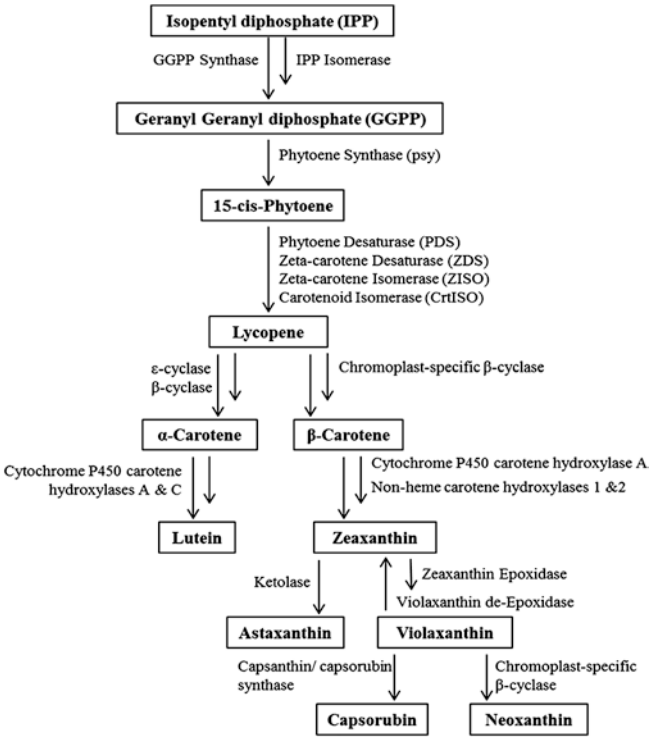

Fig. 10.3 Carotenoid biosynthesis pathway in plants. Modified from Giuliano et al. (2008)

and very low carotenoid levels. Phytoene desaturase is the second enzyme and is responsible for a two-step desaturation of phytoene to ζ (Zeta)- carotene which is then further desaturated to other forms of carotenoids such as lycopene and β-carotene. Yan et al. (2010) reported cloning of gene β-carotene hydroxylase-1 (*crtRB1*) in maize and further demonstrated a rare genetic variation in *crtRB1* to enhance β-carotene levels in maize.

Hexaploid bread wheat (*T. aestivum*) has low carotenoid levels (1.94 μg/g), whereas diploid einkorn wheat and tetraploid emmer wheat have relatively higher carotenoid content (9.62 and 6.27 μg/g, respectively), which is however, lower than that of corn (35.11 μg/g) (Panfili et al. 2004; Abdel-Aal et al. 2002, 2007). Lutein is the predominant carotenoid in wheat and comprises 80–90 % of the total carotenoid content, the remaining being zeaxanthin, β-carotene, and lutein esters (Abdel-Aal et al. 2002). Lutein content has been reported to be higher in the flour than the bran portion in all the wheat species analyzed (Abdel-Aal et al. 2002). Zhang et al. (2005) transferred yellow pigment gene (*Y*) from *Lophopyrum ponticum* to wheat cultivars. They proposed *Y* gene to be either an efficient enzyme in early steps of carotenoid biosynthetic pathway or a regulatory factor that affects several steps of the carotenoid biosynthetic pathway (Zhang et al. 2005).

Pozniak et al. (2007) mapped genes *psy1* and *psy2* on group-7 and -5 chromosomes, respectively, in durum wheat, of which *psy1* had a strong association with yellow pigment content of endosperm (Pozniak et al. 2007; Singh et al. 2009). A similar association is known in maize endosperm yellow pigment and maize *psy1* gene (Gallagher et al. 2004). Zhang and Dubcovsky (2008) isolated the *psy1-A* and *psy1-B1* genes from two durum cultivars, which was followed by the development of functional markers for flour color in wheat by He et al. (2009). Wang et al. (2009) cloned and made a phylogenetic analysis of the *psy1* gene in common wheat and related species. All the genes had six exons and five introns. Sequence divergence due to single nucleotide polymorphisms (SNPs) and insertion deletions (InDels) were present among the different clusters. Cong et al. (2010) cloned cDNA and made an expression analysis of the wheat *phytoene desaturase* (*PDS*) and ζ-carotene desaturase (*ZDS*) genes and found them to have high homology with those of other higher plant species.

10.3.3 Tocopherols and Tocotrienols (Vitamin E)

Vitamin E is a family of fat-soluble antioxidants consisting of α-, β-, γ-, and δ-tocopherols and the corresponding α-, β-, γ-, and δ- tocotrienols. Alpha-tocopherol is the form of vitamin E that is preferentially absorbed and accumulated in humans (Rigotti 2007). Compared to tocopherols, tocotrienols have been less investigated, although they show higher antioxidant potential (Sen et al. 2006). This is due to widespread occurrence of tocopherols in plants as the principal vitamin E components of leaves and seeds in most dicot species (Padley et al. 1994). On the other hand, tocotrienols typically account for the majority of the total vitamin E content in the seeds of monocots, such as rice, wheat and oats (Peterson and Qureshi 1993; Padley et al. 1994). From the human health point of view, tocotrienols have been shown to have specialized roles in protecting neurons from damage (Sen et al. 2006) and in cholesterol reduction (Das et al. 2008). Tocopherol compounds, in both durum and bread wheat are mostly present in the germ fraction (Panfili et al. 2003; Borrelli et al. 2008). Table 10.3 summarizes the content of various components of vitamin E in the grains of common cereals.

The tocopherol biosynthetic pathway in plants has been extensively studied for over 30 years (Whistance and Threlfall 1970; Grusak and DellaPenna 1999) and the enzymes and genes of the pathway have been isolated (DellaPenna 2005). With the exception of Vitamin-E-defective (VTE3) (Cheng et al. 2003), tocopherol biosynthetic enzymes share significant homology between plants and cyanobacteria, underscoring the evolutionary relationship between these organisms.

The first step in tocopherol synthesis involves the production of the aromatic head group, homogentisic acid (HGA), from p-hydroxyphenylpyruvic acid (HPP) by the enzyme p-hydroxyphenylpyruvic acid dioxygenase (HPPD), as reviewed by DellaPenna (2005). Cahoon et al. (2003) isolated *HPT* from tocotrienol-accumulating seeds of barley, wheat and rice and expressed barley *HPT* in tobacco calli

Table 10.3 Tocopherol and Tocotrienol content of major cereal flour, milling products and selected products (modified from Piironen et al. 1986)

Cereal grain	Tocopherols (T) and Tocotrienols (T3) in mg/g of fresh product							Vitamin E
	α-T	**α-T3**	**β-T**	**β-T3**	**γT**	**γT3**	**δT**	**(mg of α-T eq)**
Wheat flour	1.6	0.3	0.8	1.7	–	–	–	2.1
Wheat bran	1.6	1.5	0.8	5.6	–	–	–	2.7
Wheat germ	22.1	0.3	8.6	1.0	–	–	<0.1	25.7
Rye flour	0.6	0.4	0.3	0.6	–	–	–	0.8
Rye meal	1.0	1.4	0.3	1.1	–	–	–	1.6
Barley meal	0.3	1.6	<0.1	0.6	0.1	0.6	<0.1	0.8
Rice, brown	0.6	0.4	0.1	<0.01	0.1	0.7	<0.1	0.8
Rice, polished	<0.1	0.1	<0.1	<0.1	<0.1	0.3	<0.1	0.1
Millet, whole grain	0.1	<0.1	0.1	–	1.7	<0.1	0.6	0.3
Oats, rolled	0.8	2.0	0.1	0.3	–	<0.1	–	1.5
Oats, puffed	0.8	1.4	0.1	0.3	–	<0.1	<0.1	1.3
Semolina	0.15	0.11	0.08	1.33	–	–	–	0.28
Macaroni	0.1	0.1	0.1	0.8	–	–	–	0.2
Corn flake	<0.1	0.2	–	–	0.1	0.4	<0.1	0.1
Popcorn	0.35	0.27	0.03	–	2.61	0.23	0.15	0.71

The components present in trace amounts are shown by a '–'

using the CaMV 35S promoter. Barley *HGGT* was expressed in *Arabidopsis thaliana* leaves, which accumulated large amounts of tocotrienols upon transformation. High tocotrienol corn was designed by expressing barley *HGGT* in maize, under the control of embryo specific promoter for corn *oleosin* gene, showing that a single metabolic step was sufficient to enhance the effective level of vitamin E six-fold (Cahoon et al. 2003).

10.4 Future Perspectives

Functional food components vary across the cereal crops and within different tissues of the grain. Knowledge of the genetics, biochemistry and genomics of functional food components also differs among crop plants and is more advanced in rice and corn than in wheat, barley and oats. Moreover, large genome size of wheat, barley and oats, together with polyploidy in wheat and oats further complicate genetic and genomic analysis. High-quality sequences of wheat genome and genes are urgently needed and will greatly accelerate functional food component research.

The next challenge will be to elucidate metabolic pathways and structural and regulatory genes for functional food components. As the reviewed literature reveals, this work is already in progress and needs to be continued at an accelerated pace. Comparative genomics and bioinformatics-based approaches will be useful in leveraging information from model organisms, rice and maize to other cereal crops. However, many genes are crop-specific, so that functional genomics tools must be developed in each cereal crop plant. In this regard, TILLING appears to be a versatile tool for crops such as wheat and barley where other functional tools are not that well developed. TILLING will be useful for mining novel alleles of genes of metabolic pathways, increasing diversity in the trait of interest, as demonstrated by the directed search of specific mutants for high amylose starch. However, TILLING may not be feasible for multigene families where techniques such as RNAi may be more appropriate for knocking down specific gene activity. A transgenic approach was used to produce golden rice but public acceptance has been problematic. TILLING is a promising strategy for the targeted breeding for genes of interest with no biosafety issues because it is an entirely non-transgenic approach. Genetics, breeding and transgenic approaches have been and can be used to design cereal crops with optimum expression of functional food compounds such as β-glucan, amylose, inulin, phenolics, flavonoids, carotenoids, and vitamin E.

Wild germplasm is another untapped resource of useful genetic variation in the functional food compounds. In the past, related wild species have been used as sources of many useful genes for resistance against biotic and abiotic stresses, but they have not been used so far in improvement of cereals for their use as functional foods. Evaluating natural variation in the wild relatives of crop plants for functional food components and molecular breeding of those traits for increasing the functional food value of cereal crops should be fully explored.

10.5 Summary and Outlook

Cereals are major components of the human diet, and the content of compounds that are beneficial to human health has become a fascinating and important subject of research. With increasing knowledge of the biosynthetic pathways of functional food components, the exact roles played by the various genes involved and the factors affecting the end product, it is becoming increasingly possible to design cereal crops as functional foods, with nutritional role beyond use as a source of calories.

Acknowledgments The authors are thankful to W. Jon Raupp (Wheat Genetic and Genomic Resources Center, Kansas State University, USA) for critical reading of the article. This chapter has been submitted as Contribution No. 12-103-B from the Kansas Agricultural Experiment Station, Kansas State University, USA. Research was supported by a grant from Heartland Plant Innovations (HPI).

References

Abdel-Aal ESM, Huci P, Sosulski FW, Graf R, Gillott C, Pietrzak L (2001) Screening spring wheat for midge resistance in relation to ferulic acid content. J Agric Food Chem 49:3559–3566

Abdel-Aal ESM, Young JC, Wood PJ, Rabalski I, Hucl P, Falk D, Frégeau-Reid J (2002) Einkorn: a potential candidate for developing high lutein wheat. Cereal Chem 79:455–457

Abdel-Aal ESM, Young JC, Rabalski I, Hucl P, Fregeau-Reid J (2007) Identification and quantification of seed carotenoids in selected wheat species. J Agric Food Chem 55:787–794

Abrams SA, Hawthorne KM, Aliu O, Hicks PD, Chen C, Griffin IJ (2007) An Inulin-type fructan enhances calcium absorption primarily via an effect on colonic absorption in humans. J Nutr 137:2208–2212

Adom KK, Sorrells ME, Liu RH (2003) Phytochemical profiles and antioxidant activity of wheat varieties. J Agric Food Chem 51:7825–7834

Ahmed N, Maekawa M, Utsugi S, Himi E, Ablet H, Rikiishi K, Noda K (2003) Transient expression of anthocyanin in developing wheat coleoptile by maize *C1* and *B-peru* regulatory genes for anthocyanin synthesis. Breeding Sci 53:29–34

Alfenito MR, Souer E, Goodman CD, Buell R, Mol J, Koes R, Walbot V (1998) Functional complementation of anthoanthocyanin sequestration in the vacuole by widely divergent glutathione S-transferases. Plant Cell 10:1135–1149

Arai S (1996) Studies on functional foods in Japan- states of the art. Biosci Biotechnol Biochem 60:9–15

Astorg P (1997) Food caotenoids and cancer prevention: an overview of current research. Trends Food Sci Technol 8:406–413

Behall KM, Scholfield DJ, Hallfrisch J (2006) Whole-grain diets reduce blood pressure in mildly hypercholesterolemic men and women. J Am Diet Assoc 106:1445–1449

Beyer P, Weiss G, Kleinig H (1985) Solubilization and reconstitution of the membrane bound carotenogenic enzymes from daffodil chromoplasts. Eur J Biochem 153:341–346

Bird AR, Topping DL (2001) Resistant starches, fermentation, and large bowel health. In: Cho SS, Dreher ML (eds) Handbook of dietary fiber. Marcel Dekker, New York, pp 147–158

Bird AR, Flory C, Davies DA, Usher S, Topping DL (2004a) A novel barley cultivar (Himalaya 292) with a specific gene mutation in starch synthase IIa raises large bowel starch and short-chain fatty acids in rats. J Nutr 134:831–835

Bird AR, Jackson M, King RA, Davies DA, Usher S, Topping DL (2004b) A novel high-amylose barley cultivar (Hordeum vulgare var. Himalaya 292) lowers plasma cholesterol and alters indices of large-bowel fermentation in pigs. Br J Nutr 92:607–615

Borrelli GM, De Leonardis AM, Platani C, Troccoli A (2008) Distribution along durum wheat kernel of the components involved in semolina colour. J Cereal Sci 48:494–502

Bosch M, Mayer CD, Cookson A, Donnison IS (2011) Identification of genes involved in cell wall biogenesis in grasses by differential gene expression profiling of elongating and non-elongating maize internodes. J Exp Bot 62(10):3545–3561

Brown IL (2004) Applications and uses of resistant starch. J AOAC Int 87:727–732

Buckner B, Kelson TL, Robertson DS (1990) Cloning of the *y1* locus of maize, a gene involved in the biosynthesis of carotenoids. Plant Cell 2:867–876

Buckner B, San Miguel P, Janick-Buckner D, Bennetzen JL (1996) The *y1* gene of maize codes for phytoene synthase. Genetics 143:479–488

Burton RA, Fincher GB (2009) (1,3;1,4)-β-D-Glucans in cell walls of the Poaceae, lower plants, and fungi: a tale of two linkages. Mol Plant 2:873–882

Burton RA, Wilson SM, Hrmova M, Harvey AJ, Shirley NJ, Medhurst A, Stone BA, Newbigin EJ, Bacic A, Fincher GB (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, Shirley NJ, Mather DE, Bacic A, Fincher GB (2008) The genetics and transcriptional profiles of the cellulose synthase-like *HvCslF* gene family in barley. Plant Physiol 146:1821–1833

Burton RA, Gidley MJ, Fincher GB (2010) Heterogeneity in the chemistry, structure and function of plant cell walls. Nat Chem Biol 6:724–732

Cahoon EB, Hall SE, Ripp KG, Ganzke TS, Hitz WD, Coughlan SJ (2003) Metabolic redesign of vitamin E biosynthesis in plants for tocotrienol production and increased antioxidant content. Nat Biotechnol 21:1082–1087

Caimi PG, McCole LM, Klein TM, Kerr PS (1996) Fructan accumulation and sucrose metabolism in transgenic maize endosperm expressing *Bacillus amyloliquifaciens SacB* Gene. Plant Physiol 110:355–363

Champ M (2008) Determining the functional properties of food components in the gastrointestinal tract. In: Hamaker BR (ed) Technology of functional cereal products. WoodHead Publishing in Food Science, Technology and Nutrition. Cambridge, England, pp 126 − 154

Charalampopoulos D, Wang R, Pandiella SS, Webb C (2002) Application of cereals and cereal components in functional foods: a review. Int J Food Microbiol 79:131–141

Chawade A, Bräutigam AP, Larsson M, Vivekanand V, All Nakash M, Chen T, Olsson O (2010) Development and characterization of an oat TILLING-population and identification of mutations in lignin and beta-glucan biosynthesis genes. BMC Plant Biol 10:86

Cheng Z, Sattler S, Maeda H, Sakuragi Y, Bryant DA, DellaPenna D (2003) Highly divergent methyltransferases catalyze a conserved reaction in tocopherol and plastoquinone synthesis in cyanobacteria and photosynthetic eukaryotes. Plant Cell 15:2343–2356

Clarke B, Liang R, Morell MK, Bird AR, Jenkins CLD, Li Z (2008) Gene expression in a starch synthase IIa mutant of barley: changes in the level of gene transcription and grain composition. Funct Integr Genomics 8:211–221

Collazo P, Montoliu L, Puigdomenech P, Rigau J (1992) Structure and expression of the lignin O-methyltransferase gene from *Zea mays* L. Plant Mol Biol 20:857–867

Cone KC, Cocciolone SM, Burr FA, Burr B (1993a) Maize anthocyanin regulatory gene *pl* is a duplicate of *c1* that functions in the plant. Plant Cell 5:1795–1805

Cone KC, Cocciolone SM, Moehlenkamp CA, Weber T, Drummond BJ, Tagliani LA, Bowen BA, Perrot GH (1993b) Role of the regulatory gene *pl* in the photo-control of maize anthocyanin pigmentation. Plant Cell 5:1807–1816

Cong L, Wang C, Li Z, Chen L, Yang G, Wang Y, He G (2010) cDNA cloning and expression analysis of wheat (*Triticum aestivum* L.) phytoene and ζ-carotene desaturase genes. Mol Biol Rep 37:3351–3361

Consonni G, Geuna F, Gavazzi G, Tonelli C (1993) Molecular homology among members of the *R* gene family in maize. Plant J 3:335–346

Cox IM, Campbell MJ, Dowson D (1991) Red blood cell magnesium and chronic fatigue syndrome. Lancet 337(8744):757–760

Das S, Lekli I, Das M, Szabo G, Varadi J, Juhasz B, Bak I, Nesaretam K, Tosaki A, Powell SR, Das DK (2008) Cardioprotection with palm oil tocotrienols: comparison of different isomers. Am J Physiol Heart Circ Physiol 294:H970–H9788

Davies KM, Schwinn KE (2003) Transcriptional regulation of secondary metabolism. Funct Plant Biol 30:913–925

Deboo GB, Albertsen MC, Taylor LP (1995) Flavanone 3-hydroxylase transcripts and flavonol accumulation are temporally coordinated in maize anthers. Plant J 7:703–713

DellaPenna D (2005) A decade of progress in understanding vitamin E synthesis in Plants. J Plant Physiol 162:729–737

Deluc L, Barrieu F, Marchive C, Lauvergeat V, Decendit A, Richard T, Carde JP, Me′rillon JM, Hamdi S (2006) Characterization of a grape vine R2R3-MYB transcription factor that regulates the phenylpropanoid pathway. Plant Physiol 140:499–511

Deluc L, Bogs J, Walker AR, Ferrier T, Decendit A, Merillon JM, Robinson SP, Barrieu F (2008) The Transcription factor VvMYB5b contributes to the regulation of anthocyanin and proanthocyanidin biosynthesis in developing grape berries. Plant Physiol 147:2041–2053

Dermibas A (2005) β-glucan and mineral nutrient contents of cereals grown in Turkey. Food Chem 90:737–777

Dias AP, Grotewold E (2003) Manipulating the accumulation of phenolics in maize cultured cells using transcription factors. Biochem Eng J 14:207–216

Dixon RA, Paiva NL (1995) Stress-induced phenylpropanoid metabolism. Plant Cell 7:1085–1097

Duchateau N, Bortlik K, Simmen U, Wiemken A, Bancal P (1995) Sucrose:fructan 6-fructosyltransferase (6-SFT), a key enzyme for diverting carbon from sucrose to fructan in barley leaves. Plant Physiol 104:1249–1255

Dykes L, Rooney LW (2007) Phenolic compounds in cereals and their health benefits. Cereal Foods World 52:105–111. doi:10.1016/j.chroma.2009.08.041

Edelman J, Jefford TG (1968) The mechanism of fructan metabolism in higher plants as exemplified in *Helianthus tuberosus*. New Phytol 67:517–531

FAO Corporate Documentary Repository. World agriculture: towards 2015/2030—An FAO perspective http://www.fao.org/docrep/005/y4252e/y4252e04b.htm

Fardet A, Rock E, Rémésy C (2008) Is the *in vitro* antioxidant potential of whole-grain cereals and cereal products well reflected *in vivo*? J Cereal Sci 48:258–276

Fastnaught CE, Berglund PT, Holm ET, Fox GJ (1996) Genetic and environmental variation in β-glucan content and quality parameters of barley for food. Crop Sci 36:941–946

Fernandez-Orozco R, Li L, Harflett C, Shewry PR, Ward JL (2010) Effects of environment and genotype on phenolic acids in wheat in the HEALTHGRAIN diversity screen. J Agric Food Chem 58:9341–9352. doi:10.1021/jf100263c

Fincher GB (2009) Exploring the evolution of (1,3;1,4)-β-D-glucans in plant cell walls: comparative genomics can help! Curr Opin Plant Biol 12:140–147

Franken P, Niesbach-Klosgen U, Weydemann U, Marechal-Drouard L, Saedler H, Wienand U (1991) The duplicated chalcone synthase genes *C2* and *Whp* (white pollen) of *Zea mays* are independently regulated: evidence for translational control of *Whp* expression by the anthocyanin intensifying gene in. EMBO J 10:2605–2612

Fretzdorff B, Welge N (2003) Fructan and raffinose contents in cereals and pseudocereal grains. Getreide Mehl und Brot 57:3–8

Fujita N, Yoshida M, Asakura N, Ohdan T, Miyao A, Hirochika H, Nakamura Y (2006) Function and characterization of starch synthase I using mutants in rice. Plant Physiol 140:1070–1084

Fuller R (1989) Probiotics in man and animals. J Appl Bacteriol 66:365–378

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

Gao M, Wanat J, Stinard PS, James MG, Myers AM (2001) Characterization of *dull1*, a maize gene coding for a novel starch synthase. Plant Cell 10:399–412

Gibson GR, Roberfroid MB (1995) Dietary modulation of the human colonic microbiota: introducing the concept of prebiotics. J Nutr 125:1401–1412

Gibson GR, Beatty ER, Wang X, Cummings JH (1995) Selective stimulation of bifidobacteria in the human colon by oligofructose and inulin. Gastroenterology 108:975–982

Giuliano G, Tavazza R, Diretto G, Beyer P, Taylor MA (2008) Metabolic engineering of carotenoid biosynthesis in plants. Trends Biotechnol 26:139–145

Glei M, Hoffman T, Küster K, Hollmann J, Lindhauer MG, Pool-Zobel BL (2006) Both wheat (*Triticum aestivum*) bran arabinoxylans and gut flora-mediated fermentation products protect human colon cells from genotoxic activities of 4-hydroxynonenal and hydrogen peroxide. J Agric Food Chem 54:2088–2095

Grotewold E (ed) (2006) The science of flavonoids. Springer, New York

Grotewold E (2008) Transcription factors for predictive plant metabolic engineering: are we there yet? Curr Opin Biotech 19:138–144

Grotewold E, Peterson T (1994) Isolation and characterization of a maize gene encoding chalcone flavanone isomerase. Mol Gen Genet 242:1–8

Grusak MA, DellaPenna D (1999) Improving the nutrient composition of plants to enhance human nutrition and health. Annu Rev Plant Physiol Plant Mol Biol 50:133–161

Guarner F (2005) Inulin and oligofructose: impact on intestinal diseases and disorders. Br J Nutr 93(Suppl 1):561–565

Haas JD, Brownlie T (2001) Iron deficiency and reduced work capacity: a critical review of the research to determine a causal relationship. J Nutr 131:691S–696S

Han F, Ullrich S, Chirat S, Menteur S, Jestin L, Sarrafi A, Hayes P, Jones B, Blake T, Wesenberg D, Kleinhofs A, Kilian A (1995) Mapping of beta-glucan content and beta-glucanase activity loci in barley grain and malt. Theor Appl Genet 91:921–927

Hartmann U, Sagasser M, Mehrtens F, Stracke R, Weisshaar B (2005) Differential combinatorial interactions of cis-acting elements recognized by *R2R3-MYB*, *BZIP*, and *BHLH* factors control light-responsive and tissue-specific activation of phenylpropanoid biosynthesis genes. Plant Mol Biol 57:155–171

Hazen SP, Scott-Craig JS, Walton JD (2002) Cellulose synthase-like genes of rice. Plant Physiol 128:336–340

He XY, He ZH, Ma W, Appels R, Xia XC (2009) Allelic variants of phytoene synthase 1 (Psy1) genes in Chinese and CIMMYT wheat cultivars and development of functional markers for flour colour. Mol Breeding 23:553–563

He M, van Dam RM, Rimm E, Hu FB, Qi L (2010) Whole-grain, cereal fiber, bran, and germ intake and the risks of all-cause and cardiovascular disease-specific mortality among women with type 2 diabetes mellitus. Circulation 121:2162–2168

Heldt HW (2005) Phenylalanine ammonia lyase catalyzes the initial reaction of phenylpropanoid metabolism. Plant biochemistry. Elsevier, Amsterdam, pp 437–454

Heller W, Forkmann G (1994) Biosynthesis of flavonoids. In: Harborne JB (ed) The flavonoids, advances in research since 1986. Chapman and Hall, London, pp 499–535

Hellwege EM, Czapla S, Jahnke A, Willmitzer L, Heyer AG (2000) Transgenic potato (*Solanum tuberosum*) tubers synthesize the full spectrum of inulin molecules naturally occurring in globe artichoke (*Cynara scolymus*) roots. Proc Nat Acad Sci USA 97:8699–8704

Hellwege EM, Peeters R, Pilling J (2008) Long chain inulin. US Patent US 2008(0255249):A1

Henry RJ (1987) Pentosan and (1–3)(1–4)-beta-glucan concentrations in endosperm and whole grain of wheat, barley, oats and rye. J Cereal Sci 6:253–258

Hertog MG, Feskens EJ, Hollman PC et al (1993) Dietary antioxidant flavonoids and risk of coronary heart disease: the Zutphen elderly study. Lancet 342:1007–1011

Himi E, Noda K (2004) Isolation and location of three homoeologous dihydroflavonol-4-reductase (*DFR*) genes of wheat and their tissue-dependent expression. J Exp Bot 55:365–375

Himi E, Noda K (2005) Red grain colour gene (*R*) of wheat is a Myb-type transcription factor. Euphytica 143:239–242

Himi E, Maekawa M, Miura H, Noda K (2011) Development of PCR markers for *Tamyb*10 related to *R*-1, red grain color gene in wheat. Theor Appl Genet 122:1561–1576

Huynh B-L, Wallwork H, Stangoulis JCR, Graham RD, Willsmore KL, Olson S, Mather DE (2008) Quantitative trait loci for grain fructan concentration in wheat (*Triticum aestivum* L.). Theor Appl Genet 117:701–709

Institute of Medicine. Food and Nutrition Board (2001) Dietary reference intakes for Vitamin A, Vitamin K, Arsenic, Boron, Chromium, Copper, Iodine, Iron, Manganese, Molybdenum, Nickel, Silicon, Vanadium and Zinc. National Academy Press, Washington

International Barley Genome Sequencing Consortium (2012) A physical, genetic and functional sequence assembly of the barley genome. Nature 491:711–716

Ito T, Saito K, Sugawara M, Mochida K, Nakakuki T (1999) Effect of raw and heat-moisture-treated high-amylose corn starches on the process of digestion in the rat digestive tract. J Sci Food Agric 79:1203–1207

Itoh K, Ozaki H, Okada K, Hori H, Takeda Y, Mitsui T (2003) Introduction of Wx transgene into rice wx mutants leads to both high- and low-amylose rice. Plant Cell Physiol 44:473–480

Izydorczyk MS, Dexter JE (2008) Barley β-glucans and arabinoxylans: molecular structure, physicochemical properties, and uses in food products. Food Res Int 41:850–868

Jaskari J, Salovaara H, Mattilla-Sandholm T, Putanen K (1993) The effect of oat β-glucan on the growth of selected *Lactobacillus* spp. and *Bifidobacterium* spp. In: Aalto-Kaarlehto T, Salovaara H (ed.) Proceedings of the 25th Nordic Cereal Congress University of Helsinki, Helsinki, pp 242–244

Jenkins CLD, Lewis D, Bushell R, Belobrajdic DP, Bird AR (2011) Chain length of cereal fructans isolated from wheat stem and barley grain modulates in vitro fermentation. J Cereal Sci 53(2):188–191

Jones JM (2007) Mining whole grains for functional components. Food Sci Technol Bull Funct Foods 4:67–86

Kahlon TS, Keagy PM (2003) Functional foods: an overview. Cereal Foods World 48:112–115

Kawakami A, Yoshida M (2002) Molecular characterization of sucrose:sucrose 1-fructosyltransferase and sucrose:fructan 6-fructosyltransferase associated with fructan accumulation in winter wheat during cold hardening. Biosci Biotechnol Biochem 66:2297–2305

Kawakami A, Yoshida M (2005) Fructan:fructan 1-fructosyltransferase, a key enzyme for biosynthesis of graminan oligomers in hardened wheat. Planta 223:90–104

Keegstra K, Walton J (2006) β-Glucans- brewer's bane, dietician's delight. Science 311:1872–1873

Kenn DA, Dagg AHS, Stuart IM (1993) Effect of environment and genotype on the fermentability of malt produced from four Australian barley varieties. Am Soc Brew Chem 51:119–122

Kervinen T, Peltonen S, Teeri TH, Karjalainen R (1998) Differential expression of phenylalanine ammonia-lyase genes in barley induced by fungal infection or elicitors. New Phytol 139:293–300

Khlestkina EK, Röder MS, Pshenichnikova TA, Simonov AV, Salina EA, Börner A (2008a) Genes for anthocyanin pigmentation in wheat: review and microsatellite-based mapping. In: Verrity JF, Abbington LE (eds) Chromosome mapping research developments. Nova Science Publishers, New York, pp 155–175

Khlestkina EK, Röder MS, Salina EA (2008b) Relationship between homoeologous regulatory and structural genes in allopolyploid genome- a case study in bread wheat. BMC Plant Biol 8:88

Kirubuchi-Otobe C, Nagamine T, Yangisawa T, Ohnishi M, Yamaguchi I (1997) Production of hexaploid wheats with waxy endosperm character. Cereal Chem 74:72–74

Kleessen B, Sykura B, Zunft HJ, Blaut M (1997) Effects of inulin and lactose on faecal microfloa, microbial activity and bowel habit in elderly constipated persons. Am J Clin Nutr 65:1397–1402

Klepacka J, Fornal L (2006) Ferulic acid and its position among the phenolic compounds of wheat. Crit Rev Food Sci Nutr 46:639–647

Koes R, Verweij W, Quattrocchio F (2005) Flavonoids: a colorful model for the regulation and evolution of biochemical pathways. Trends Plant Sci 10:236–242

Kohlmeier L, Hastings SB (1995) Epidemiologic evidence of a role of carotenoids in cardiovascular disease prevention. Am J Clin Nutr 62:1370S–1376S

Li HP, Liao YC (2003) Isolation and characterization of two closely linked phenylalanine ammonia-lyase genes from wheat. Yi Chuan Xue Bao 30:907–912

Li WL, Faris JD, Chittoor JM, Leach JE, Hulbert SH, Liu DJ, Chen PD, Gill BS (1999) Genomic mapping of defense response genes in wheat. Theor Appl Genet 98:226–233

Li F, Murillo C, Wurtzel T (2007) Maize *Y9* encodes a product essential for 15-cis-ζ-carotene isomerizaion. Plant Physiol 144:1181–1189

Li L, Shewry PR, Ward JL (2008) Phenolic acids in wheat varieties in the HEALTHGRAIN diversity screen. J Agric Food Chem 56:9732–9739

Liu RH (2007) Whole grain phytochemicals and health. J Cereal Sci 46:207–219

Liyana-Pathirana CM, Shahidi F (2006) Importance of insoluble-bound phenolics to antioxidant properties of wheat. J Agric Food Chem 54:1256–1264

Lüscher M, Erdin C, Sprenger N, Hochstrasser U, Boller T, Wiemken A (1996) Inulin synthesis by a combination of purified fructosyltransferases from tubers of *Helianthus tuberosus*. FEBS Lett 385:39–42

Ma QH, Xu Y, Lin ZB, He P (2002) Cloning of cDNA encoding COMT from wheat which is differentially expressed in lodging-sensitive and -resistant cultivars. J Exp Bot 53:2281–2282

Manach C, Scalbert A, Morand C, Rémésy C, Jime′nez L (2004) Polyphenols: food sources and bioavailability. Am J Clin Nutrion 79:727–747

Manickavelu A, Kawaura K, Imamura H, Mori M, Ogihara Y (2011) Molecular mapping of quantitative trait loci for domestication traits and ß-glucan content in a wheat recombinant inbred line population. Euphytica 177:179–190

Mao J, Burt AJ, Ramputh AL-I, Simmonds J, Cass L, Hubbard K, Miller S, Altosaar I, Arnason JT (2007) Diverted secondary metabolism and improved resistance to European corn borer (Ostrinia nubilalis) in maize (Zea mays L.) transformed with wheat oxalate oxidase. J Agric Food Chem 55:2582–2589

Marrs KA, Alfenito MR, Lloyd AM, Walbot V (1995) A glutathione S-transferase involved in vacuolar transfer encoded by the maize gene *Bronze*-2. Nature 375:397–400

Martin C, Prescott A, Mackay S, Bartlett J, Vrijlandt E (1991) Control of anthocyanin biosynthesis in flowers of *Antirrhinum majus*. Plant J 1:37–49

Martin C, Jin H, Schwinn K (2001) Mechanisms and applications of transcriptional control of phenylpropanoid metabolism. In: Romeo J, Saunders J, Matthews B (eds) Regulation of phytochemicals by molecular techniques. Elsevier Science Ltd, Oxford, pp 155–170

Meldgaard M (1992) Expression of chalcone synthase, dihydroflavonol reductase, and flavanone-3-hydroxylase in mutants of barley deficient in anthocyanin and proanthocyanidin biosynthesis. Theor Appl Genet 83:695–706

Menga V, Fares C, Troccoli A, Cattivelli L, Baiano A (2010) Effects of genotype, location and baking on the phenolic content and some antioxidant properties of cereal species. Int J Food Sci Technol 45:7–16

Middleton E, Kandaswami C, Theoharides TC (2000) The effects of plant flavonoids on mammalian cells: implications for inflammation, heart disease, and cancer. Pharmacol Rev 52:673–751

Minami E, Tanaka Y (1993) Nucleotide sequence of the gene for phenylalanine ammonia-lyase of rice and its deduced amino acid sequence. Biochim Biophys Acta 1171:321–322

Mitchell RAC, Dupree P, Shewry PR (2007) A novel bioinformatics approach identifies candidate genes for the synthesis and feruloylation of Arabinoxylan. Plant Physiol 144:43–53

Mol J, Grotewold E, Koes R (1998) How genes paint flowers and seeds. Trends Plant Sci 3:212–217

Morell M, Kosar-Hashemi B, Samuel M, Chandler P, Rahman S, Buelon A, Batey I, Li Z (2003) Identification of the molecular basis of mutations at the barley *sex*6 locus and their novel starch phenotype. Plant J 34:172–184

Munkvold JD, Greene RA, Bermudez-Kandianis CE, La Rota CM, Edwards H, Sorrells SF, Dake T, Benscher D, Kantety R, Linkiewicz AM, Dubcovsky J, Akhunov ED, Dvorák J, Miftahudin, Gustafson JP, Pathan MS, Nguyen HT, Matthews DE, Chao S, Lazo GR, Hummel DD, Anderson OD, Anderson JA, Gonzalez-Hernandez JL, Peng JH, Lapitan N, Qi LL, Echalier B, Gill BS, Hossain KG, Kalavacharla V, Kianian SF, Sandhu D, Erayman M, Gill KS, McGuire PE, Qualset CO, Sorrells ME (2004) Group 3 chromosome bin maps of wheat and their relationship to rice chromosome 1. Genetics 168:639–650

Nagaraj VJ, Altenbach D, Galati V, Lüscher M, Meyer AD, Boller T, Wiemken A (2004) Distinct regulation of sucrose:sucrose-1-fructosyltransferase (1-SST) and

sucrose:fructan-6-fructosyltransferase (6-SFT), the key enzymes of fructan synthesis in barley leaves: 1-SST as the pacemaker. New Phytol 161:735–748
Nakamura T, Yamamori M, Hirano H, Hidaka S, Nagamine T (1995) Production of waxy (amylose-free) wheats. Mol Genet Genomics 248:253–259
Nemeth C, Freeman J, Jones HD, Sparks C, Pellny TK, Wilkinson MD, Dunwell J, Anderson AAM, Aman P, Guillon F, Saulnier L, Mitchell RAC, Shewry PR (2010) Down-regulation of the *CSLF*6 gene results in decreased (1,3;1,4)-β-D-glucan in endosperm of wheat. Plant Physiol 152:1209–1218
Neyrinck AM, Possemiers S, Druart C, Van de Wiele T, De Backer F et al (2011) Prebiotic effects of wheat arabinoxylan related to the increase in Bifidobacteria, Roseburia and Bacteroides/Prevotella in diet-induced obese mice. PLoS ONE 6(6):e20944. doi:10.1371/journal.pone.0020944
Neyrinck AM, Van Hèe VF, Piront N, De Backer F, Toussaint O, Cani PD, Delzenne NM (2012) Wheat-derived arabinoxylan oligosaccharides with prebiotic effect increase satietogenic gut peptides and reduce metabolic endotoxemia in diet-induced obese mice. Nutr Diabetes 2:e28. doi:10.1038/nutd.2011.24
Nishi A, Nakamura Y, Tanaka N, Satoh H (2001) Biochemical and genetic analysis of the effects of amylose-extender mutation in rice endosperm. Plant Physiol 127:459–472
Oikawa A, Joshi HJ, Rennie EA, Ebert B, Manisseri C et al (2010) An integrative approach to the identification of Arabidopsis and rice genes involved in xylan and secondary wall development. PLoS ONE 5(11):e15481
Padley FB, Gunstone FD, Harwood JL (1994) Major vegetable fats. In: Gunstone FD, Harwood JL, Padley FB (eds) The lipid handbook, 2nd edn. Chapman & Hall, London, p 130
Paine JA, Shipton CA, Chaggar S, Howells RM, Kenndey MJ, Vernon G, Wright S, Hinchliffe E, Adams JL, Silverstone AL, Drake R (2005) Improving the nutritional value of golden rice through increased pro-vitamin A content. Nat Biotechnol 23:482–487
Panfili G, Fratianni A, Irano M (2003) Normal phase high performance liquid chromatography method for the determination of tocopherols and tocotrienols in cereals. J Agric Food Chem 51:3940–3944
Panfili G, Fratianni A, Irano M (2004) Improved normal-phase high-performance liquid chromatography procedure for the determination of carotenoids in cereals. J Agric Food Chem 52:6373–6377
Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J et al (2009) The Sorghum bicolor genome and the diversification of grasses. Nature 457:551–556
Paz-Ares J, Wienand U, Peterson PA, Saedler H (1986) Molecular cloning of the *c* locus of *Zea mays*: a locus regulating the anthocyanin pathway. The EMBO J 5:829–833
Paz-Ares J, Ghosal D, Wienand U, Peterson PA, Saedler H (1987) The regulatory C1 locus of *Zea mays* encodes a protein with homology to MYB-related proto-oncogene products and with structural similarities to transcriptional activators. The EMBO J 6:3553–3558
Pennisi E (2009) Steak with a side of beta-glucans. Science 326:1058–1059
Peterson DM, Qureshi AA (1993) Genotype and environmental effects on tocols of barley and oats. Cereal Chem 70:157–162
Piironen V, Syväoja E-L, Varo P, Salminen K, Koivistoinen P (1986) Tocopherols and tocotrienols in cereal products from Finland. Cereal Chem 63:78–81
Pilu R, Piazza P, Petroni K, Ronchi A, Martin C, Tonelli C (2003) *pl-bol3*, a complex allele of the anthocyanin regulatory *pl1* locus that arose in a naturally occurring maize population. Plant J 36:510–521
Poulsen M, Molck AM, Jacobsen BL (2002) Different effects of short and long chained fructans on large intestinal physiology and carcinogen induced aberrant crypt foci in rats. Nutr Cancer 42:194–205
Pozniak CJ, Knox RE, Clarke FR, Clarke JM (2007) Identification of QTL and association of a phytoene synthase gene with endosperm colour in durum wheat. Theor Appl Genet 114:525–537
Prasad AS (1998) Zinc deficiency in humans: a neglected problem. J Am Coll Nutr 17(6):542–543

Quattrocchio F, Wing JF, Leppen HTC, Mol JNM, Koes RE (1993) Regulatory genes controlling anthocyanin pigmentation are functionally conserved among plant species and have distinct sets of target genes. Plant Cell 5:1497–1512

Rahman S, Bird A, Regina A, Li Z, Ral P, McMaugh S, Topping D, Morell M (2007) Resistant starch in cereals: exploiting genetic engineering and genetic variation. J Cereal Sci 46:251–260

Reddy BS, Hamid R, Rao CV (1997) Effect of dietary oligofructose and inulin on colonic preneoplastic aberrant crypt foci inhibition. Carcinogenesis 18:1371–1374

Regina A, Bird A, Topping D, Bowden S, Freeman J, Barsby T, Kosar-Hashemi B, Li Z, Rahman S, Morell M (2006) High-amylose wheat generated by RNA interference improves indices of large-bowel health in rats. Proc Nat Acad Sci USA 103:3546–3551

Regina A, Kosar-Hashemi B, Ling S, Li Z, Rahman S, Morell M (2010) Control of starch branching in barley defined through differential RNAi suppression of starch branching enzyme IIa and IIb. J Exp Bot 61:1469–1482

Rhodes MJC, Price KR (1997) Identification and analysis of plant phenolic antioxidants. Eur J Cancer Prev 6:518–521

Rigotti A (2007) Absorption, transport, and tissue delivery of vitamin E. Mol Aspects Med 28:423–436

Roberfroid M (2007) Prebiotics: the concept revisited. J Nutr 137:830S–837S

Ryoo N, Yu C, Park C-S, Baik M-Y, Park IM, Cho M-H, Bhoo SH, An G, Hahn T-R, Jeon J-S (2007) Knockout of a starch synthase gene *OsSSIIIa*/*Flo5* causes white-core floury endosperm in rice (*Oryza sativa* L.). Plant Cell Rep 26:1083–1095

Schneeman BO (1999) Fiber, inulin and oligofructose: similarities and differences. Nutritional and health benefits of inulin and oligofructose. J Nutr 129:1424S–1427S

Sen C, Khanna S, Roy S (2006) Tocotrienols: vitamin E beyond tocopherols. Life Sci 78:2088–2098

Serpen A, Kmen VG, Karago A, Hamit K (2008) Phytochemical quantification and total antioxidant capacities of Emmer (*Triticum dicoccon* Schrank) and Einkorn (*Triticum monococcum* L.) wheat landraces. J Agric Food Chem 56:7285–7292

Sestili F, Janni M, Doherty A, Botticella E, D'Ovidio R, Masci S, Jones HD, Lafiandra D (2010) Increasing the amylose content of durum wheat through silencing of the SBEIIa genes. BMC Plant Biol 10:14

Shannon JC, Garwood DL (1984) Genetics and physiology of starch development. In: Whistler RL, BeMiller JN, Paschall EF (eds) Starch: chemistry and technology, 2nd edn. Academic Press, Orlando, pp 26–86

Shelton DR, Lee WJ (2000) Cereal carbohydrates. In: Kulp K, Ponte JG (eds) Cereal science and technology. Marcel Dekker, USA, pp 385–414

Shewry PR (2008) Improving the nutritional quality of cereals by conventional and novel approaches. In: Hamaker BR (ed) Technology of functional cereal products, WoodHead Publishing in Food Science, Technology and Nutrition. Cambridge, England, pp 159 – 183

Shewry PR, Piironen V, Lampi A-M, Edelmann M, Kariluoto S, Nurmi T, Fernandez-Orozco R, Ravel C, Charmet G, Andersson AAM, Aman P, Boros D, Gebruers K, Dornez E, Courtin CM, Delcour JA, Rakszegi M, Bedo Z, Ward JL (2010) The HEALTHGRAIN wheat diversity screen: effects of genotype and environment on phytochemicals and dietary fiber components. J Agric Food Chem 58:9291–9298

Shin YM, Park HJ, Yim SD, Baek NI, Lee CH, An G, Woo YM (2006) Transgenic rice lines expressing maize *C1* and *R-S* regulatory genes produce various flavonoids in the endosperm. Plant Biotechnol J 4:303–315

Simopoulos AP (1991) Omega-3 fatty acids in health and disease and in growth and development. Am J Clin Nutr 54:438–463

Singh A, Reimer S, Pozniak CJ, Clarke FR, Clarke JM, Knox RE, Singh AK (2009) Allelic variation at *Psy1-A1* and association with yellow pigment in durum wheat grain. Theor Appl Genet 118:1539–1548

Slade AJ, Fuerstenberg SI, Loeffler D, Steine MN, Facciotti D (2005) A reverse genetic, nontransgenic approach to wheat crop improvement by TILLING. Nat Biotechnol 23:75–81

Snart J, Bibiloni R, Grayson T, Lay C, Zhang H, Allison GE, Laverdiere JK, Temelli F, Vasanthan T, Bell R, Tannock GW (2006) Supplementation of the diet with high-viscosity beta-glucan results in enrichment for lactobacilli in the rat cecum. Appl Environ Microbiol 72:1925–1931

Sosulski F, Krygier K, Hogge L (1982) Free, esterified, and insoluble-bound phenolic acids. 3. Composition of phenolic acids in cereal and potato flours. J Agric Food Chem 30:337–340

Sprenger N, Bortlik K, Brandt A, Boller T, Wiemken A (1995) Purification, cloning, and functional expression of sucrose:fructan 6-fructosyltransferase, a key enzyme of fructan synthesis in barley. Proc Nat Acad Sci USA 92:11652–11656

Stalikas CD (2007) Extraction, separation, and detection methods for phenolic acids and flavonoids. J Sep Sci 30:3268–3295

Stuart IM, Loi L, Fincher GB (1988) Varietal and environmental variations in (1 → 3, 1 → 4)-β-glucan levels and (1 → 3, 1 → 4)-β-glucanase potential in barley: Relationships to malting quality. J Cereal Sci 7:61–71

Tian Z, Qian Q, Liu Q, Yan M, Liu X, Yan C, Liu G, Gao Z, Tang S, Zeng D, Wang Y, Yu J, Gu J, Li J (2009) Allelic diversities in rice starch biosynthesis lead to a diverse array of rice eating and cooking qualities. Proc Nat Acad Sci USA 106:21760–21765

Topping DL, Clifton PM (2001) Short chain fatty acids and human colonic function—roles of resistant starch and non-starch polysaccharides. Physiol Rev 81:1031–1064

Urahara T, Tsuchiya K, Kotakw T, Tohno-oka T, Komae K, Kawada N et al (2004) A β(1 → 4)-xylosyltransferase involved in the synthesis of arabinoxylans in developing barley endosperms. Physiol Plant 122:169–180

Van den Ende W, Laere A (1996) De novo synthesis of fructans from sucrose in vitro by a combination of two purified enzymes (sucrose: sucrose 1-fructosyl transferase and fructan: fructan 1-fructosyl transferase) from chicory roots (*Cichorium intybus* L.). Planta 200:335–342

Van den Ende W, Clerens S, Vergauwen R, Riet LV, Laere AV, Yoshida M, Kawakami A (2003) Fructan 1-Exohydrolases. β-(2,1)-trimmers during graminan biosynthesis in stems of wheat: Purification, characterization, mass mapping and cloning of two Fructan 1-Exohydrolase isoforms. Plant Physiol 131:621–631

Van Riet L, Nagaraj V, Van den Ende W, Clerens S, Wiemken A, Van Laere A (2006) Purification, cloning and functional analysis of a fructan 6-exohydrolase from wheat (*Triticum aestivum* L.). J Exp Bot 57:213–223

Verbeke W, Scholderer J, Lähteenmäki L (2009) Consumer appeal of nutrition and health claims in three existing product concepts. Appetite 52:684–692

Vom Endt D, Kijne J, Memelink J (2002) Transcription factors controlling plant secondary metabolism: What regulates the regulators? Phytochemistry 61:107–114

Wang J, He X, He Z, Wang H, Xia X (2009) Cloning and phylogenetic analysis of phytoene synthase (*psy*1) genes in common wheat and related species. Hereditas 146:208–219

Wanner LA, Li G, Ware D, Somssich IE, Davis KR (1995) The phenylalanine ammonia-lyase gene family in *Arabidopsis thaliana*. Plant Mol Biol 27:327–338

Weaver LM, Herrmann KM (1997) Dynamics of the shikimate pathway in plants. Trends Plant Sci 2:346–351

Weedon BCL, Moss GP (1995) Structure and nomenclature. In: Britton G, Pfander H, Liaaen-Jensen S (eds) Carotenoids. Spectroscopy, vol 1B. Birkhauser, Verlag, Basel, pp 27–44

Whistance GR, Threlfall DR (1970) Biosynthesis of phytoquinones; homogentisic acid: a precursor of plastoquinones, tocopherols and alpha-tocopherolquinone in higher plants, green algae and blue-green algae. Biochem J 117:593–600

Wienand U, Weydemann U, Niesbach-Klösgen U, Peterson PA, Saedler H (1986) Molecular cloning of the *c*2 locus of *Zea Mays*, the gene coding for chalcone synthase. Mol Gen Genet 203:202–207

Wong JC, Lambert RJ, Wurtzel ET, Rocheford TR (2004) QTL and candidiate genes phytoene synthase and zeta-carotene desaturase associated with accumulation of carotenoids in maize. Theor Appl Genet 108:349–359

Yamamori M, Fujita S, Hayakawa K, Matsuki J, Yasui T (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

Yamamori M, Kato M, Yui M, Kawasaki M (2006) Resistant starch and starch pasting properties of a starch synthase IIa-deficient wheat with apparent high amylase. Aust J Agric Res 57:531–535

Yan J, Kandianis CB, Harjes CE, Bai L, Kim E-H, Yang X, Skinner DJ, Fu Z, Mitchell S, Li Q, Fernandez MGS, Zaharieval Babul R, Fu Y, Palacios N, Li J, DellaPenna D, Brutnell T, Buckler ES, Warburton ML, Rocheford T (2010) Rare genetic variation at *Zea mays crtRB*1 increases β-carotene in maize grain. Nat Genet 42:322–328

Yang DH, Yeoup CB, Kim JS, Kim JH, Yun PY, Lee YK, Lim YP, Lee MC (2005) cDNA cloning and sequence analysis of the rice cinnamate-4-hydroxylase gene, acytochrome P450-dependent monooxygenase involved in the general phenylpropanoid pathway. J Plant Biol 48:311–318

Ye X, Al-Babili S, Klöti A, Zhang J, Lucca P, Beyer P, Potrykus I (2000) Engineering the provitamin A (beta-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. Science 287:303–305

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 X, Colleoni C, Ratushna V, Sirghie-Colleoni M, James MG, Myers AM (2004) Molecular characterization demonstrates that the *Zea mays* gene *sugary*2 codes for the starch synthase isoform SSIIa. Plant Mol Biol 54:865–879

Zhang W, Lukaszewski AJ, Kolmer J, Soria MA, Goyal S, Dubcovsky J (2005) Molecular characterization of durum and common wheat recombinant lines carrying leaf rust resistance (*Lr*19) and yellow pigment (*Y*) genes from *Lophopyrum ponticum*. Theor Appl Genet 111:573–582

Zhang J, Huang S, Fosu-Nyarko J, Dell B, McNeil M, Waters I, Moolhuijzen P, Conocono E, Appels R (2008) The genome structure of the 1-*FEH* genes in wheat (*Triticum aestivum* L.): new markers to track stem carbohydrates and grain filling QTLs in breeding. Mol Breeding 22:339–351

Zhou JM, Lee E, Kanapathy-Sinnaiaha F, Park Y, Kornblatt JA, Lim Y, Ibrahim RK (2010) Structure-function relationships of wheat flavones O-methyltransferase: homology modeling and site-directed mutagenesis. BMC Plant Biol 10:156

Zhu Q, Dabi T, Beeche A, Yamamoto R, Lawton MA, Lamb C (1995) Cloning and properties of a rice gene encoding phenylalanine ammonia-lyase. Plant Mol Biol 29:535–550

Zhu C, Naqvi S, Breitenbach J, Sandmann G, Cristou P, Capell T (2008) Combinatorial genetic transformation generates a library of metabolic phenotypes for the carotenoid pathway in maize. Proc Nat Acad USA 105: 18232–18237

Chapter 11
QTL Mapping: Methodology and Applications in Cereal Breeding

Pushpendra K. Gupta, Pawan L. Kulwal and Reyazul R. Mir

11.1 Introduction

Quantitative trait loci (QTL) mapping in crop plants has now become routine due to the progress made in this area during the last two decades. Although, initial QTL studies mainly focused on the identification of QTLs for only some important quantitative traits (QTs) in any individual crop, QTLs could later be identified for majority of the QTs in each of a number of crops, in many cases leading to cloning of individual QTLs. Consequently, in different crops, more than 10,000 QTLs/genes (also described as marker-trait associations; MTAs) were already detected by 2008 (over 5,000 QTL have been detected in rice alone; http://www.gramene.org/qtl/), suggesting enormous research activity in this area in the recent past (Bernardo 2008), although not many of these studies could be utilized for crop improvement. It is also apparent that majority of these studies have been conducted in cereals, primarily because of their relative importance in our food security system. Advances in the techniques of QTL mapping were also facilitated by developments that took place in the field of genomics research including statistical genomics (Liu 1998; Mauricio 2001; Borevitz and Chory 2004). Keeping in view the importance of the subject and the enormous research activity

P. K. Gupta (✉)
Molecular Biology Laboratory, Department of Genetics and Plant Breeding, CCS University, Meerut 250004, India
e-mail: pkgupta36@gmail.com

P. L. Kulwal
State Level Biotechnology Centre, Mahatma Phule Agricultural University, Rahuri, Ahmednagar 413722, Maharashtra, India

R. R. Mir
Division of Plant Breeding and Genetics, Shere-Kashmir University of Agricultural Sciences and Technology of Jammu (SKUAST-J), Chatha, 180009 Jammu, Jammu and Kashmir, India

P. K. Gupta and R. K. Varshney (eds.), *Cereal Genomics II*,
DOI: 10.1007/978-94-007-6401-9_11,

witnessed in this area of research, several reviews on QTL analysis in plants appeared in the past (Tanksley 1993; Mackay 2001; Doerge 2002; Hackett 2002; Collard et al. 2005; Jansen 2007; Nordborg and Weigel 2008; Moose and Mumm 2008; Zhang and Gai 2009; Myles et al. 2009; Van Eeuwijk et al. 2010).

In the first edition of edited volume entitled 'Cereal Genomics', two chapters were devoted to QTL mapping; one was focused on QTLs and genes for disease resistance in barley and wheat (Jahoor et al. 2004), while the other was focused on QTLs and genes for tolerance to abiotic stress in cereals (Tuberosa and Salvi 2004). After the publication of this edited volume, significant advances have been made during the last 8 years, so that in this chapter, we give an updated version of the various resources (including mapping populations), methods and their applications in cereals, which became available for linkage-based simple QTL mapping (QM), joint linkage mapping, linkage disequilibrium (LD)-based association mapping (AM) and joint-linkage association mapping (JLAM) in cereals (Fig. 11.1). Genetical genomics along with eQTLs (expression QTLs), pQTLs (protein QTLs) and mQTLs (metabolite QTLs) have also been discussed briefly. While discussing linkage-based QM, we also discuss the desirability of using either joint linkage analysis or meta-QTL analysis (involving several mapping populations together) to emphasize the importance of QTL studies for the same trait in a crop conducted using a number of biparental mapping populations. Two other areas of research which are discussed briefly in this review include QTL studies undertaken to study quantitative disease resistance and genetics of domestication syndrome in cereals.

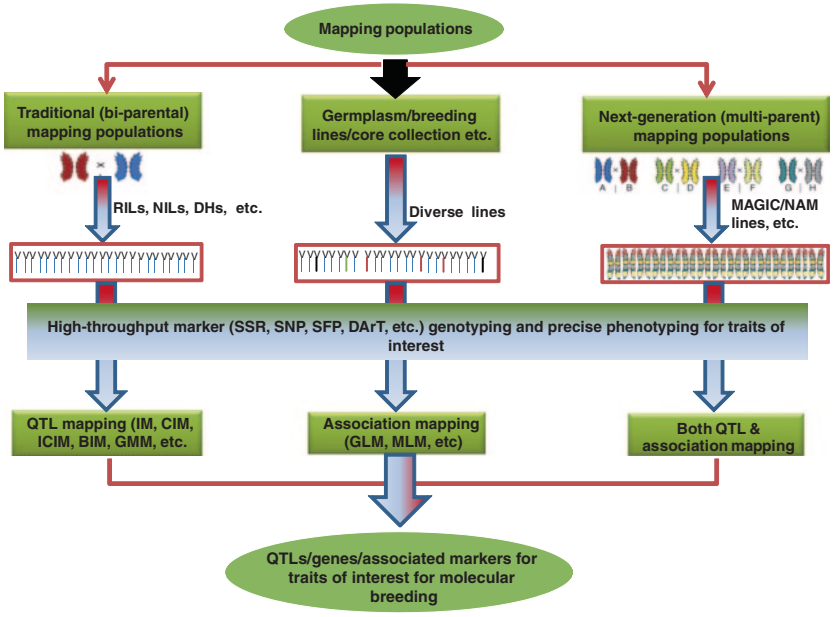

Fig. 11.1 Systematic representation of various steps involved in QTL identification

Finally, towards the end of this chapter, we summarize the results available on cloning of QTL in cereals. However, keeping in view the enormous literature published on the subject, we find it neither possible nor desirable to collect and present all the available literature in this article.

11.2 QTL Mapping Based on Linkage Analysis

11.2.1 Mapping Populations for QTL Analysis in Cereals

The importance of mapping populations for QTL analysis is widely known. Generally individual mapping populations for QTL analysis are developed using two or more parental genotypes, which differ for the trait of interest. In some other cases, mapping populations are developed primarily for construction of linkage maps, but are subsequently also utilized for QTL analysis, so that the parents of such a mapping population are diverse, but were not selected keeping in mind any trait of interest. International Triticeae Mapping Population ($ITMI_{pop}$) in wheat is one such example, which was developed for construction of genetic maps, but was subsequently used for many QTL studies, since parental genotypes of ITMI were chosen to be diverse and consequently the recombinant inbred lines (RILs) segregated for a number of traits of economic importance, thus permitting QTL analysis for these traits (Röder et al. 1998; Borner et al. 2002; Kulwal et al. 2003, 2004). In some of these studies, it was observed that significant variation for a trait of interest was available among RILs, even when the parents did not differ for the trait. This may happen when the parents differ genetically for the trait of interest, but may not differ phenotypically for this trait, so that similar phenotypes result due to the presence of different sets of QTL for the same trait in the two parents. In our own studies, such a situation was observed, when we utilized ITMI population for QTL analysis for pre-harvest sprouting tolerance (PHST) (Kulwal et al. 2004). However, during the last couple of decades, use of different types of mapping populations including biparentals, germplasm and breeding populations has been suggested, each having its own advantages and limitations in terms of their use in QTL mapping (see following sections and Fig. 11.2), suggesting that the choice of material also depends upon the objective of the study.

Bi-Parental Mapping Populations

F_2, BC, RIL and DH Populations

Biparental mapping populations that are used for QTL analysis include F_2, backcross (BC_1), RIL and doubled haploid (DH) populations, although initially F_2 populations

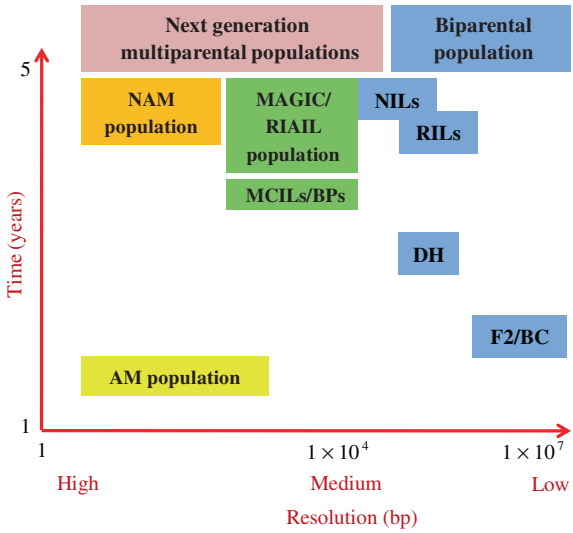

Fig. 11.2 Hypothetical figure showing comparison of various mapping populations used in QTL mapping in terms of resolution and research time (modified after Yu and Buckler 2006)

that were actually developed with the aim of constructing linkage maps, were also utilized for QTL mapping (Fig. 11.1). Some of the earlier high density maps in rice (Harushima et al. 1998), barley (Graner et al. 1991), maize (Gardiner et al. 1993), wheat (Gill et al. 1991) and sorghum (Whitkus et al. 1992) were also based on F_2 mapping populations, which were also used for QTL analysis. Since a F_2 population is the product of only one cycle of recombination and RILs result from several such cycles of recombination, a linkage map prepared using F_2 population is not as fine as that obtained using RILs. This problem can be overcome by using sufficiently large F_2 population (Hackett 2002; Collard et al. 2005), but this incurs substantial costs, so that aletrnative strategies have, therefore, been developed through intermating F_2 individuals for one or more generations, sometimes leading to the development of of the so-called advanced intermated lines, the AILs, as done in maize (Darvasi and Soller 1995). Moreover, since a linkage map is often used for QTL mapping, and since an F_2 population cannot be replicated, it offers limited utility in terms of studying the genetic architecture of the trait. Thus, majority of the linkage and QTL mapping studies carried out in cereals made use of RIL populations, mainly due to higher mapping resolution and due to the possibility, which they offer for the study of genotype × environment interactions. However, the time required for the development of RILs often discourages researchers from developing these mapping populations.

In view of the time needed for developing RILs, DH populations are often preferred, since they still retain some of the advantages of RILs and can be produced within a reasonable period of time (Gupta et al. 2010a). Techniques for production of DH populations have been standardized in most cereals and these populations are now being increasingly used for QTL analysis. This has been successfully achieved in almost all cereals including wheat (Cadalen et al. 1998; Perretant et al. 2000), barley (Graner et al. 1991; Kleinhofs et al. 1993; Langridge et al. 1995),

rice (Lu et al. 1996; Xu et al. 1997) and oats (Tanhuanpää et al. 2008, 2010). Thus, both, RILs and DH populations represent eternal resources for QTL mapping and therefore often described as 'immortal populations', although immortalized F_2 populations have also been developed and utilized in crops like rice and maize, with the sole objective of studying dominant gene effects for heterosis breeding (Hua et al. 2003; Tang et al. 2010 see later).

Advanced Generation Populations

Advanced generations following a cross between two diverse parents may also give suitable populations for QTL analysis. One such example is advanced backcross population, which can be utilized for simultaneous QTL analysis and marker assisted selection (Tanksley and Nelson 1996; for more details on the use of these populations, consult Advanced Backcross QTL Analysis, discussed later in this chapter).

Immortalized F_2 Populations and NC III Design Populations

One of the limitations of using eternal RIL/DH mapping populations is that these populations cannot be used for studying the dominant genetic effects of QTLs, which is necessary when heterosis is exploited for enhanced crop productivity. We know that F_2 or backcross (BC) populations, which can be used for the study of dominant genetic effects, cannot be subjected to multi-location trials, nor can the trials be repeated over years to study QTL × environment interactions. Therefore, efforts have been made to develop immortalized F_2 (IF_2) populations both in self-pollinated and cross-pollinated crops (rice and maize), where heterosis is being exploited (Hua et al. 2003; Tang et al. 2010). These immortalized F_2 populations are derived from RILs/DH lines of one or more mapping populations by intercrossing RILs/DH lines in pairs, so that a large number of F_1 hybrids obtained from hundreds of RIL/DH pairs (without repetition) will mimic a F_2 population. Such an IF_2 population can be repeatedly produced again and again depending upon the need and can therefore be used for multi-location trials over years, thus providing an opportunity for detection of QTLs involved in dominant, dominant × dominant and dominant × additive effects. These populations can also be used for a study of the mechanism of heterosis in a particular crop.

In maize, North Carolina III (NCIII) design has also been used for developing populations that are suitable for study of dominant effects. For instance, in a recent study, crosses of hybrids between RILs (belonging to three mapping populations), each with both the corresponding parental inbred lines, have been used for developing populations of F_1's in maize which are comparable to immortalized F_2 populations utilized in rice (Larièpe et al. 2012). In this study, a high correlation between heterosis and heterozygosity of Hybrids for a number of markers confirmed the complex genetic basis and the role of dominance in heterosis.

Next-Generation Multi-Parental Mapping Populations

Next-generation mapping populations include a variety of mapping populations, which should overcome several limitations associated with traditional bi-parental and association mapping populations (Morrell et al. 2012; Fig. 11.1). Some of the advantages of these populations over bi-parental and association mapping populations include the following: (1) increase in the rate of effective recombinations per generation, (2) higher level of genetic variability, (3) higher resolution, (4) effective sampling of rare alleles, (5) no problem of population structure and (6) better estimation of allelic effects. These advantages associated with the next generation multiparental mapping populations have been attributed to the involvement of multiple parents in both controlled crosses (as in QTL mapping) or inter-mating over a number of generations (as in association mapping). 'Multiparent Advanced Generation Intercross' (MAGIC) populations, 'Nested Association Mapping' (NAM) populations, Multiline Cross Inbred Lines (MCILs) and Recombinant Inbred Advanced Intercross Lines (RIAILs) are the most important examples of these next generation multi-parental populations that are being developed in cereals. Similarly, in case of *Arabidopsis*, eight founder accessions were crossed to produce six sets of recombinant inbred lines, which together make an *Arabidopsis* multiparent RIL (AMPRIL) population (Huang et al. 2011). In a recent study, these multiparental mapping populations (e.g., MAGIC and AMPRIL) were compared with biparental mapping populations for their use in QTL mapping (Rakshit et al. 2012).

'Multiparent Advanced Generation Intercross' (MAGIC) Populations

In case of a MAGIC population, multiple parents (involving 4–20 parents) are used for the development of a mapping population (Cavanagh et al. 2008; Lehmensiek et al. 2009). One such MAGIC population in model plant system Arabidopsis consisted of 527 RILs that were developed using a heterogeneous stock derived from 19 intermated accessions; this population helped in mapping of a number of already known QTLs with high precision (Kover et al. 2009). MAGIC populations are also being developed in a number of cereal crops including wheat (for details, see Gupta et al. 2010b), rice and sorghum (CGIAR Generation Challenge Programme 2009). At the International Rice Research Institute (IRRI), Philippines, efforts are underway to develop a number of rice MAGIC populations based on sixteen diverse founder lines involving eight each from the *indica* and *japonica* eco-geographic races of *Oryza sativa* (Bandillo et al. 2011). Although, these populations offer a definite advantage over any of the other types of populations, one needs to subgroup these populations based on their adaptive traits (like plant height, days to maturity) before using them for phenotyping for traits like drought and heat stress.

'Nested Association Mapping' Populations

A NAM population was developed in maize by crossing one common parent B73 with 25 diverse maize founder inbreds followed by selfing the F_1's

to generate 25 F_2 populations. These F_2 populations were advanced through single seed descent (SSD) to generate 25 RIL populations each with 200 RILs, thus making a NAM population of 5,000 RILs, which captured an estimated number of 136,000 recombination events. This allowed the use of both linkage analysis and association mapping in a single, unified mapping population, since it takes into account both historic and recent recombination events (Yu et al. 2008). Besides this, NAM population also offers the advantage of high allele richness coupled with high resolution and high statistical power, without the associated disadvantages of either linkage mapping or association mapping (McMullen et al. 2009). Since its development, this NAM population has been used in several linkage and association mapping studies (see association mapping section), involving genetic dissection of flowering time, leaf architecture, northern and southern leaf blights in maize (Buckler et al. 2009; Kump et al. 2011; Poland et al. 2011; Tian et al. 2011). Using the data recorded on about a million plants, these studies identified a large number of small effect QTLs and also the epistatic and/or environmental interactions for each of the trait studied.

Multiline Cross Inbred Lines

The MAGIC and NAM populations discussed above are specially designed for the purpose of QTL mapping. However, multiparental inbred lines are also generated during regular routine plant breeding programs, although not in a manner like MAGIC or NAM. These multiline cross inbred lines and/or breeding populations can also be utilized for QTL mapping, because these are available without any extra cost and effort and can have direct relevance to crop improvement (Wurschum 2012). The advantage of using such material is that the mating is controlled by the breeder, so that the ancestry is known and the phenotypic data is available for free.

Unfortunately, most of the methods and computer programs of QTL mapping have been designed for simple-line crosses or multi-line crosses that are based on a regular mating system. These methods, therefore, are not suitable for multi-line cross inbred lines produced through routine and complicated mating designs of commercial plant breeding. Newer methods have, therefore, been proposed for conducting QTL analysis using these multi-line cross inbred lines (Xie et al. 1998; Yi and Xu 2002). Software like MCQTL and INTERQTL have also been developed for this purpose (Crepieux et al. 2004; Jourjon et al. 2005). While using multiline cross inbred lines (including breeding populations), one needs to ensure that some common checks were included in the trials so as to make the phenotypic data balanced one, as all the lines are not evaluated every year (for more details see Wurschum 2012). In maize, using 404 multi-cross inbred lines, eight QTL were mapped for a male flowering trait, described as ‘growing degree day heat units to pollen shedding’ (GDUSHD). These QTLs contributed 80% of the variance observed among the inbred lines (Zhang et al. 2005).

Recombinant Inbred Advanced Intercross Lines

As mentioned above, like MAGIC populations, Recombinant Inbred Advanced Intercross Lines (RIAILs) represent a new class of next generation multiparental mapping populations (Rockman and Kruglyak 2008). Some other similar advanced intercross populations have been described as intermated recombinant inbred populations (IRIP) or intermated recombinant inbred lines (IRIL) (see Liu et al. 1996; Rockman and Kruglyak 2008). Like MAGIC populations, RIAILs also involve intercrossing of multiple parental genotypes and form a single population with large number of RIAILs (see Morrell et al. 2012). The use of multiple parents and advanced intercrosses allows increase in the number of recombination breakpoints in these RIAILs (Rockman and Kruglyak 2008). Therefore, RIAIL mapping populations have the merits of both AILs and RILs, and among plant systems, have already been produced and utilized for QTL mapping in Arabidopsis (Liu et al. 1996) and maize (Lee et al. 2002).

11.2.2 Methods of Linkage-Based QTL Mapping

A number of linkage-based approaches have been used for QTL mapping in cereals. These approaches include single marker analysis (SMA) and QTL interval mapping, the latter in its turn including single-locus and two-locus analyses. The most common and simple approach of mapping genes makes use of linear regression and was used as early as 1920s (Sax 1923). However, due to the availability of molecular markers and newer statistical tools, significant improvement in methods used for QTL mapping has been possible in recent years. Also, the focus shifted from marker analysis to simple, composite and multiple interval mapping (Lander et al. 1987; Lander and Botstein 1989; Zeng 1994; Kao et al. 1999). Emphasis has also shifted from single-locus analysis to two-locus analysis involving epistatic QTLs (with and without main effects) and from maximum likelihood approach to the use of Bayesian approach in QTL mapping in plants. These approaches of QTL mapping are discussed in brief in the following sections.

Bulk Segregant Analysis

In the initial years, when marker resources were only being developed and linkage mapping was time-consuming, marker-trait associations were generally determined using bulk segregant analysis (BSA) (Michelmore et al. 1991), which is still considered as a rapid approach for detecting linkage of a marker with a QTL for a trait of interest. Several significant marker-trait associations (MTAs) that were detected in wheat using BSA were later confirmed through QTL interval mapping. For example, in wheat, through BSA, Prasad et al. (1999) found a microsatellite marker mapped on chromosome 2DL associated with grain protein content, which

was later confirmed through simple interval mapping (SIM) and composite interval mapping (CIM) based on the genetic map developed for the concerned mapping population (Prasad et al. 2003). Similarly, the marker associated with grain weight identified by Varshney et al. (2000) through BSA was confirmed by Kumar et al. (2006). In maize, BSA has been used for the mapping of QTL for drought resistance (Quarrie et al. 1999). Recently, BSA also led to the identification of two QTLs for shoot fly tolerance loci in sorghum (Apotikar et al. 2011), fine mapping of a QTL for drought resistance in rice (Salunkhe et al. 2011), and an eQTL for leaf rust resistance in barley (Chen et al. 2010).

Other Linkage-Based Approaches for QTL Mapping

During 1990s and later, with the availability of newer molecular markers in all cereal crops (SSRs, AFLPs, SNPs, DArT), high density linkage maps were developed (see databases Graingenes, Panzea and Gramene). Framework maps using mapping populations were also developed for a number of cereal crops and used for QTL interval mapping. Both, SIM and CIM were conducted using a variety of software's like MAPMAKER/QTL and QTL Cartographer (Lander and Botstein 1989; Zeng 1994). The shortcomings of SIM and the improvements provided by CIM have been discussed in several articles (Doerge 2002; Mackay 2001; Mauricio 2001; Hackett 2002; Gupta and Kulwal 2006). Later, multiple interval mapping (MIM; Kao et al. 1999) and inclusive composite interval mapping (ICIM; Li et al. 2007) were also developed, thus leading to successive improvement in the dissection of genetic system for a variety of traits in all crops including cereals.

In most linkage-based approaches, generally maximum likelihood (ML) approach involving Expectation-Maximization (EM) algorithm is used to find out the likelihood of the observed trait distribution with and without QTL effect. The likelihood ratio (LR) and/or LOD scores estimated through ML approach are used for finding out the presence/absence of a QTL at a specified position in an interval (Lander and Botstein 1989). A method for determining the threshold value of LOD score for the given set of data was also suggested (Churchill and Doerge 1994; Doerge and Churchill 1996) in order to avoid detecting false QTLs. However, this EM approach was initially considered to be computationally demanding, so that regression approach for interval mapping was also developed (Haley and Knott 1992; Martinez and Curnow 1992). However, with improved computation facility available everywhere now, this seems to be no longer an issue.

As mentioned above, the composite interval mapping (CIM) approach proposed by Zeng (1994) has been the method of choice for QTL mapping involving bi-parental crosses. However, one of the drawbacks associated with this approach is that the results are influenced by the choice of cofactors (background markers) used and there is no set rule to select for the cofactors. The method of inclusive composite interval mapping (ICIM) proposed by Li et al. (2007) takes into account the significant cofactors and calculates their effects using stepwise regression. As this is done before the interval mapping is conducted and as the effects are fixed during the genome

scanning, it eliminates the arbitrariness of cofactor selection. Through simulation, Li et al. (2007) compared CIM using QTL Cartographer with that of ICIM using their own software and observed that the average LOD profiles of ICIM showed increased and clear peaks around most of the pre-defined QTL, as against those detected by CIM especially on chromosomes with multiple QTLs. Similar results were also obtained by Li et al. (2010), when they compared interval mapping (IM) with ICIM. The utility of this method was also demonstrated for mapping digenic epistatic QTLs (Li et al. 2008), as has been done for flowering time in maize (Buckler et al. 2009). A comparison of some of the methods of QTL mapping is presented in Table 11.1.

Interacting Epistatic QTLs

In the initial QTL interval mapping studies, QTL × QTL and QTL × environment (QE) interactions were not examined, although these interactions are known to be significant for majority of quantitative traits in all major crops. The approaches like multiple interval mapping (Kao et al. 1999; Zeng et al. 1999) and several other approaches addressed this issue, but in most of these cases, we could examine interactions among only those QTL, which had their own main effects (M-QTLs) and were detected through CIM. However, there can be QTLs, which have no main effects and influence a trait only through epistatic interactions. These QTLs are specifically described as epistatic QTL or E-QTL. QTLNetwork developed by Jun Zhu at Zhejiang University (Hangzhou, China) has the ability to detect these E-QTLs and the associated interactions. Several studies involving detection of E-QTLs have now been conducted in cereals including those in rice (Li et al. 1997, 2003; Fan et al. 2005), wheat (Kulwal et al. 2004, 2005; Kumar et al. 2007; Mohan et al. 2009), barley (Xu and Jia 2007), and maize (Ma et al. 2007). It is thus obvious that no molecular breeding program can ignore the occurrence of these interacting QTLs.

Multi-Trait Mapping

In nature, often more than two traits are correlated and the ability to detect a common QTL for more than one trait (pleiotropic QTL) can accelerate marker-assisted selection program, if the traits are positively correlated. In majority of the QTL mapping studies, several traits are analyzed separately, although some of these traits are correlated. Advantages of multi-trait analysis in the detection of linked or pleiotropic QTLs have long been advocated (Korol et al. 1995, 1998). It was also shown empirically and by simulation studies that use of correlation information can increase the power and precision of QTL detection. In barley, Korol et al. (1998) demonstrated the usefulness of multi-trait mapping in increasing the power of QTL detection using an example of two correlated traits (α-amylase and malt extract). Some QTL mapping programs like MultiQTL have also been developed for this purpose. Provisions for multiple trait mapping have also been made in QTL Cartographer and other software.

Table 11.1 Comparison between different methods of QTL mapping

Particular	BSA/SMA	SIM	CIM	MIM	BIM	AM
Linkage map	Not required	Required	Required	Required	Required	Not required
Exact position of QTL can be find out	No	Yes	Yes	Yes	Yes	No/Yes (depends on availability of map)
Precision of QTL effect	No	Not as precise as that of CIM	Precise	Precise	Precise	Precise
Interaction effect of QTL detected	No	No	Yes	Yes	Yes	Yes
Marker information utilized during analysis	Single marker	Flanking marker	Whole genome	Whole genome	Whole genome	Whole genome
Cofactors taken into account	No	No	Yes	Yes	Yes	Yes
Computationally intensive	No	No	Yes	Yes	Yes	Yes
Chances of false positives	High	High	Less	Less	Least	Low/high (depending on criteria used for significance)

A number of multi-trait QTL studies have also been conducted in wheat and sorghum. For example, in wheat, Kulwal et al. (2003) used multi-trait composite interval mapping (MCIM) for detection of a common QTL on chromosome 2D, which controlled three correlated traits including early growth habit, days to heading and days to maturity. Similarly, Kumar et al. (2007) used this approach and detected common QTLs on different chromosomes for yield and yield contributing traits in two mapping populations of wheat. Recently in sorghum, Apotikar et al. (2011) used MCIM and detected two pleiotropic QTLs for shoot fly resistance component traits. In all these studies, it is interesting to note that MCIM also detected a common QTL for correlated traits, which was not detected for the individual traits following CIM. However, caution should be exercised to know whether these correlations are due to pleiotropy or due to the presence of closely linked loci in the same genomic region. More studies involving high density maps are needed in future to address this concern.

Bayesian Approach for QTL Mapping

Bayesian approach of QTL mapping has some advantages over the so-called frequentist approaches described above. In this method, prior information is incorporated in a very specific manner and is combined with information from the observed data to generate the posterior distribution over the parameter values according to Bayes' rule. Besides this, it also offers straightforward interpretation of the results (for reviews, see Shoemaker et al. 1999; Beaumont and Rannala 2004). In the past, due to its computationally intensive nature, its application was rather restricted and used mostly in animal systems. However, with the advancements in the computation power, this is no longer an issue and the use of Bayesian approach is widely gaining momentum in almost every field of genetics including plant breeding. In one of its early application in QTL mapping in crop plants, Satagopan et al. (1996) used this approach to detect QTLs using Markov Chain Monte Carlo (MCMC) for estimating the locations and effect parameters for multiple QTLs with pre-specified number of QTLs in a DH progeny of *Brassica napus*. Later, with the increasing popularity of this approach, new methods/models were proposed. Using barley as an example, Xu (2003a) proposed a Bayesian regression method to estimate simultaneously the genetic effects associated with markers of the entire genome. Similarly, Yi et al. (2003a, b) gave a search strategy for mapping interacting QTLs using Bayesian approach.

In order to take into account the effects of multiple QTLs and marker × environment interactions, Bauer et al. (2009) used a Bayesian multi-locus multi-environmental method of QTL mapping in barley and compared their results with those of restricted maximum likelihood (REML) single-locus method; they observed that Bayesian approach was computationally more demanding than the REML method. Wang et al. (2009) proposed a Bayesian multiple interval method for mapping QTL underlying endosperm traits in cereals and observed that Bayesian method proposed by them estimates multiple QTLs and their effects, and distinguishes the two dominant effects of endosperm QTL. Recently, Sharma et

al. (2011) used empirical Bayes method in wheat for root traits to estimate additive and epistatic effects for all possible marker pairs simultaneously in a single model in order to minimize the error variance and to detect interaction effects between loci with no main effect. Besides its use in QTL interval mapping involving biparental populations, recently its application in pedigreed populations from the ongoing breeding programs has also been demonstrated (Bink et al. 2008).

One of the few limitations of Bayesian approach often pointed out is that the choice of the prior distribution in Bayesian framework is too subjective and that two researchers using the same data could reach different conclusions, if they use different priors (Shoemaker et al. 1999). Disregarding its limitations, there is no doubt that in future Bayesian methods will bring new insights into the genetics of complex traits.

Genetical Genomics and Expression QTLs (eQTLs)

A large numbers of QTL studies involving phenotypic traits have been conducted in plants; sometimes, these QTLs are referred to as PhQTL (Jansen et al. 2009). However, in the age of genomics, one would like to use genomics data for QTL analysis. An approach of expression QTL mapping (eQTL) involving use of microarray also called as genetical genomics is one such powerful approach (Jansen and Nap 2001). In this approach, expression profile of each gene in a mapping population is used as a quantitative trait, so that for each gene (cDNA) or gene product analyzed in the segregating population, QTL analysis can pinpoint the region (eQTL) of the genome involved in its expression either in *cis* or in *trans* orientation with respect to the gene involved (for a review, see Hansen et al. 2008).

The eQTL studies have been carried out in a number of cereals. The success of this technique was first shown in maize by Schadt et al. (2003). In wheat, using a rather small DH population (39 DH lines), Jordan et al. (2007) observed the presence of *cis*—and *trans*-acting eQTLs controlling seed development. In barley, Potokina et al. (2008a, b) reported genome-wide occurrence of limited pleiotropy of *cis*-regulatory mutations. Chen et al. (2010) used this approach to study the partial resistance to barley leaf rust and identified strong candidate genes underlying phenotypic QTL for resistance to leaf rust. However in maize, using IBM population and the hybrids between the individual RILs and both the parents, Swanson-Wagner et al. (2009) found that over three-quarters of ~4000 eQTL, which they detected are *trans* acting. The eQTL analysis thus would certainly help in understanding the genetics and the biosynthetic pathways of complex traits at the molecular level. Although, it is resource and cost-intensive approach, it is going to be increasingly used in future.

Protein Quantity Loci (pQTLs)

Regulatory genes which are responsible for the occurrence of variability in a specific protein are described as protein quantity loci (pQLs; Damerval et al. 1994). The variation in protein abundance can be used as a molecular phenotype and QTL

mapping can be performed, which became possible due to several technological advances (Jansen et al. 2009). In actual practice, pQLs were detected for drought tolerance in maize (de Vienne et al. 1999), and for metabolism and disease/defense-related processes in barley (Katja et al. 2011). More such studies will certainly help in understanding the molecular mechanism underlying complex traits.

Metabolite QTL (mQTL)

Like proteins, metabolites can also be mapped through QTL analysis using metabolite profiles in mapping population of an individual species. This can be done by measuring the abundance of a specific metabolite in all the lines of a mapping population and using it as a phenotypic trait. Such a quantitative genetic analysis has been used for molecular dissection of several secondary metabolite biosynthetic pathways in plant systems through QTL mapping of the abundance of individual plant metabolites (Szalma et al. 2005; Keurentjes et al. 2006; Schauer et al. 2006; Meyer et al. 2007). For instance, >1,200 'metabolite QTLs,' were detected in tomato using introgression lines that were generated by crossing an elite cultivar, *Solanum lycopersicum* var. Roma, with a wild or 'ancestral' tomato plant, *Solanum pennellii* (Schauer et al. 2006). Similarly, in Arabidopsis also, using a mapping population consisting of 210 RILs, 438 QTLs affecting 243 metabolites were identified, when composite interval mapping (CIM) was conducted for 557 metabolites (Rowe et al. 2008). In *Eucalyptus* also, more than 80 metabolite QTLs (mQTLs), representing variation in 22 known metabolites were identified using a backcross population (van Dyk et al. 2011).

In cereals and grasses also, some studies on the genetics of the accumulation of individual metabolites have been undertaken, although no major study on the identification of QTLs for a large number of metabolites has ever been undertaken. For instance, resistance to Fusarium head blight (FHB) disease in wheat was found to be associated with abundance of 27 metabolites (Hamzehzarghani et al. 2008), accumulation of ABA was found to be associated with leaf size and tiller number in rice and wheat (Quarrie et al. 1997), and QTLs for the main steps of nitrogen (N) metabolism in developing ear and their co-localization with QTLs for kernel yield and putative candidate genes were reported in maize (Canas et al. 2012). In future, more such studies will be conducted in cereals and will facilitate study of metabolites biosynthetic pathways and their use for improvement of yield and quality.

QTL Mapping for Dynamic Traits

Quantitative traits for which phenotypic values change over time during the period of growth are called dynamic traits. Mapping of QTLs for such traits have also been termed as time-related mapping (TRM) as against time-fixed mapping (TFM) for the traits on which the data is recorded at fixed time or stage (Wu et al. 1999). It is logical to think that during the process of development, different QTLs express at different times, although they may have same effect at the end. Wu and Lin (2006) termed it as functional mapping. Examples of such traits can be plant height, tiller

number, pre-harvest sprouting, disease incidence, etc. It is therefore important that phenotypic data on such traits should be recorded at different stages of growth so that it will help in revealing the expression dynamics of individual QTL as done in rice, while detecting 5 QTLs for tiller number (Wu et al. 1999). Following the same approach, Takai et al. (2005) detected two QTLs strongly associated with increased grain filling percentage per panicle in rice. In wheat, this approach was recently used in an association mapping study for pre-harvest sprouting tolerance, where different durations of after-ripening period were used for QTL mapping. Although, the PHS data could not be recorded at regular intervals on the same spikes, functionality of the trait was studied as a response to different durations of after-ripening period (Kulwal et al. 2012). The advantage of this approach is that, one can find the stage at which heritability of the trait is highest, so that the QTLs identified at this stage will be more useful for a breeding program involving MAS.

QTL Mapping for Ordinal Traits

Many quantitative traits in plants are ordinal in nature meaning that observations on such traits are recorded in several ordered categories. Generally the data on disease incidence and many other traits in plants in most cases is recoded on a scale of 0 through 5 or 0 through 9 or 0 through 100 based on the severity of the incidence; this data is subjected to QTL analysis like any other quantitative trait. However, it has been suggested that different statistical approaches should be used to analyse the data on quantitative traits and ordinal/categorical traits (Li J. et al. 2006). One of the reasons for similar treatment of all these different types of traits in earlier studies was partly the lack of availability of statistical tools to deal with these traits. Using simulated data, Hackett and Weller (1995) suggested an ordinal model to estimate the parameters more accurately, especially when the number of categories is small or when only one linked marker is available. Rao and Xu (1998) also tried to address this issue using example of four-way crosses through simulation. Li J. et al. (2006) proposed multiple interval mapping (MIM) for ordinal traits and implemented this strategy in QTL Cartographer and Yi et al. (2007) used Bayesian framework for mapping of interacting QTL for ordinal traits.

QTL Analysis for Quantitative Disease Resistance and QRLs

It is now widely accepted that disease resistance is a quantitative trait, which in each case involves few major QTLs and many minor QTLs influencing the level of resistance in a quantitative manner. While major QTLs are sometimes described as R genes, both major and minor QTLs are together described as quantitative resistance loci or QRLs (Young 1996). Some QRLs have also been reported to be coincident with major R-genes for disease resistance in cereals (Wisser et al. 2005; Friedt and Ordon 2007). Hundreds of these so-called QRLs have already been identified in different crop plants including cereals, and some of these were even cloned through map-based cloning (see section on cloning of QTLs). A distinction has also been

made between, quantitative disease resistance (QDR) and qualitative disease resistance, sometimes even erroneously (Poland et al. 2009; St.Clair 2010), and the term QTL has also often been erroneously used as a synonym to QDR (St.Clair 2010). We feel that it is unnecessary to make a distinction between QRLs and QTLs.

The principle of linkage-based QTL mapping, when extended to QDR and QRLs, certainly facilitated a better understanding of the roles of specific resistance loci in providing race-specificity and partial resistance (Young 1996; Poland et al. 2009; Kou and Wang 2010; St.Clair 2010). The notable examples of QDR include resistance against blast disease of rice, different types of rusts in cereals, Fusarium head blight in wheat and southern and northern leaf blights in maize. Generally the data on QDR is recorded visually leading to variation in the data recorded by different persons, so that the results of QTL mapping are not always accurate. In one of the interesting studies in maize, however, data on northern leaf blight gave consistent results among raters with respect to QTL identification, although estimated allelic effects of these QTLs differed (Poland and Nelson 2011).

A thorough understanding of QDR will also help in the development of durable resistance in high yielding crop cultivars using DNA markers tightly linked with QRLs controlling the QDR (Young 1996; Poland et al. 2009; St.Clair 2010). A long list of these QRLs and the associated markers is now available in cereals, which can be effectively used for MAS for developing durable resistance in cereals. However, QDR and QRLs represent a relatively new terminology, as evident from the fact that while dealing with the QTLs and genes for disease resistance in wheat and barley in the first edition of Cereal Genomics, Jahoor et al. (2004) did not use this terminology.

Advanced Backcross QTL Analysis

QTL detection and use of identified QTL in the breeding program are often treated as separate activities. Perhaps, this is the reason, why not many QTL mapping studies have successfully been translated into the varietal development programs. To address this concern and to exploit the potential of the wild/unadapted germplasm in breeding program, a method of QTL mapping called advanced backcross QTL (AB-QTL) analysis was proposed by Tanksley and Nelson (1996). The goal of this approach was the simultaneous detection and transfer of useful QTLs from the wild/unadapted relatives to a popular cultivar for improvement of a trait. In this method, a backcross population (BC_2, BC_3) is developed from a cross between the superior cultivar/variety and a wild species carrying the desirable target trait, and molecular markers are used to monitor the transfer of QTLs. QTL analysis in this approach is postponed until an advanced backcross generation (BC_2, BC_3, etc.).

AB-QTL analysis was initially used in tomato to improve the fruit phenotypes (Tanksley et al. 1996; Fulton et al. 1997), but was later used in several cereal crops. For example, in wheat, Huang et al. (2003) used this approach to detect QTLs for yield and yield components in a BC_2F_2 population derived from a cross between the German winter wheat variety ‘Prinz’ and the synthetic wheat line W-7984. Out of a total 40 QTLs detected, alleles for 24 of them belonging to synthetic wheat W-7984 were associated with a positive and desirable effect. Similarly,

Narasimhamoorthy et al. (2006) used this approach for detecting QTLs for yield and yield components in a backcross population developed from a cross between hard red winter wheat variety Karl 92 and the synthetic wheat line TA 4152-4.

In rice, using AB-QTL, Moncada et al. (2001) observed that 56% of the detected QTLs were derived from *O. rufipogon*, despite the fact that phenotypic performance of the wild germplasm would not suggest its value as a breeding parent for transfer of QTL alleles of higher value from wild rice (*Oryza rufipogon*). This also shows the power of using this approach in QTL mapping and introgression of the QTL in the breeding program. Thomson et al. (2003) also used this approach to introgress QTLs from *O. rufipogon* to US tropical Japonica cultivar Jefferson. In maize, Ho et al. (2002) used this strategy to identify QTLs of agronomic importance in a cross between two elite inbreds of maize. In barley also, Pillen et al. (2003) observed that *H. vulgare*. ssp. *spontaneum* genotype ISR101-23 was associated with a yield increase of 7.7% averaged over the environments tested.

Mapping As You Go (MAYG)

It is well known and documented now that the detection of QTLs and estimation of the QTL effects are influenced by the genetic background and the environment. This is one reason why MAS programs are not always successful. This limitation can be largely overcome by continuously revising the estimates of QTL effects by remapping the QTLs in the new and elite germplasm generated in each cycle of selection. This approach was described as "Mapping As You Go" (MAYG) by Podlich et al. (2004) and requires that the results of QTL analysis are revised/updated over time during the MAS program so that the QTL estimates remain valid for each cycle of the breeding program.

Meta-QTL Analysis and its Application in Cereals

During the last two decades, the activity involving QTL mapping in plants increased exponentially and as many as ~ 34,300 papers (up to 2010) were published in this area (Danan et al. 2011). It is also well known that there is multiplicity of reports on the same trait in the same crop, so that the use of different parental combinations and/or different environments often resulted in identification of QTLs on the same individual chromosomes involving same genomic regions (Rong et al. 2007). It may, therefore, be necessary to know whether the QTLs identified in a specific genomic region in one study correspond to those detected in the same genomic region in other studies. This issue can be resolved through meta-QTL analysis.

Initially, Chardon et al. (2004) applied Goffinet and Gerber's approach for identification of "hot-spots" for flowering time in maize. A new statistical method based on Gaussian mixture model leading to the development of MetaQTL software was also proposed by Veyrieras et al. (2007). These statistical methods have been utilized for conducting meta-analysis in several crops including wheat, maize, rice and barley, where meta-QTL analysis was conducted for a variety of traits (Table 11.2).

Table 11.2 A summary of meta-QTL analysis studies conducted in cereals

Crop and trait	Number of QTL	Number of metaQTL	Reference
Wheat			
Ear emergence	127	19	Griffiths et al. (2009)
Earliness	84	18	Hanocq et al. (2007)
Fusarium head blight	79	18	Haberle et al. (2009)
Fusarium head blight	77	21	Loffler et al. (2009)
Crop height	104	16	Griffiths et al. (2012)
Yield contributing traits	257	55	Zhang et al. (2010)
Grain dietary fiber content	12	3	Quraishi et al. (2011)
Pre-harvest sprouting	50	8	Tyagi and Gupta (2012)
Grain weight	92	23	Unpublished
Rice			
Drought tolerance	401	32	Khowaja et al. (2009)
Blast resistance	347	165	Ballini et al. (2008)
Root growth	165	9	Norton et al. (2008)
Maize			
Flowering time	313	62	Chardon et al. (2004)
Silage quality	59 (digestibility) and 150 (cell wall composition)	26 (digestibility) and 42 (cell wall composition)	Truntzler et al. (2010)
Plant height	1,201	40	(Wang et al. 2011)
Drought tolerance	399	75	Hao et al. (2010)
Grain yield component	138	16	Li et al. (2011)
NUE	190	37	Liu et al. (2012)
N-remobilization and post silking N-uptake	608	72	(Coque et al. 2008)

Mixed-Model Analysis of Multi-Environment Data

In the Interacting Epistatic QTLs section above, we discussed the importance of QTL × environment interactions (QE) in QTL mapping. Similarly, in Multi-Trait Mapping, we also discussed the use of multi-trait QTL analysis that permits detection of tightly linked or pleiotropic QTL for a number of correlated traits. Most QTL studies, however, are generally conducted each on a single trait using phenotypic data recorded in a single environment. Even when trait data is recorded at several locations and/or years, often only means over locations/years are used for QTL analysis. Only few studies are available, where multi-trait and/or multi-environment (MTME) data is collected and used for QTL analysis. In a plant breeding program, however, a breeder records data on several traits across several locations and years; it is also widely known that genotype by environment interactions (GEI) are generally significant for complex traits with variable heritability. In order to overcome this limitation, mixed models approach was used for analysis of MT or ME data for QTL analysis in barley (Piepho 2000; Malosetti et al. 2004, 2006), maize (Boer et

al. 2007) and rice (Emrich et al. 2008). Later Malosetti et al. (2008) proposed mixed models for an efficient QTL analysis using MTME data, and found it to be very effective while analyzing the data on five traits in maize (Malosetti et al. 2008). The software Genstat has a provision for conducting such an analysis.

11.2.3 Domestication Related QTLs in Cereals

Crop domestication is a co-evolutionary process, where a plant species undergoes transition from a wild to cultivated environments (Purugganan and Fuller 2009; Glemin and Bataillon 2009). In recent years, the advent of molecular marker technology and advances in genomics technologies helped in the genetic dissection of complex domestication traits via QTL analysis (Peleg et al. 2011). The QTLs for domestication traits have largely been identified through 'top-down approaches' (sometimes also called phenotype-first approaches) where we move from phenotype to candidate genes using both linkage-based interval mapping and LD-based association mapping. However, a 'bottom-up approach' can also be used, where we first identify genes with signatures of adaptation using population genetics tools and then identify the phenotype to which these genes contribute. In crops like rice, maize and barley, where significant progress has been made towards the development of genomics resources like ESTs, microarrays, targeted mutagenesis lines, genetic linkage maps, genome sequence, etc., the linking of a candidate gene to phenotype through bottom-up approach is no longer a challenge (Wright et al. 2005; Peleg et al. 2011). The various steps involved in these two approaches are presented in Fig. 11.3.

QTLs involved in domestication have not been uniformly distributed throughout the genome but rather clustered in gene-rich regions of a genome, which correspond to hot spots of recombination (Peng et al. 2003; Ross-Ib arra 2005, 2007). Large number of QTLs have already been identified in cereals like wheat, maize, rice, barley sorghum and pearl millet for a variety of traits including seed size, glumes softness (free threshing), rachis stiffness (shattering), panicle length, plant height, number of tillers, heading date, dormancy, etc. During 1980s Steve Tanksley's group at Cornell University started QTL analysis of fruit mass (a domestication trait in tomato) in a cross between wild and domesticated tomato, localizing six QTLs. Later they successfully isolated a genomic region containing a major QTL *fruit-weight2.2* (*fw2.2*). Doebley and coworkers also isolated major genes that govern phenotypic differences between maize and its wild derivative teosinte (Doebley et al. 1990). These genes in maize included teosinte branched1 (*tb1*), a gene controlling lateral branching (Doebley et al. 1995), and teosinte glume architecture (*tga*), the latter contributing to differences in inflorescence architecture (Wang et al. 2005). The domestication gene "*teosinte branched1* or *tb1*" is the first maize domestication gene and is also one of the most important domestication gene that has been cloned through transposon tagging approach. In case of rice, *Sh4* was cloned as the major shattering QTL, explaining ~69% of phenotypic variance between a traditional *indica* cultivar and the annual wild progenitor *O. nivara* (Li et al. 2006a).

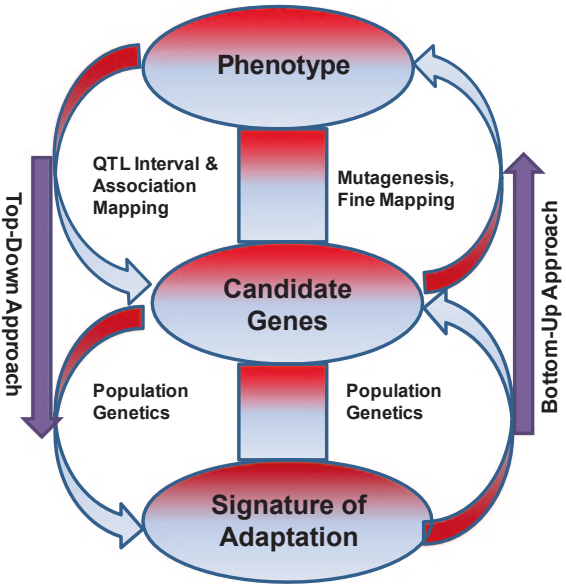

Fig. 11.3 Major steps involved in *top-down* and *bottom-up* approaches of plant domestication (modified from Ross-Ib arra et al. 2007)

Table 11.3 List of domestication QTLs/genes cloned in cereals

Crop	Domestication trait	Gene responsible
Rice	Shattering, threshing	*Sh, qSH1*
	Plant architecture, inflorescence structure	*prog1*
	Grain/seed cover, size, and coloration	*Bh4, Rc, Rd, Phr1, qSW5, GS3*
Wheat	Shattering, threshing	*Q*
Maize	Plant architecture, inflorescence structure	*tb1*
	Grain/seed cover, size, and coloration	*tga1*
Barley	Shattering, threshing	*Nud*
	Plant architecture, inflorescence structure	*vrs1*

Similarly, another shattering QTL, *qSH1*, accounting for ~69% of phenotypic variance between *indica* and temperate *japonica* cultivars, was also cloned (Konishi et al. 2006). A list of QTLs involved in the control of different domestication traits and their details have been tabulated elsewhere (see Pozzi et al. 2004) and their information is also available on some databases like Gramene (Youens-Clark et al. 2011). The domestication QTL, which have been cloned so far are listed in Table 11.3.

11.3 QTL Mapping Based on LD (Association Mapping)

In plants, till recently, identification of QTLs mainly relied on linkage analysis using segregating populations, each derived from a cross between two inbreds. This method proved successful for identification of QTLs for many traits of

interest in a variety of crops. However, this approach has a low resolution, since only limited numbers of recombination events are involved and a limited number of alleles are sampled, so that it is relatively difficult to compare the results of one study with that of another study. Moreover, multiple populations may be required for analysis of many traits which ultimately adds to the expense of generating, genotyping, and phenotyping these populations.

Linkage disequilibrium-based association mapping (AM) is an alternative strategy to identify marker-trait associations (MTAs) and has been used extensively in human and animal genetic experiments, where large segregating populations are not available. Association mapping is also receiving an increasing attention as a method, which complements linkage-based interval mapping in crops (Buckler and Thornsberry 2002; Breseghello and Sorrells 2006a). Association mapping has a number of advantages over other mapping techniques including the potential for increased QTL resolution, and an increased sampling of molecular variation (for reviews, see Buckler and Thornsberry 2002; Flint-Garcia et al. 2003; Gupta et al. 2005; Yu and Buckler 2006; Nordborg and Weigel 2008; Zhu et al. 2008; Stich and Melchinger 2010).

A large number of association mapping studies have been undertaken in plants including cereals. A list of some important association mapping studies carried out in major cereals (bread wheat, rice, barley and maize) is given in Table 11.4. The individual studies either involved a candidate gene approach or a genome-wide association study (GWAS). Each of these two approaches could be either population-based or family-based, although most association studies in plants are population based, in contrast to many association studies in humans, which are largely family-based. List of some of these studies can be found in earlier reviews (Gupta et al. 2005; Hall et al. 2010; Stich and Melchinger 2010). Population structure has often been considered a limitation in association mapping in plants (reviewed by Gupta et al. 2005; Nordborg and Weigel 2008), but methods have been developed to overcome this limitation (Pritchard et al. 2000; Price et al. 2006; Kang et al. 2008). The mixed linear model (MLM) approach was also proposed for controlling false positives, which simultaneously incorporates both population structure and cryptic relationship or kinship (Yu et al. 2005).

One of the earliest AM studies in plants, which took into account population structure, involved gene *Dwarf8* in maize (Thornsberry et al. 2001). Almost all AM studies carried out hereafter also took into account the population structure, if present. Breseghello and Sorrells (2006b) performed AM for kernel size and milling quality traits on a set of 95 cultivars of soft winter wheat which were released since 1980 and were representative of the variability of the elite soft winter wheat germplasm in the eastern US. Using MLM, they identified significant markers for kernel size on chromosome 2D, 5A, and 5B of wheat. In another study, involving the use of NAM population, Tian et al. (2011) showed that variations at the *liguleless* gene contributed to more upright leaves in maize. Using the same NAM population, Kump et al. (2011) and Poland et al. (2011) identified SNPs associated with variation for southern and northern leaf blights in maize, respectively. There has been a steady increase in the number of such studies in crop plants including most cereals. The development of high throughput marker genotyping involving markers like diversity arrays technology (DArT) and single nucleotide polymorphisms (SNPs) has

Table 11.4 A summary of studies involving association mapping (AM) and joint linkage-association mapping (JLAM) in major cereals

Crop and trait	Population size (range)	No. of markers (range)	Reference
Bread wheat			
Grain quality traits	95–207	62 to 115	Breseghello and Sorrells (2006b), Zheng et al. (2009), Reif et al. (2011)
Agronomic traits	96–230	85–874	Yao et al. (2009), Neumann et al. (2011), Mir et al. (2012)
Disease resistance*	44–567	91 to >1,600	Crossa et al. (2007), Tommasini et al. (2007), Adhikari et al. (2011), Miedaner et al. (2011), Yu et al. (2011, 2012)
Aluminium resistance	1,055	178 DArT	Raman et al. (2010)
Flowering time	235	6 candidate genes	Rousset et al. (2011)
Pre-harvest sprouting tolerance	96–242	250–1,166	Jaiswal et al. (2012), Kulwal et al. (2012), Rehman Arif et al. (2012)
Rice			
Agronomic traits	128–517	86–3.6 m	Wen et al. (2009), (de Borba et al. 2010), (Huang et al. 2010, (Zhao et al. 2011), (Li et al. 2012), (Zhou et al. 2012)
Straighthead disorder	547	75 SSR	Agrama and Yan (2009)
Stigma and spikelet characteristics	90	109	Yan et al. (2009)
Silica concentration in rice hulls	174	164	Bryant et al. (2011)
Aluminum tolerance	383	44,000	Famoso et al. (2011)
Disease resistance	Up to 217	155–316	Yoshida et al. (2009), Jia et al. (2012)
Barley			
Morphological and agronomical traits	102– 615	42–1,536	Kraakman et al. (2004, 2006); Rostoks et al. (2006), Stracke et al. (2009), (Cockram et al. 2010), Sun et al. (2011), Wang et al. (2012), Rode et al. (2012)

(continuied)

Table 11.4 (continued)

Crop and trait	Population size (range)	No. of markers (range)	Reference
Abiotic stress tolerance (salinity, drought, winter hardiness)	48– 188	22– 3,072	Eleuch et al. (2008), (von Zitzewitz et al. 2011), Wu et al. (2011), Varshney et al. (2012)
Disease resistance	318– 768	1,536 to > 3,000	Roy et al. (2010), Massman et al. (2011)
Maize			
Flowering time	92	Candidate gene	Thornsberry et al. (2001)
Biochemical parameters	86–350	Candidate gene; 1,229	Szalma et al. (2005), (Setter et al. 2011)
Aluminum tolerance	282	21 Candidate genes	Krill et al. (2010)
Disease resistance	~ 5,000	1,106–1.6 million	Poland et al. (2011)**, (Kump et al. 2011)**
Leaf architecture	4892	1.6 million SNPs	Tian et al. (2011)**
Kernel composition	4,699 RILs + 282 ILs	1,106 SNPs	Cook et al. (2012)**

*Stem rust, leaf rust, yellow rust, powdery mildew, *Stagnospora nodurum* glume blotch, Fusarium head blight
**JLAM studies

facilitated AM studies. More recently, the development of genotyping by sequencing (GBS) technology has also been successfully applied to wheat and barley, and should be the future technology of choice for AM studies in cereals (Poland et al. 2012). There is little doubt that association mapping will be increasingly used in future leading to better understanding of the genetics of complex traits and their possible use for improvement of crops in general and for improvement of cereals in particular.

11.4 Joint Linkage and Association Mapping

We know that both linkage-based interval mapping and LD-based association mapping, discussed above, have their own advantages and limitations when used alone (Wu et al. 2002; reviewed in Gupta et al. 2005; Nordborg and Weigel 2008). It was therefore proposed to integrate these two approaches into one approach called joint linkage-association mapping or JLAM in brief (Wu and Zeng 2001; Wu et al. 2002; Meuwissen et al. 2002; Myles et al. 2009). Using ML estimates and based on extensive simulations, Wu et al. (2002) showed that this method allows for simultaneous estimation of a number of genetic and genomic parameters including the allele frequencies of each individual QTL, QTL effect and location, and association of a QTL with a known marker locus. Later, this method was also extended for the multi-trait data (Meuwissen and Goddard 2004; Stich et al. 2008). In a recent study, Lu et al. (2010) used this method in maize involving NAM population. For identification of QTL underlying drought tolerance; they identified 18 new QTLs, which were not identified by either of the two methods individually (parallel mapping). Keeping in view the merits of JLAM, this seems to be the future approach of choice for genetic analysis of quantitative traits.

11.5 Cloning of QTLs

The QTLs identified through QTL interval mapping are generally located in wide intervals (10–20 cM), which may sometimes contain several hundreds of genes. Therefore, it will be desirable to move close to the target QTL and clone it, so that perfect functional markers may be developed for introgression of respective QTL/gene through molecular breeding. Positional cloning (often also called map based cloning; Jander et al. 2002) of QTLs involves localization of a QTL to shortest possible marker interval (QTL fine mapping) followed by identification of corresponding interval on the DNA sequence (QTL physical mapping). Candidate genes genetically and physically co-segregating with the QTL are then identified and/or selected for evaluation (Salvi and Tuberosa 2005). Majority of the QTLs isolated so far have been cloned through positional cloning, although this is believed to be the most tedious and time-consuming method for QTL cloning. Other approaches like LD-based association mapping and functional genomics which bypasses the tedious procedures of positional cloning can lead to

identification of candidate genes and are now also becoming popular (Salvi and Tuberosa 2005, 2007). In general, QTL cloning involves the following steps (see Krattinger et al. 2009a): (1) development of a large mapping population (>1,500 plants) derived from a cross between two NILs for the target QTL; (2) reducing the QTL interval using this population; (3) identification of a contig covering the QTL region by screening the closely linked molecular markers with a large insert library (e.g., BAC or YAC library) and fingerprinting the candidate BAC or YAC clone; (4) sequencing the contig and using the sequence for identification of the gene, and (5) validation of candidate gene(s) to test its effect on phenotype.

During the last few years, many reports have become available on cloning of QTLs in all major cereal crops for different traits including difficult traits like root and drought tolerance (see Keller et al. 2007; Salvi and Tuberosa 2007; also see Table 11.5). All these QTLs were first identified through interval mapping and later cloned through positional cloning. One of the earliest cloned QTL in maize is the one for plant architecture (Doebley et al. 1995, 1997). Other examples from maize include cloning of a QTL for glume architecture (*Tga1*; Wang et al. 2005) and a major flowering time QTL "*Vgt1*" found associated with drought tolerance (Salvi et al. 2007). Similarly, QTLs for glume architecture (*Tga1*) and plant architecture have also been cloned in maize (Wang et al. 2005; Doebley et al. 1995, 1997). Efforts are also being made to clone two major wheat QTLs on chromosome bins 1.06 and 2.04 (root-ABA1) affecting root architecture and a number of agronomic traits, including grain yield (Tuberosa and Salvi 2007).

However, the maximum numbers of QTLs so far have been cloned in rice and wheat. Some of the important traits, for which QTLs have been cloned in rice include the following: (1) heading date (*Hd1/Se1, Hd6/CK2, Hd3a, Ehd1*), (2) grain size and length (*GS3*), (3) grain number (*Gn1/CKX2*), (4) regenerability (*PSR1*), (5) seed shattering (*qSH-1/RPL* and *sh4*), (6) salt tolerance (*SKC1*), (7) submergence tolerance (*Sub1*), and (8) UV resistance (*qUVR-10*). Similarly, important QTLs cloned in wheat through positional/map-based cloning include the following: (1) *Gpc-B1* associated with increased grain protein, zinc, and iron content (Uauy et al. 2006), (2) *Yr36* (*WKS1*), which confers resistance to a broad spectrum of stripe rust races at relatively high (25° to 35°C) temperatures (Fu et al. 2009), (3) *Lr34* which codes for an ABC transporter that confers durable resistance to multiple fungal pathogens causing leaf rust in wheat (Krattinger et al. 2009b).

All the QTLs which have been cloned so far were earlier identified mostly as major QTL in biparental mapping populations. The success behind cloning of these QTLs lies in the fact that they were initially mapped either on the exact position or very close to the real position (Price 2006). It also implies that for successful cloning of a QTL, the correct positioning of QTL in the marker interval is very important and that an error of few cM can account for huge physical region of the genome, particularly in a crop like wheat. The newer genomics approaches like association mapping, and next generation sequencing technologies which are routinely used now by plant geneticists hold great promises to accelerate the science of QTL cloning for important traits in cereal crops using natural populations. These advances will also make it possible to target the QTLs other than those with a major effect (Salvi and Tuberosa 2007).

Table 11.5 List of QTLs and QRLs cloned in important cereals

Crop	QTL	Trait	Function	Reference
Wheat				
	Gpc1	Grain protein content	Transcription factor	Uauy et al. (2006)
	Lr34	Leaf rust, stripe rust and powdery mildew	Encodes a protein resembling pleiotropic drug resistance-like ABC transporter.	Krattinger et al. (2009b)
	Yr36	Stripe rust	Encodes a kinase-START protein	Fu et al. (2009)
Rice				
	Hd1	Flowering time	Transcription factor	Yano et al. (2000)
	Hd6	Flowering time	Protein kinase	Takahashi et al. (2001)
	Hd3a	Flowering time	Unknown	Kojima et al. (2002)
	Ehd1	Heading time	B-type response regulator	Doi et al. (2004)
	Gn1	Grain number	Cytokinin oxidase	Ashikari et al. (2005)
	PSR1	Regenerability	Nitrite reductase	Nishimura et al. (2005)
	SKC1	Salt tolerance	HKT Transporter	Ren et al. (2005)
	qUVR-10	UV resistance	CPD Photlyse	Ueda et al. (2005)
	GS3	Grain size and length	VWFC membrane protein	Fan et al. (2006)
	Sub1	Submergence tolerance	Transcription factor	Xu et al. (2006)
	sh4	Seed shattering	Transcription factor	Li et al. (2006a)
	GW2	Grain width and weight	RING-type E3 ubiquitin ligase	Song et al. (2007)
	pi21	Blast resistance	Proline-rich protein	Fukuoka et al. (2009)
	Sdr4	Dormancy and domestication	Regulator	Sugimoto et al. (2010)
	qSH1	Seed shattering	BEL-1 homeobox	Konishi et al. (2006)
Maize	*tb1*	Plant architecture	Transcription factor	Doebley et al. (1997)
	tga1	Glume architecture	Transcription factor	Wang et al. (2005)
	Vgt1	Flowering time	Transcription factor	Salvi et al. (2007)
Barley	*Ppd-H1*	Photoperiod response	Pseudo-response regulator	Turner et al. (2005)

Some of the QRLs involved in QDR have also been cloned in a number of crops including cereals. These cloning reports are expected to increase in future in all crops including cereals, since more and more crop genomes are being sequenced and genomic resources are increasingly becoming available. For MAS and for the production of transgenics, it is desirable that the ideal functional marker should be the actual causal gene(s) and/or QTN (Michelmore 2003; Moose and Mumm 2008). Although, large numbers of QRLs have been reported, very few of them have been cloned, followed by successful functional validation of the causal gene(s) and identification of causal QTN(s). In cereals, so far only three QRLs conferring QDR have been cloned, including one for blast resistance (*pi21*) in rice (Fukuoka et al. 2009) and two slow rust resistances QRLs (*Lr34* and *Yr36*) in wheat (Fu et al. 2009; Krattinger et al. 2009b; Table 11.5). Further insights about cloning of these QRLs are available elsewhere (St.Clair 2010; Kou and

Wang 2010). It is interesting to note that the causal genes for all these three cloned QRLs are structurally different from the R-genes (St.Clair 2010). Efforts for cloning a major QTL for resistance against Fusarium head blight (*Fhb1*) in wheat are also underway (Liu et al. 2008). Cloning of more such QRLs and identification of causal genes(s) will help in determining the nature of genes underlying QDR.

11.6 Computer Software for QTL Mapping

The success of any QTL mapping experiment depends on the time and the cost it involves and the computation demand it has. During the last two decades after the development of MAPMAKER for construction of linkage maps in 1987, a number of user-friendly efficient computer software have been developed for a faster QTL mapping involving both linkage-based interval mapping and LD-based association mapping. A majority of available programs, each has provisions for majority of approaches discussed in this chapter (Table 11.6; http://www.rqtl.org/; verified

Table 11.6 List of software available for QTL and association mapping

Name of software	Features	Reference
SAS program	ANOVA, AM	Knapp and Bridges (1990)
MAPMAKER/QTL	SIM	Lincoln et al. (1993)
MapManager QTX	SIM, CIM	Manly and Olson (1999)
MQTL	CIM	Tinker and Mather (1995)
MAPQTL	SIM, CIM	van Ooigen and Maliepaard (1996)
MultiQTL	SIM, MIM	www.multiqtl.com
PLAB QTL	SIM, CIM, Epistatic QTL	Utz and Melchinger (1996)
Qgene	SIM	Nelson (1997)
Multimapper	BIM	Sillanpaa and Arjas (1998)
QTLMapper /QTL network	CIM, Epistatic QTL	Wang et al. (1999)
QTL express	SIM, CIM	Seaton et al. (2002)
R/qtl	SIM, CIM, Epistatic QTL	Broman et al. (2003)
QTL cartographer	SIM, CIM, MIM, BIM, Ordinal trait	(Wang et al. 2011)
Genotype matrix mapping	SIM, CIM, Epistatic QTL	Isobe et al. (2007)
MapDisto	ANOVA	http://mapdisto.free.fr/MapDisto/
IciMapping	ICIM	Li et al. (2007)
MetaQTL	Meta-analysis	Veyrieras et al. (2007)
TASSEL	AM, LD, PCA	Bradbury et al. (2007)
BIMBAM	AM	Servin and Stephens (2007)
FlexQTL	BIM	Bink Marco and van Eeuwijk (2009)
GAPIT	AM, genomic prediction	Lipka et al. (2011)

May 22, 2012). The choice of the software also however, depends on the objective and the method, which the researcher considers appropriate for the dataset. Among these software, "R" is one such free software, which has the ability of statistical computing and graphics (http://www.r-project.org/) and can run on a variety of platforms. The advantage of using R is that, one can write the script/code for any analysis and can distribute the code to anybody. Because of this unique feature, in recent years, the focus has shifted from using stand-alone computer programs to R package to perform all sorts of statistical analyses including QTL mapping and association mapping. One such example is the R/qtl, which can perform QTL mapping using a variety of statistical approaches (Broman et al. 2003; http://www.rqtl.org/).

11.7 Limitations of QTL Mapping

The phenotypic variation explained by a given QTL is the function of the size of the mapping population, marker density, trait heritability and the QTL mapping approach that is used for QTL analysis. In a simulation study, it was shown that the average estimates of phenotypic variances associated with an identified QTL were greatly overestimated if only 100 progeny were evaluated, slightly overestimated if 500 progeny were evaluated, and fairly close to the actual magnitude when 1000 progeny were evaluated; this is popularly known as Beavis effect (Beavis 1998). It was also shown that while using small populations, the QTL effects are overestimated and the statistical power of QTL analysis to detect minor QTL is compromised (for further details, see Xu 2003b).

Other limitations of biparental QTL mapping include failure to detect loci for which the parents do not differ, so that several loci controlling the trait of interest will escape detection. Also, the goal of QTL interval mapping is to identify loci rather than allele mining, so that the breeder will not be able to identify the most desirable and novel alleles for the breeding program; often the most desirable alleles may not be represented in the parents. This limitations is largely overcome in association mapping, which otherwise suffers with another limitation of high rate of false positives. Another limitation of QTL interval mapping is overestimation of QTL effects for most complex traits with low heritability, when mapping populations with small to moderate size are used for QTL analysis; resistance to abiotic stresses is one such group of traits of agronomic value.

11.8 Summary and Outlook

During the last two decades, significant progress has been made in the field of QTL mapping, thus giving a new direction to the studies involving inheritance and genetic dissection of complex quantitative traits both in plant and animal systems including humans. This has become possible due to significant developments in

two major research areas; first the twin areas of molecular markers and genome sequencing, and second the subject area of statistical genomics (Morrell et al. 2012). The marker-trait associations (MTAs) detected through QTL mapping also facilitated indirect marker-aided selection (MAS) in plant breeding programs leading to the development of a new area of science popularly described as molecular breeding. The use of indirect MAS as a component of conventional plant breeding has already led to the development of dozens of improved cultivars in cereals alone (Gupta et al. 2010a, b; Xu 2010), and the research activity in this area is steadily growing. The statistical tools are also being further refined by development of new concepts, experimental designs, algorithms and software to facilitate further the research activity in the area of QTL mapping. In this manner, we are gradually overcoming the currently known limitations of the available methods, which have been widely discussed. For instance we know that both linkage analysis and LD-based association mapping have their own limitations when used independently, and efforts are, therefore, being made to overcome these limitations (Myles et al. 2009). One such approach is the development of a variety of experimental designs for joint linkage association mapping (JLAM), which was initially proposed more than ten years ago (Wu and Zeng 2001; Wu et al. 2002), and is being regularly improved by comparing different models that are currently available (Myles et al. 2009; Würschum et al. 2012) and has been successfully utilized in maize (see Table 11.4).

In the area of QTL mapping including both linkage-based and LD-based methods, an improved method in the form of inclusive composite interval mapping (ICIM) has become available, and further progress is being made through simulation and empirical studies. Several software are now available, which allow study of not only the two-locus QTL × QTL interactions, often involving QTLs, which do not have their own independent main effects, but also the QTL × environment interactions. In future, the study of higher order interactions will certainly be possible, to elucidate the complex genetic networks that are involved in the expression of complex quantitative traits.

In association studies, it is known that genotype-phenotype associations can be partly due to population structure and kinship leading to the discovery of large proportion of false positives. The problem of rare frequency alleles is also receiving attention, and will certainly be fully resolved in future. We know that large number of QTLs detected through association mapping often explain only a very small proportion of the phenotypic variation (<5–10%), perhaps due to a large number of minor QTLs and other major but rare QTLs escaping detection. We already have the provision of using Q + K matrices to address the problem of structure and relatedness, but the problem of rare alleles perhaps remains to be fully resolved.

Another related but important growing area of interest is genomic selection (GS) or genome-wide selection (GWS), which does not make a part of this review on QTL mapping, but certainly makes a part of molecular breeding, to which QTLs contribute (Meuwissen et al. 2001). This may address the problem of minor and rare QTLs, which together may contribute substantially to the total

phenotypic variation. However, construction of a suitable training population and the computational part involving estimation of breeding values of alleles at different genetically mapped marker loci remains a challenge. In view of this, not many studies have been conducted involving genomic selection in cereals, but this area of research is likely to grow in future. The availability of low-cost and high-throughput genotyping including genotyping by sequencing (GBS) will certainly facilitate research in the field of genomics and association mapping. Ed Buckler (2012) estimates that following GBS, the cost of genotyping has already gone down to US $ 10–20 per sample to genotype it for thousands of variations (markers), thus providing a powerful approach for future association studies.

In future, one would also witness a shift from detection of PhQTL to that of eQTL, pQL and mQTL. This information would be used for elucidating the biosynthetic pathways underlying quantitative traits of agroinomic value. QTL analysis will also include mapping the contribution of imprinting or epigenetic changes to morphological traits (epigenetic QTLs), which is a fast growing area of research.

In summary, we can say that considerable progress has already been made in the area of QTL mapping and molecular breeding in cereals, but the field of research is still growing and will keep statisticians, molecular biologists and plant breeders busy in the years to come.

References

Adhikari TB, Jackson EW, Gurung S, Hansen JM, Bonman JM (2011) Association mapping of quantitative resistance to *Phaeosphaeria nodorum* in spring wheat landraces from the USDA national small grains collection. Phytopathol 11:1301–1310

Agrama HA, Yan WG (2009) Association mapping of straighthead disorder induced by arsenic in *Oryza sativa*. Plant Breeding 128:551–558

Apotikar DB, Venkateswarlu D, Ghorade RB, Wadaskar RM, Patil JV, Kulwal PL (2011) Mapping of shoot fly tolerance loci in sorghum using SSR markers. J Genet 90:59–66

Ashikari M, Sakakibara H, Lin S, Yamamoto T, Takashi T, Nishimura A, Angeles ER, Qian Q, Kitano H, Matsuoka M (2005) Cytokinin oxidase regulates rice grain production. Science 309:741–745

Ballini E, Morel JB, Droc G, Price A, Courtois B, Notteghem JL, Tharreau D (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

Bandillo NB, Muyco PA, Redona E, Gregorio G, Singh KK, Leung H (2011) Population development through multiparent advanced generation intercrosses (MAGIC) among diverse genotypes to facilitate gene discovery for various traits in rice. Phill J Crop Sci 36:32–33

Bauer AM, Hoti F, von Korff M, Pillen K, Leon J, Sillanpaa MJ (2009) Advanced backcross-QTL analysis in spring barley (*H. vulgare* ssp. spontaneum) comparing a REML versus a Bayesian model in multi-environmental field trials. Theor Appl Genet 119:105–123

Beaumont MA, Rannala B (2004) The Bayesian revolution in genetics. Nat Rev Genet 5:251–261

Beavis WD (1998) QTL analyses: power, precision, and accuracy. In: Paterson AH (ed) Molecular dissection of complex traits. CRC Press, New York, pp 145–162

Bernardo R (2008) Molecular markers and selection for complex traits in plants: Learning from the last 20 years. Crop Sci 48:1649–1664

Bink MCAM, Boer MP, ter Braak CJF, Jansen J, Voorrips RE, van de Weg WE (2008) Bayesian analysis of complex traits in pedigreed plant populations. Euphytica 161:85–96

Bink Marco CAM, van Eeuwijk FA (2009) A Bayesian QTL linkage analysis of the common dataset from the 12th QTL-MAS workshop. BMC Proc 3:S4

Boer MP, Wright D, Feng L, Podlich DW, Luo L, Cooper M, van Eeuwijk FA (2007) A mixed-model quantitative trait loci (QTL) analysis for multiple-environment trial data using environmental covariables for QTL-by-environment interactions, with an example in maize. Genetics 177:1801–1813

de Borba TCO, Brondani RPV, Breseghello F et al (2010) Association mapping for yield and grain quality traits in rice (*Oryza sativa* L.). Genet Mol Biol 33:515–524

Borevitz JO, Chory J (2004) Genomics tools for QTL analysis and gene discovery. Curr Opin Plant Biol 7:132–136

Borner A, Schumann E, Furste A, Coster H, Leithold B, Roder MS, Weber WE (2002) Mapping of quantitative trait loci determining agronomic important characters in hexaploid wheat (*Triticum aestivum* L). Theor Appl Genet 105:921–936

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635

Breseghello F, Sorrells ME (2006a) Association analysis as a strategy for improvement of quantitative traits in plants. Crop Sci 46:1323–1330

Breseghello F, Sorrells ME (2006b) Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. Genetics 172:1165–1177

Broman KW, Wu H, Sen Ś, Churchill GA (2003) R/qtl: QTL mapping in experimental crosses. Bioinformatics 19:889–890

Bryant R, Proctor A, Hawkridge M, Jackson A, Yeater K, Counce P, Yan W, McClung A, Fjellstrom R (2011) Genetic variation and association mapping of silica concentration in rice hulls using a germplasm collection. Genetica 139:1383–1398

Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ et al (2009) The genetic architecture of maize flowering time. Science 325:714–718

Buckler ES, Thornsberry JM (2002) Plant molecular diversity and applications to genomics. Curr Opin Plant Biol 5:107–111

Buckler ES (2012) Uniting the world's maize diversity for detection of complex traits and accelerating breeding. 4th international conference on quantitative genetics: understanding variation in complex traits (Edinburgh, UK; 17–22 June 2012). Book of abstracts, p 29

Cadalen T, Sourdille P, Charmet G, Tixier MH, Gay G, Boeuf C, Bernard S, Leroy P, Bernard M (1998) Molecular markers linked to genes affecting plant height in wheat using a doubled-haploid population. Theor Appl Genet 96:933–940

Canas RA, Quillere I, Gallais A, Hirel B (2012) Can genetic variability for nitrogen metabolism in the developing ear of maize be exploited to improve yield? New Phytol 194:440–452

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

CGIAR Generation Challenge Programme (2009) 2009 project updates. Generation challenge programme, Texcoco, Mexico

Chardon F, Virlon B, Moreau L, Falque M, Joets J, Decousset L, Murigneux A, Charcosset A (2004) Genetic architecture of flowering time in maize as inferred from quantitative trait loci meta-analysis and synteny conservation with the rice genome. Genetics 168:2169–2185

Chen X, Hackett CA, Niks RE, Hedley PE, Booth C, Druka A, Marcel TC, Vels A, Bayer M, Milne I, Morris J, Ramsay L, Marshall D, Cardle L, Waugh R (2010) An eQTL analysis of partial resistance to *Pucciniahordei* in barley. PLoS ONE 5:e8598

Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. Genetics 138:963–971

Cockram J, White J, Zuluaga DL, Smith S et al (2010) Genome-wide association mapping to candidate polymorphism resolution in the unsequenced barley genome. Proc Natl Acad Sci USA 107:21611–21616

Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. Euphytica 142:169–196

Cook JP, McMullen MD, Holland JB, Tian F, Bradbury P, Ross-Ibarra J, Buckler ES, Flint-Garcia SA (2012) Genetic architecture of maize kernel composition in the nested association mapping and inbred association panels. Plant Physiol 158:824–834

Coque M, Martin A, Veyrieras JB, Hirel B, Gallais A (2008) Genetic variation for N-remobilization and postsilking N-uptake in a set of maize recombinant inbred lines. Theor Appl Genet 117:729–747

Crepieux S, Lebreton C, Servin B, Charmet G (2004) Quantitative trait loci (QTL) detection in multicross inbred designs: recovering QTL identical-by-descent status information from marker data. Genetics 168:1737–1749

Crossa J, Burgueno J, Dreisigacker S, Vargas M, Herrera-Foessel SA, Lillemo M, Singh RP, Trethowan R, Warburton M, Franco J, Reynolds M, Crouch JH, Ortiz R (2007) Association analysis of historical bread wheat germplasm using additive genetic covariance of relatives and population structure. Genetics 177:1889–1913

Damerval C, Maurice A, de Josse JM, Vienne D (1994) Quantitative trait loci underlying gene product variation: a novel perspective for analyzing regulation of genome expression. Genetics 137:289–301

Danan S, Jean-Baptiste V, Véronique L (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

Darvasi A, Soller M (1995) Advanced intercross lines, an experimental population for fine genetic mapping. Genetics 141:1199–1207

de Vienne D, Leonardi A, Damerval C, Zivy M (1999) Genetics of proteome variation for QTL characterization: application to drought-stress responses in maize. J Exp Bot 50:303–309

Doebley J, Stec A, Gustus C (1995) *Teosinte branched1* and the origin of maize: evidence for epistasis and the evolution of dominance. Genetics 141:333–346

Doebley J, Stec A, Wendel J, Edwards M (1990) Genetic and morphological analysis of a maize-teosinte F_2 population: implications for the origin of maize. Proc Natl Acad Sci USA 87:9888–9892

Doebley J, Stec A, Hubbard L (1997) The evolution of apical dominance in maize. Nature 386:485–488

Doerge RW (2002) Mapping and analysis of quantitative trait loci in experimental populations. Nat Rev Genet 3:43–52

Doerge RW, Churchill GA (1996) Permutation tests for multiple loci affecting a quantitative character. Genetics 142:285–294

Doi K, Izawa T, Fuse T, Yamanouchi U, Kubo T, Shimatani Z, Yano M, Yoshimura A (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

Eleuch L, Jilal A, Grando S, Ceccarelli S, Schmising MK, Tsujimoto H, Hajer A, Daaloul A, Baum M (2008) Genetic diversity and association analysis for salinity tolerance, heading date and plant height of barley germplasm using simple sequence repeat markers. J Integr Plant Biol 50:1004–1014

Emrich K, Price A, Piepho HP (2008) Assessing the importance of genotype x environment interaction for root traits in rice using a mapping population III: QTL analysis by mixed models. Euphytica 161:229–240

Famoso AN, Zhao K, Clark RT, Tung C-W, Wright MH, Bustamante C, Kochian LV, McCouch SR (2011) Genetic architecture of aluminum tolerance in rice (*Oryza sativa*) determined through genome-wide association analysis and QTL mapping. PLoS Genet 7:e1002221

Fan CC, Yu XQ, Xing YZ, Xu CG, Luo LJ, Zhang Q (2005) The main effects, epistatic effects and environmental interactions of QTLs on the cooking and eating quality of rice in a doubled-haploid line population. Theor Appl Genet 110:1445–1452

Fan C, Xing Y, Mao H, Lu T, Han B, Xu C, Li X, Zhang Q (2006) GS3, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. Theor Appl Genet 112:1164–1171

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

Friedt W, Ordon F (2007) Molecular markers for gene pyramiding and disease resistance breeding in barley. In: Varshney RK, Tuberosa T (eds) Genomics-assisted crop improvement, vol 2. Genomics applications in crops, Springer, Berlin, pp 81–101

Fu DL, Uauy C, Distelfeld A, Blechl A, Epstein L, Chen XM, Sela H, Fahima T, Dubcovsky J (2009) A kinase-START gene confers temperature-dependent resistance to wheat stripe rust. Science 323:1357–1360

Fukuoka S, Saka N, Koga H, Ono K, Shimizu T, Ebana K, Hayashi N, Takahashi A, Hirochika H, Okuno K, Yano M (2009) Loss of function of a proline-containing protein confers durable disease resistance in rice. Science 325:998–1001

Fulton TM, Beck-Bunn T, Emmatty D, Eshed Y, Lopez J, Petiard V, Uhlig J, Zamir D, Tanksley SD (1997) QTL analysis of an advanced backcros of *Lycopersicon peruvianum* to the cultivated tomato and comparisons with QTLs found in other wild species. Theor Appl Genet 95:881–894

Gardiner JM, Coe EH, Melia-Hancock S, Hoisington DA, Chao S (1993) Development of a core RFLP map in maize using an immortalized F_2 population. Genetics 154:917–930

Gill KS, Lubbers EL, Gill BS, Raupp WJ, Cox TS (1991) A genetic linkage map of *Triticum tauschii* (DD) and its relationship to the D genome of bread wheat (AABBDD). Genome 34:362–374

Glemin S, Bataillon T (2009) A comparative view of the evolution of grasses under domestication. New Phytol 183:273–290

Graner A, Jahoor A, Schondelmaier J, Siedler H, Pillen K, Fischbeck G, Wenzel G, Herrmann RG (1991) Construction of an RFLP map of barley. Theor Appl Genet 83:250–256

Griffiths S, Simmonds J, Leverington M, Wang Y, Fish L, Sayers L, Alibert L, Orford S, Wingen L, Herry L, Faure S, Laurie D, Bilham L, Snape J (2009) Meta-QTL analysis of the genetic control of ear emergence in elite European winter wheat germplasm. Theor Appl Genet 119:383–395

Griffiths S, Simmonds J, Leverington M, Wang Y, Fish L, Sayers L, Alibert L, Orford S, Wingen L, Snape J (2012) Meta-QTL analysis of the genetic control of crop height in elite European winter wheat germplasm. Mol Breeding 29:159–171

Gupta PK, Kulwal PL (2006) Methods of QTL analysis in crop plants: present status and future prospects. In: Trivedi PC (ed) Biotechnology and biology of plants. Avishkar Publishers, Jaipur, pp 1–23

Gupta PK, Rustgi S, Kulwal PL (2005) Linkage disequilibrium and association studies in higher plants: present status and future prospects. Plant Mol Biol 57:461–485

Gupta PK, Kumar J, Mir RR, Kumar A (2010a) Marker-assisted selection as a component of conventional plant breeding. Plant Breeding Rev 33:145–217

Gupta PK, Langridge P, Mir RR (2010b) Marker-assisted wheat breeding: present status and future possibilities. Mol Breeding 26:145–161

Haberle J, Holzapfel J, Schweizer G, Hartl L (2009) A major QTL for resistance against Fusarium head blight in European winter wheat. Theor Appl Genet 119:325–332

Hall D, Tegstrom C, Ingvarsson PK (2010) Using association mapping to dissect the genetic basis of complex traits in plants. Brief Funct Genomics 9:157–165

Hackett CA (2002) Statistical methods of QTL mapping in cereals. Plant Mol Biol 48:585–599

Hackett CA, Weller JI (1995) Genetic mapping of quantitative trait loci for traits with ordinal distributions. Biometrics 51:1252–1263

Haley CS, Knott SA (1992) A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity 69:315–324

Hamzehzarghani H, Paranidharan V, Abu-Nada Y, Kushalappa AC, Mamer O, Somers D (2008) Metabolic profiling to discriminate wheat near isogenic lines, with quantitative trait loci at chromosome 2DL, varying in resistance to fusarium head blight. Can J Plant Sci 88:789–797

Hanocq E, Laperche A, Jaminon O, Laine AL, Gouis JL (2007) Most significant genome regions involved in the control of earliness traits in bread wheat, as revealed by QTL meta-analysis. Theor Appl Genet 114:569–584

Hansen BG, Halkier BA, Kliebenstein DJ (2008) Identifying the molecular basis of QTLs: eQTLs add a new dimension. Trends Plant Sci 13:72–77

Hao Z, Li X, Liu X, Xie C, Li M, Zhang D, Zhang S (2010) Meta-analysis of constitutive and adaptive QTL for drought tolerance in maize. Euphytica 174:165–177

Harushima Y, Yano M, Shomura A, Sato M et al (1998) A high-density rice genetic linkage map with 2275 markers using a single F2 population. Genetics 148:479–494

Ho J, McCouch S, Smith M (2002) Improvement of hybrid yield by advanced backcross QTL analysis in elite maize. Theor Appl Genet 105:440–448

Hua J, Xing Y, Wu W, Xu C, Sun X et al (2003) Single-locus heterotic effects and dominance by dominance interactions can adequately explain the genetic basis of heterosis in an elite rice hybrid. Proc Natl Acad Sci USA 100:2574–2574

Huang XQ, Coster H, Ganal MW, Roder MS (2003) Advanced backcross QTL analysis for the identification of quantitative trait loci alleles from wild relatives of wheat (*Triticum aestivum* L.). Theor Appl Genet 106:1379–13

Huang X, Paulo M-J, Boer M, Effgen S, Keizer P, Koornneef M, van Eeuwijk FA (2011) Analysis of natural allelic variation in *Arabidopsis* using a multiparent recombinant inbred line population. Proc Natl Acad Sci USA 108:4488–4493

Huang X, Wei X, Sang T, Zhao Q, Feng Q et al (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. Nat Genet 42:961–967

Isobe S, Nakaya A, Tabata S (2007) Genotype matrix mapping: searching for quantitative trait loci interactions in genetic variation in complex traits. DNA Res 14:217–225

Jahoor A, Eriksen L, Backes G (2004) QTLs and genes for disease resistance in barley and wheat. In: Gupta PK, Varshney RK (eds) Cereal genomics. Kluwer Academic Publishers, The Netherlands, pp 199–251

Jaiswal V, Mir RR, Mohan A, Balyan HS, Gupta PK (2012) Association mapping for pre-harvest sprouting tolerance in common wheat (*Triticum aestivum* L.). Euphytica 188:89–102

Jander G, Norris SR, Rounsley SD, Bush DF, Levin IM, Last RL (2002) Arabidopsis map-based cloning in the post-genome era. Plant Physiol 129:440–450

Jansen RC, Nap J-P (2001) Genetical genomics: the added value from segregation. Trends Genet 17:388–391

Jansen RC (2007) Quantitative trait loci in inbred lines. In: Handbook of statistical genetics, 3rd edn. Wiley, New York. ISBN: 978-0-470-05830-5

Jansen RC, Tesson BM, Fu J, Yang Y, McIntyre LM (2009) Defining gene and QTL networks. Curr Opin Plant Biol 12:241–246

Jia L, Yan W, Zhu C, Agrama HA, Jackson A et al (2012) Allelic analysis of sheath blight resistance with association mapping in rice. PLoS ONE 7(3):e32703

Jordan MC, Somers DJ, Banks TW (2007) Identifying regions of the wheat genome controlling seed development by mapping expression quantitative trait loci. Plant Biotechnol J 5:442–453

Jourjon M-F, Jasson S, Marcel J, Ngom B, Mangin B (2005) MCQTL: multi-allelic QTL mapping in multi-cross design. Bioinformatics 21:128–130

Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, Eskin E (2008) Efficient control of population structure in model organism association mapping. Genetics 178:1709–1723

Kao CH, Zeng ZB, Teasdale RD (1999) Multiple interval mapping for quantitative trait loci. Genetics 152:1203–1216

Katja W, Pietsch C, Strickert M, Matros A, Roder MS, Weschke W, Wobus U, Mock H-P (2011) Mapping of quantitative trait loci associated with protein expression variation in barley grains. Mol Breeding 27:301–314

Keller B, Bieri S, Bossolini E, Yahiaoui N (2007) Cloning genes and QTLs for disease resistance in cereals. In: Varshney RK, Tuberosa R (eds) Genomics assisted crop improvement, vol 2. Genomics applications in crops, Springer, Berlin, pp 103–128

Keurentjes JJ, Fu J, de Vos CH, Lommen A, Hall RD, Bino RJ, van der Plas LH, Jansen RC, Vreugdenhil D, Koornneef M (2006) The genetics of plant metabolism. Nat Genet 38:842–849

Khowaja FS, Gareth NJ, Brigitte C, Adam PH (2009) Improved resolution in the position of drought-related QTLs in a single mapping population of rice by meta-analysis. BMC Genomics 10:276

Kleinhofs A, Kilian A, Saghai Maroof MA, Biyashev RM, Hayes P, Chen FQ, Lapitan N, Fenwick A, Blake TK, Kanazin V, Ananiev E, Dahleen L, Kudrna D, Bollinger J, Knapp SJ, Liu B, Sorrells M, Heun M, Franckowiak JD, Hoffman D, Skadsen R, Steffenson BJ (1993)

A molecular, isozyme and morphological map of the barley (*Hordeum vulgare*) genome. Theor Appl Genet 86:705–712

Knapp SJ, Bridges WC (1990) Using molecular markers to estimate quantitative trait locus parameters; power and genetic variances for unreplicated and replicated progeny. Genetics 126:769–777

Kojima S, Takahashi Y, Kobayashi Y, Monna L, Sasaki T, Araki T, Yano M (2002) *Hd3a*, a rice ortholog of the Arabidopsis FT gene, promotes transition to flowering downstream of *Hd1* under short-day conditions. Plant Cell Physiol 43:1096–1105

Konishi S, Izawa T, Lin SY, Ebana K, Fukuta Y, Sasaki T, Yano M (2006) An SNP caused loss of seed shattering during rice domestication. Science 312:1392–1396

Korol AB, Ronin YI, Kirzhner VM (1995) Interval mapping of quantitative trait loci employing correlated trait complexes. Genetics 140:1137–1147

Korol AB, Ronin YI, Nevo E, Hays PM (1998) Multi-interval mapping of correlated trait complexes. Heredity 80:273–284

Kou Y, Wang S (2010) Broad-spectrum and durability: understanding of quantitative disease resistance. Curr Opin Plant Biol 13:181–185

Kover PX, Valdar W, Trakalo J, Scarcelli N, Ehrenreich IM, Purugganan MD, Durrant C, Mott R (2009) Multiparent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana*. PLoS Genet 5:e1000551

Kraakman ATW, Martínez F, Mussiraliev B, van Eeuwijk FA, Niks RE (2006) Linkage disequilibrium mapping of morphological, resistance, and other agronomically relevant traits in modern spring barley cultivars. Mol Breeding 17:41–58

Kraakman ATW, Niks RE, Van den Berg PMMM, Stam P, Van Eeuwijk FA (2004) Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. Genetics 168:435–446

Krattinger SG, Lagudah ES, Spielmeyer W, Singh RP, Huerta-Espino J, McFadden H, Bossolini E, Selter LL, Keller B (2009a) A putative ABC transporter confers durable resistance to multiple fungal pathogens in wheat. Science 323:1360–1362

Krattinger S, Wicker T, Keller B (2009b) Map-based cloning of genes in triticeae (wheat and barley). In: Feuillet C, Muehlbauer GJ (eds) Genetics and genomics of the triticeae, plant genetics and genomics: crops and model 7. Springer, Berlin, pp 337–357

Krill AM, Kirst M, Kochian LV, Buckler ES, Hoekenga OA (2010) Association and linkage analysis of aluminum tolerance genes in maize. PLoS ONE 5:e9958

Kulwal PL, Ishikawa G, Benscher D, Feng Z, Yu L-X, Jadhav A, Mehetre S, Sorrells ME (2012) Association mapping for pre-harvest sprouting resistance in white winter wheat. Theor Appl Genet 125:793–805

Kulwal PL, Singh R, Balyan HS, Gupta PK (2004) Genetic basis of pre-harvest sprouting tolerance using single-locus and two-locus QTL analyses in bread wheat. Funct Integr Genomics 4:94–101

Kulwal PL, Roy JK, Balyan HS, Gupta PK (2003) QTL analysis for growth and leaf characters in bread wheat. Plant Sci 164:267–277

Kulwal PL, Kumar N, Kumar A, Gupta RK, Balyan HS, Gupta PK (2005) Gene networks in hexaploid wheat: interacting quantitative trait loci for grain protein content. Funct Integr Genomics 5:254–259

Kumar N, Kulwal PL, Balyan HS, Gupta PK (2007) QTL analysis for yield and yield contributing traits in two mapping populations of bread wheat. Mol Breeding 19:163–177

Kumar N, Kulwal PL, Gaur A, Tyagi AK, Khurana JP, Khurana P, Balyan HS, Gupta PK (2006) QTL analysis for grain weight in bread wheat. Euphytica 151:135–144

Kump KL, Bradbury PJ, Wisser RJ, Buckler ES, Belcher AR, Oropeza-Rosas MA, Zwonitzer JC, Kresovich S, McMullen MD, Ware D, Balint-Kurti PJ, Holland JB (2011) Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. Nat Genet 43:163–169

Lander ES, Botstein D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185–199

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174–181

Langridge P, Karakousis A, Collins N, Kretschmer J, Manning S (1995) A consensus linkage map of barley. Mol Breeding 1:389–395

Larièpe A, Mangin B, Jasson S, Combes V, Dumas F, Jamin P, Lariagon C, Jolivot D, Madur D, Fiévet JB, Gallais A, Dubreuil P, Charcosset A, Moreau L (2012) The genetic basis of heterosis: multiparental quantitative trait loci mapping reveals contrasted levels of apparent overdominance among traits of agronomical interest in maize (*Zea mays* L.). Genetics 190:795–811

Lee M, Sharopova N, Beavis WD, Grant D, Katt M, Blair D, Hallauer A (2002) Expanding the genetic map of maize with the intermated B73 3 Mo17 (IBM) population. Plant Mol Biol 48:453–461

Lehmensiek A, Bovill W, Wenzl P, Langridge P, Rudi A (2009) Genetics and genomics of the triticeae. In: Feuillet C, Muehlbauer GJ (eds) Plant genetics and genomics: crops and models 7, DOI 10.1007/978-0-387-77489-3_7

Li C, Zhou A, Sang T (2006a) Rice domestication by reducing shattering. Science 311:1936–1939

Li ZK, Pinson SRM, Park WD, Paterson AH, Stansel JW (1997) Epistasis for three yield components in rice *Oryza sativa* L. Genetics 145:453–465

Li ZK, Yu SB, Lafitte HR, Huang N, Courtois B, Hittalmani S, Vijayakumar CHM, Liu GF, Wang GC, Shashidhar HE, Zhuang JY, Zheng KL, Singh VP, Sidhu JS, Srivantaneeyakul S, Khush GS (2003) QTL× environment interactions in rice. I. Heading date and plant height. Theor Appl Genet 108:141–153

Li C, Zhou A, Sang T (2006b) Genetic analysis of rice domestication syndrome with the wild annual species, *Oryza nivara*. New Phytol 170:185–194

Li H, Hearne S, Banziger M, Li Z, Wang J (2010) Statistical properties of QTL linkage mapping in biparental genetic populations. Heredity 105:257–267

Li H, Ribaut J-M, Li Z, Wang J (2008) Inclusive composite interval mapping (ICIM) for digenic epistasis of quantitative traits in biparental populations. Theor Appl Genet 116:243–260

Li H, Ye G, Wang J (2007) A modified algorithm for the improvement of composite interval mapping. Genetics 175:361–374

Li J, Wang S, Zeng ZB (2006) Multiple-interval mapping for ordinal traits. Genetics 173:1649–1663

Li JZ, Zhang ZW, Li YL, Wang QL, Zhou YG (2011) QTL consistency and meta-analysis for grain yield components in three generations in maize. Theor Appl Genet 122:771–782

Li X, Yan W, Agrama H, Jia L et al (2012) Unraveling the complex trait of harvest index with association mapping in rice (*Oryza sativa* L.). PLoS ONE 7:e29350

Lincoln S, Daly M, Lander E (1993) Mapping genes controlling quantitative traits using MAPMAKER/QTL. Version 1.1, 2nd edn. Whitehead Institute for Biomedical Research, Technical report

Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, Gore M, Buckler ES, Zhang Z (2011) User manual of GAPIT: genome association and prediction integrated tool. http://www.maizegenetics.net/gapit

Liu BH (1998) Statistical genomics: linkage mapping and QTL analysis. CRC Press, Boca Raton

Liu S, Pumphrey MO, Gill BS, Trick HN, Zhang JX, Dolezel J, Chalhoub B, Anderson JA (2008) Toward positional cloning of *Fhb1*, a major QTL for Fusarium head blight resistance in wheat. Cereal Res Commun 36:195–201

Liu R, Zhang H, Zhao P, Zhang Z, Liang W, Tian Z, Zheng Y (2012) Mining of candidate maize genes for nitrogen use efficiency by integrating gene expression and QTL data. Plant Mol Biol Rep 30:297–308

Liu SC, Kowalski SP, Lan TH, Feldmann KA, Paterson AH (1996) Genome-wide high-resolution mapping by recurrent intermating using Arabidopsis thaliana as a model. Genetics 142:247–258

Loffler M, Schon CC, Miedaner T (2009) Revealing the genetic architecture of FHB resistance in hexaploid wheat (*Triticum aestivum* L.) by QTL meta-analysis. Mol Breeding 23:473–488

Lu C, Shen L, Tan Z, Xu Y, He P, Chen Y, Zhu L (1996) Comparative mapping of QTLs for agronomic traits of rice across environments using a doubled haploid population. Theor Appl Genet 93:1211–1217

Lu Y, Zhang S, Shah T, Xie C, Hao Z, Li X, Farkhari M, Ribaut J-M, Cao M, Rong T, Xu Y (2010) Joint linkage–linkage disequilibrium mapping is a powerful approach to detecting quantitative trait loci underlying drought tolerance in maize. Proc Natl Acad Sci USA 107:19585–19590

Ma XQ, Tang JH, Teng WT, Yan JB, Meng YJ, Li JS (2007) Epistatic interaction is an important genetic basis of grain yield and its components in maize. Mol Breeding 20:41–51

Mackay TFC (2001) The genetic architecture of quantitative traits. Annu Rev Genet 33:303–339

Malosetti M, Boer MP, Bink MCAM, van Eeuwijk FA (2006) Multi-trait QTL analysis based on mixed models with parsimonious covariance matrices. In: Proceedings of the 8th world congress on genetics applied to livestock production, August 13–18, Belo Horizonte, MG, Brasil. http://www.wcgalp8.org.br/wcgalp8. Article 25–04

Malosetti M, Ribaut JM, Vargas M, Crossa J, van Eeuwijk FA (2008) A multi-trait multi-environment QTL mixed model with an application to drought and nitrogen stress trials in maize (*Zea mays* L.). Euphytica 161:241–257

Malosetti M, Voltas J, Romagosa I, Ullrich SE, van Eeuwijk FA (2004) Mixed models including environmental covariables for studying QTL by environment interaction. Euphytica 137:139–145

Manly KF, Olson JM (1999) Overview of QTL mapping software and introduction to map manager QTL. Mamm Genome 10:327–334

Mauricio R (2001) Mapping quantitative trait loci in plants: uses and caveats for evolutionary biology. Nat Rev Genet 2:370–381

Martinez O, Curnow RN (1992) Estimating the locations and the sizes of the effects of quantitative trait loci using flanking markers. Theor Appl Genet 85:480–488

Massman J, Cooper B, Horsley R, Neate S, Dill-Macky R, Chao S, Dong Y, Schwarz P, Muehlbauer GJ, Smith KP (2011) Genome-wide association mapping of Fusarium head blight resistance in contemporary barley breeding germplasm. Mol Breeding 27:439–454

McMullen MD, Stephen K, Hector SV, Peter B et al (2009) Genetic properties of the maize nested association mapping population. Science 325:737–740

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

Meuwissen THE, Karlsen A, Lien S, Olsaker I, Goddard ME (2002) Fine mapping of a quantitative trait locus for twinning rate using combined linkage and linkage disequilibrium mapping. Genetics 161:373–379

Meuwissen THE, Goddard ME (2004) Mapping multiple QTL using linkage disequilibrium and linkage analysis information and multitrait data. Genet Sel Evol 36:261–279

Meyer RC, Steinfath M, Lisec J et al (2007) The metabolic signature related to high plant growth rate in *Arabidopsis thaliana*. Proc Natl Acad Sci USA 104:4759–4764

Michelmore WR, Paran I, Kesseli RV (1991) Identification of marker linked to diseases resistance genes by bulked segregant analysis, a rapid method to detect the markers in specific genetic region by using the segregating population. Proc Natl Acad Sci USA 88:9828–9832

Michelmore RW (2003) The impact zone: genomics and breeding for durable disease resistance. Curr Opin Plant Biol 6:397–404

Miedaner T, Wurschum T, Maurer HP, Korzun V, Ebmeyer E, Reif JC (2011) Association mapping for Fusarium head blight resistance in European soft winter wheat. Mol Breeding 28:647–655

Mir RR, Kumar N, Jaiswal V, Girdharwal N, Prasad M, Balyan HS, Gupta PK (2012) Genetic dissection of grain weight in bread wheat through quantitative trait locus interval and association mapping. Mol Breeding 29:963–972

Mohan A, Kulwal PL, Singh S, Kumar V, Mir RR, Kumar J, Prasad M, Balyan HS, Gupta PK (2009) Genome-wide QTL analysis for pre-harvest sprouting tolerance in bread wheat. Euphytica 168:319–329

Moncada P, Martinez CP, Borrero J, Chatel M, Gauch Jr-H, Guimareaes E, Tohmem J, McCouch SR (2001) Quantitative trait loci for yield and yield components in an *Oryza sativa* × *Oryza rufipogon* BC_2F_2 population evaluated in an upland environment. Theor Appl Genet 102:41–52

Moose SP, Mumm RH (2008) Molecular plant breeding as the foundation for 21st century crop improvement. Plant Physiol 147:969–977

Morrell PL, Buckler ES, Ross-Ibarra J (2012) Crop genomics: advances and applications. Nat Rev Genet 13:85–96

Myles S, Peiffer J, Brown PJ, Ersoz ES, Zhang Z, Costich DE, Buckler ES (2009) Association mapping: critical considerations shift from genotyping to experimental design. Plant Cell 21:2194–2202

Narasimhamoorthy B, Gill BS, Fritz AK, Nelson JC, Brown-Guedira GL (2006) Advanced backcross QTL analysis of a hard winter wheat × synthetic wheat population. Theor Appl Genet 112:787–796

Nelson JC (1997) QGENE: software for marker-based genomic analysis and breeding. Mol Breeding 3:239–245

Neumann K, Kobiljski B, Dencic S, Varshney RK, Borner A (2011) Genome-wide association mapping: a case study in bread wheat (*Triticum aestivum* L.). Mol Breeding 27:37–58

Nishimura A, Ashikari M, Lin S, Takashi T, Angeles ER, Yamamoto T, Matsuoka M (2005) Isolation of a rice regeneration quantitative trait loci gene and its application to transformation systems. Proc Natl Acad Sci USA 102:11940–11944

Nordborg M, Weigel D (2008) Next-generation genetics in plants. Nature 456:720–723

Norton GJ, Aitkenhead MJ, Khowaja FS, Whalley WR, Price AH (2008) A bioinformatic and transcriptomic approach to identifying positional candidate genes without fine mapping: an example using rice root-growth QTLs. Genomics 92:344–352

Peleg Z, Fahima T, Korol AB, Abbo S, Saranga Y (2011) Genetic analysis of wheat domestication and evolution under domestication. J Expt Bot 62:5051–5061

Peng J, Ronin Y, Fahima T, Roder MS, Li Y, Nevo E, Korol A (2003) Domestication quantitative trait loci in Triticum dicoccoides, the progenitor of wheat. Proc Natl Acad Sci USA 100:2489–2494

Perretant MR, Cadalen T, Charmet G, Sourdille P, Nicolas P, Boeuf C, Tixier MH, Branlard G, Bernard S (2000) QTL analysis of bread-making quality in wheat using a doubled haploid population. Theor Appl Genet 100:1167–1175

Piepho HP (2000) A mixed-model approach to mapping quantitative trait loci in barley on the basis of multiple environment data. Genetics 156:2043–2050

Pillen K, Zacharias A, Leon J (2003) Advanced backcross QTL analysis in barley (*Hordeum vulgare* L.). Theor Appl Genet 105:1253–125

Podlich DW, Winkler CR, Cooper M (2004) Mapping as you go: an effective approach for marker-assisted selection of complex traits. Crop Sci 44:1560–1571

Poland JA, Nelson RJ (2011) In the eye of the beholder: the effect of rater variability and different rating scales on QTL mapping. Phytopath 101:290–298

Poland JA, Balint-Kurti PJ, Wisser RJ, Pratt RC, Nelson RJ (2009) Shades of gray: the world of quantitative disease resistance. Trends Plant Sci 11:21–29

Poland JA, Bradbury PJ, Buckler ES, Nelson RJ (2011) Genome-wide nested association mapping of quantitative resistance to northern leaf blight in maize. Proc Natl Acad Sci USA 108:6893–6898

Poland JA, Brown PJ, Sorrells ME, Jannink J-L (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS ONE 7:e32253

Potokina E, Druka A, Luo Z, Moscou M, Wise R, Waugh R, Kearsey M (2008a) Tissue-dependent limited pleiotropy affects gene expression in barley. Plant J 56:287–296

Potokina E, Druka A, Luo Z, Wise R, Waugh R, Kearsey M (2008b) Gene expression quantitative trait locus analysis of 16000 barley genes reveals a complex pattern of genome-wide transcriptional regulation. The Plant J 53:90–101

Pozzi C, Rossini L, Vecchietti A, Salamini F (2004) Gene and genome changes during domestication of cereals. In: Gupta PK, Varshney RK (eds) Cereal genomics. Springer, Berlin, pp 165–198

Prasad M, Kumar N, Kulwal PL, Röder M, Balyan HS, Dhaliwal HS, Gupta PK (2003) QTL analysis for grain protein content using SSR markers and validation studies using NILs in bread wheat. Theor Appl Genet 106:659–667

Prasad M, Varshney RK, Kumar A, Balyan HS, Sharma PC, Edwards KJ, Singh H, Dhaliwal HS, Roy JK, Gupta PK (1999) A microsatellite marker associated with a QTL for grain protein content on chromosome arm 2DL of bread wheat. Theor Appl Genet 99:341–345

Price AH (2006) Believe it or not, QTLs are accurate. Trends Plant Sci 11:213–216

Price A, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38:904–909

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

Purugganan MD, Fuller DQ (2009) The nature of selection during plant domestication. Nature 457:843–848

Quraishi UM, Murat F, Abrouk M, Pont C, Confolent C, Oury FX, Ward J, Boros D, Gebruers K, Delcour JA, Courtin CM, Bedo Z, Saulnier L, Guillon F, Balzergue S, Shewry PR, Feuillet C, Charmet G, Salse J (2011) Combined meta-genomics analyses unravel candidate genes for the grain dietary fiber content in bread wheat (*Triticum aestivum* L.). Funct Integr Genomics 11:71–83

Quarrie SA, Laurie DA, Zhu J, Lebreton C, Semikhodskii A, Steed A, Witsenboer H, Calestani C (1997) QTL analysis to study the association between leaf size and abscisic acid accumulation in droughted rice leaves and comparisons across cereals. Plant Mol Biol 35:155–165

Quarrie SA, Vesna LJ, Dragan K, Andy S, Sofija P (1999) Bulk segregant analysis with molecular markers and its use for improving drought resistance in maize. J Expt Bot 50:1299–1306

Rakshit S, Rakshit A, Patil JV (2012) Multiparent intercross populations in analysis of quantitative traits. J Genet 91:111–117

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

Rao S, Xu S (1998) Mapping quantitative trait loci for ordered categorical traits in four-way crosses. Heredity 81:214–224

Rehman Arif MA, Neumann K, Nagel M, Kobiljski B, Lohwasser U, Borner A (2012) An association mapping analysis of dormancy and pre-harvest sprouting in wheat. Euphytica 188:409–417

Reif JC, Gowda M, Maurer HP, Longin CFH, Korzun V, Ebmeyer E, Bothe R, Pietsch C, Wurschum T (2011) Association mapping for quality traits in soft winter wheat. Theor Appl Genet 122:961–970

Ren ZH, Gao JP, Li LG, Cai XL, Huang W, Chao DY, Zhu MZ, Wang ZY, Luan S, Lin HX (2005) A rice quantitative trait locus for salt tolerance encodes a sodium transporter. Nat Genet 37:1141–1146

Rockman MV, Kruglyak L (2008) Breeding designs for recombinant inbred advanced intercross lines. Genetics 179:1069–1078

Rode J, Ahlemeyer J, Friedt W, Ordon F (2012) Identification of marker-trait associations in the German winter barley breeding gene pool (*Hordeum vulgare* L.). Mol Breeding 30:831–843

Röder MS, Korzun V, Wendehake K, Plaschke J, Tixier M-H, Leroy P, Ganal MW (1998) A microsatellite map of wheat. Genetics 149:2007–2023

Rong J, Feltus FA, Waghmare VN, Pierce GJ, Chee PW, Draye X, Saranga Y, Wright RJ, Wilkins TA, May OL, Smith CW, Gannaway JR, Wendel JF, Paterson AH (2007) Meta-analysis of polyploid cotton QTL shows unequal contributions of subgenomes to a complex network of genes and gene clusters implicated in lint fiber development. Genetics 176:2577–2588

Ross-Ibarra J (2005) Quantitative trait loci and the study of plant domestication. Genetica 123:197–204

Ross-Ibarra J, Morrell PL, Gaut BS (2007) Plant domestication, a unique opportunity to identify the genetic basis of adaptation. Proc Natl Acad Sci USA 104:8641–8648

Rousset M, Bonnin I, Remoue C et al (2011) Deciphering the genetics of flowering time by an association study on candidate genes in bread wheat (*Triticum aestivum* L.). Theor Appl Genet 123:907–926

Rostoks N, Ramsay L, MacKenzie K, Cardle L et al (2006) Recent history of artificial outcrossing facilitates whole-genome association mapping in elite inbred crop varieties. Proc Natl Acad Sci USA 103:18656–18661

Rowe HC, Hansen BG, Halkier BA, Kliebenstein DJ (2008) Biochemical networks and epistasis shape the *Arabidopsis thaliana* metabolome. Plant Cell 20:1199–1216

Roy JK, Smith KP, Muehlbauer GJ, Chao S, Close TJ, Steffenson BJ (2010) Association mapping of spot blotch resistance in wild barley. Mol Breeding 26:243–256

Salunkhe AS, Poornima R, Prince KS, Kanagaraj P, Sheeba JA, Amudha K, Suji KK, Senthil A, Babu RC (2011) Fine mapping QTL for drought resistance traits in rice (*Oryza sativa* L.) using bulk segregant analysis. Mol Biotechnol 49:90–95

Salvi S, Tuberosa R (2005) To clone or not to clone plant QTLs: present and future challenges. Trends Plant Sci 10:297–304

Salvi S, Tuberosa R (2007) Cloning QTLs in plants. In: Varshney RK, Tuberosa R (eds) Genomics-assisted crop improvement, vol 1., Genomics approaches and platformsSpringer, Berlin, pp 207–225

Salvi S, Sponza G, Morgante M, Tomes D, Xiaomu NX, Fengler KA, Meeley R, Ananiev EV, Svitashev S, Bruggemann E, Li B, Haney CF, Radovic S, Zaina G, Rafalski J-A, Tingey SV, Miao G-H, Phillips RL, Tuberosa R (2007) Conserved non-coding genomic sequences controlling flowering time differences in maize. Proc Natl Acad Sci USA 104:11376–11381

Satagopan JM, Yandell BS, Newton MA, Osborn TC (1996) A Bayesian approach to detect quantitative trait loci using Markov Chain Monte Carlo. Genetics 144:805–816

Sax K (1923) The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. Genetics 8:552–560

Schadt EE, Monks SA, Drake TA, Lusis AJ, Che N, Colinayo V, Ruff TG, Milligan SB, Lamb JR, Cavet G, Linsley PS, Mao M, Stoughton RB, Friend SH (2003) Genetics of gene expression surveyed in maize, mouse and man. Nature 422:297–302

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

Seaton G, Haley CS, Knott SA, Kearsey M, Visscher PM (2002) QTL Express: mapping quantitative trait loci in simple and complex pedigrees. Bioinformatics 18:339–34

Servin B, Stephens M (2007) Imputation-based analysis of association studies: candidate regions and quantitative traits. PLoS Genet 3:e114

Setter TL, Yan J, Warburton M, Ribaut J-M, Xu Y, Mark S, Buckler ES, Zhang Z, Gore MA (2011) Genetic association mapping identifies single nucleotide polymorphisms in genes that affect abscisic acid levels in maize floral tissues during drought. J Expt Bot 62:701–716

Sharma S, Xu S, Ehdaie B, Hoops A, Close TJ, Lukaszewski AJ, Waines JG (2011) Dissection of QTL effects for root traits using a chromosome arm-specific mapping population in bread wheat. Theor Appl Genet 122:759–769

Shoemaker JS, Painter IS, Weir BS (1999) Bayesian statistics in genetics: a guide for the uninitiated. Trends Genet 15:354–358

Sillanpaa MJ, Arjas E (1998) Bayesian mapping of multiple quantitative trait loci from incomplete inbred line cross data. Genetics 148:1373–1388

Song X-J, Huang W, Shi M, Zhu M-Z, Lin H-X (2007) A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. Nat Genet 39:623–630

Stich B, Melchinger AE (2010) An introduction to association mapping in plants. CAB Reviews 5:039

Stich B, Piepho H-P, Schulz B, Melchinger AE (2008) Multi-trait association mapping in sugar beet (*Beta vulgaris* L.). Theor Appl Genet 117:947–954

Stracke S, Haseneyer G, Veyrieras JB, Geiger HH, Sauer S, Graner A, Piepho HP (2009) Association mapping reveals gene action and interactions in the determination of flowering time in barley. Theor Appl Genet 118:259–273

Swanson-Wagner RA, DeCook R, Jia Y, Bancroft T, Ji T, Zhao X, Nettleton D, Schnable PS (2009) Paternal dominance of trans-eQTL influences gene expression patterns in maize hybrids. Science 326:1118–1120

St.Clair DA (2010) Quantitative disease resistance and quantitative resistance loci in breeding. Annu Rev Phytopath 48:247–268

Sugimoto K, Takeuchi Y, Ebana K, Miyao A, Hirochika H, Hara N, Ishiyama K, Kobayashi M, Ban Y, Hattori T, Yano M (2010) Molecular cloning of Sdr4, a regulator involved in seed dormancy and domestication of rice. Proc Natl Acad Sci USA 107:5792–5797

Sun D, Ren W, Sun G, Peng J (2011) Molecular diversity and association mapping of quantitative traits in Tibetan wild and worldwide originated barley (*Hordeum vulgare* L.) germplasm. Euphytica 178:31–43

Szalma SJ, Buckler ES, Snook ME, McMullen MD (2005) Association analysis of candidate genes for maysin and chlorogenic acid accumulation in maize silks. Theor Appl Genet 110:1324–1333

Tanhuanpää P, Kalendar R, Schulman AH, Kiviharju E (2008) The first doubled haploid linkage map for cultivated oat. Genome 51:560–569

Tanhuanpää P, Manninen O, Kiviharju E (2010) QTLs for important breeding characteristics in the doubled haploid oat progeny. Genome 53:482–493

Takai T, Yoshimichi F, Tatsuhiko S, Takeshi H (2005) Time-related mapping of quantitative trait loci controlling grain-filling in rice (*Oryza sativa* L.). J Expt Bot 56:2107–2118

Tanksley SD (1993) Mapping polygenes. Annu Rev Genet 27:205–233

Tanksley SD, Grandillo S, Fulton TM, Zamir D, Eshed Y, Petiard V, Lopez J, Beck BT (1996) Advanced backcross QTL analysis in a cross between an elite processing line of tomato and its wild relative *L. pimpinellifolium*. Theor Appl Genet 92:213–222

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

Tang J, Yan J, Ma X, Teng W, Wu W, Dai J, Dhillon BS, Melchinger AE, Li J (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

Takahashi Y, Shomura A, Sasaki T, Yano M (2001) Hd6, a rice quantitative trait locus involved in photoperiod sensitivity, encodes the a-subunit of protein kinase CK2. Proc Natl Acad Sci USA 98:7922–7927

Tinker NA, Mather DE (1995) MQTL: software for simplified composite interval mapping of QTL in multiple environments. J Quant Trait Loci 1:2

Thomson MJ, Tai TH, McClung AM, Lai X-H, Hinga ME, Lobos KB, Xu Y, Martinez CP, McCouch SR (2003) Mapping quantitative trait loci for yield, yield components and morphological traits in an advanced backcross population between *Oryza rufipogon* and the *Oryza sativa* cultivar Jefferson. Theor Appl Genet 107:479–493

Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ES (2001) *Dwarf8* polymorphisms associate with variation in flowering time. Nat Genet 28:286–289

Tian F, Bradbury PJ, Brown PJ, Hung H, Sun Q, Flint-Garcia S, Rocheford TR, McMullen MD, Holland JB, Buckler ES (2011) Genome-wide association study of leaf architecture in the maize nested association mapping population. Nat Genet 43:159–162

Tommasini L, Schnurbusch T, Fossati D, Mascher F, Keller B (2007) Association mapping of Stagonospora nodorum blotch resistance in modern European winter wheat varieties. Theor Appl Genet 115:697–708

Truntzler M, Barrière Y, Sawkins MC, Lespinasse D, Betran J, Charcosset A, Moreau L (2010) Meta-analysis of QTL involved in silage quality of maize and comparison with the position of candidate genes. Theor Appl Genet 121:1465–1482

Tuberosa R, Salvi S (2004) QTLs and genes for tolerance to abiotic stress in cereals. In: Gupta PK, Varshney RK (eds) Cereal genomics. Kluwer Academic Publishers, The Netherlands, pp 253–315

Tuberosa R, Salvi S (2007) From QTLs to genes controlling root traits in maize. In: Spiertz JHJ, Struik PC, van Laar HH (eds) Scale and complexity in plant systems research: gene-plant-crop relations. Springer, Berlin, pp 15–24

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

Tyagi S, Gupta PK (2012) Meta-analysis of QTLs involved in pre-harvest sprouting tolerance and dormancy in bread wheat. Triticeae Genomics Genet 3:9–24

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

Ueda T, Sato T, Hidema J, Hirouchi T, Yamamoto K, Kumagai T, Yano M (2005) *qUVR-10*, a major quantitative trait locus for ultraviolet-B resistance in rice, encodes cyclobutane pyrimidine dimer photolyase. Genetics 171:1941–1950

Utz H, Melchinger A (1996) PLABQTL: a program for composite interval mapping of QTL. J Quant Trait Loci 2:1

van Dyk MM, Kullan ARK, Mizrachi E, Hefer CA, van Rensburg LJ, Tschaplinski TJ, Cushman KC, Engle NE, Tuskan GA, Jones N, Kanzler A, Myburg AA (2011) Genetic dissection of transcript, metabolite, growth and wood property traits in an F2 pseudo-backcross pedigree of *Eucalyptus grandis* x *E. urophylla*. BMC Proc 5:O7

van Eeuwijk FA, Bink MCAM, Chenu K, Chapman SC (2010) Detection and use of QTL for complex traits in multiple environments. Cur Opin Plant Biol 13:193–205

van Ooijen JW, Maliepaard C (1996) MapQTL™ version 3.0: software for the calculation of QTL positions on genetic maps. Plant Research International, Wageningen

Varshney RK, Paulo MJ, Grando S, van Eeuwijk FA, Keizer LCP, Guo P, Ceccarelli S, Kilian A, Baum M, Graner A (2012) Genome wide association analyses for drought tolerance related traits in barley (*Hordeum vulgare* L.). Field Crops Res 126:171–180

Varshney RK, Prasad M, Roy JK, Kumar N, Singh H, Dhaliwal HS, Balyan HS, Gupta PK (2000) Identification of eight chromosomes and a microsatellite marker on 1AS associated with QTLs for grain weight in bread wheat. Theor Appl Genet 100:1290–1294

Veyrieras J-B, Goffinet B, Charcosset A (2007) MetaQTL: a package of new computational methods for the meta-analysis of QTL mapping experiments. BMC Bioinfor 8:49

von Zitzewitz J, Cuesta-Marcos A, Condon F, Castro AJ, Chao S, Corey A, Filichkin T, Fisk SP, Gutierrez L, Haggard K, Karsai I, Muehlbauer GJ, Smith KP, Veisz O, Hayes PM (2011) The genetics of winterhardiness in barley: perspectives from genome-wide association mapping. Plant Genome 4:76–91

Wang DL, Zhu J, Li ZK, Paterson AH (1999) Mapping QTLs with epistatic effects and QTL × environment interactions by mixed linear model approaches. Theor Appl Genet 99:1255–1264

Wang H, Wagler T, Li B, Zhao Q, Vigouroux Y, Faller M, Bomblies K, Lukens L, Doebley J (2005) The origin of the naked grains of maize. Nature 436:714–719

Wang M, Jiang N, Jia T, Leach L, Cockram J, Waugh R, Ramsay L, Thomas B, Luo Z (2012) Genome-wide association mapping of agronomic and morphologic traits in highly structured populations of barley cultivars. Theor Appl Genet 124:233–246

Wang S, Basten CJ, Zeng Z-B (2011) Windows QTL Cartographer 2.5. Department of Statistics, North Carolina State University, Raleigh, NC. (http://statgen.ncsu.edu/qtlcart/WQTLCart.htm)

Wang Y, Yao J, Zhang ZF, Zheng YL (2006b) The comparative analysis based on maize integrated QTL map and meta-analysis of plant height QTLs. Chin Sci Bull 51:2219–2230

Wang Y-M, Kong W-Q, Tang Z-X, Lu X, Xu C-W (2009) Bayesian statistics-based multiple interval mapping of qtl controlling endosperm traits in cereals. Acta Agron Sinica 35:1569–1575

Wen W, Mei H, Feng F, Yu S, Huang Z, Wu J, Chen L, Xu X, Luo L (2009) Population structure and association mapping on chromosome 7 using a diverse panel of Chinese germplasm of rice (*Oryza sativa* L.). Theor Appl Genet 119:459–470

Whitkus R, Doebley J, Lee M (1992) Comparative genome mapping of sorghum and maize. Genetics 132:1119–1130

Wisser RJ, Sun Q, Hulbert SH, Kresovich S, Nelson RJ (2005) Identification and characterization of regions of the rice genome associated with broad-spectrum, quantitative disease resistance. Genetics 169:2277–2293

Wright SI, Bi IV, Schroeder SG, Yamasaki M, Doebley JF, McMullen MD, Gaut BS (2005) The effects of artificial selection on the maize genome. Science 308:1310–1314

Wu W-R, Li W-M, Tang D-Z, Lu H-R, Worland AJ (1999) Time-related mapping of quantitative trait loci underlying tiller number in rice. Genetics 151:297–303

Wu R, Lin M (2006) Functional mapping—how to map and study the genetic architecture of dynamic complex traits. Nat Rev Genet 7:229–237

Wu D, Qiu L, Xu L, Ye L, Chen M et al (2011) Genetic variation of *HvCBF* genes and their association with salinity tolerance in Tibetan annual wild barley. PLoS ONE 6:e22938

Wu RL, Zeng Z-B (2001) Joint linkage and linkage disequilibrium mapping in natural populations. Genetics 157:899–909

Wu R, Chang-Xing M, George C (2002) Joint linkage and linkage disequilibrium mapping of quantitative trait loci in natural populations. Genetics 160:779–792

Wurschum T (2012) Mapping QTL for agronomic traits in breeding populations. Theor Appl Genet 125:201–210

Xie C, Gessler DDG, Xu S (1998) Combining different line crosses for mapping quantitative trait loci using the identical by descent-based variance component method. Genetics 149:1139–1146

Xu S (2003a) Estimating polygenic effects using markers of the entire genome. Genetics 163:789–801

Xu S (2003b) Theoretical basis of the Beavis effect. Genetics 165:2259–2268

Xu S, Jia Z (2007) Genome wide analysis of epistatic effects for quantitative traits in barley. Genetics 175:1955–1963

Xu Y, Zhu L, Xiao J, Huang N, McCouch SR (1997) Chromosomal regions associated with segregation distortion of molecular markers in F2, backcross, doubled haploid, and recombinant inbred populations in rice (*Oryza sativa* L.). Mol Gen Genet 253:535–545

Xu K, Xu X, Fukao T, Canlas P, Maghirang-Rodriguez R, Heuer S, Ismail MA, Bailey-Serres J, Ronald PC, Mackill DJ (2006) *Sub1A* is an ethylene response factor-like gene that confers submergence tolerance to rice. Nature 442:705–708

Xu Y (2010) Molecular plant breeding. CABI, Oxfordshire

Yan WG, Li Y, Agrama HA, Luo D, Gao F, Lu X, Ren G (2009) Association mapping of stigma and spikelet characteristics in rice (*Oryza sativa* L.). Mol Breeding 24:277–292

Yano M, Katayose Y, Ashikari M, Yamanouchi U, Monna L, Fuse T, Baba T, Yamamoto K, Umehara Y, Nagamura Y, Sasaki T (2000) *Hd1*, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the Arabidopsis flowering time gene CONSTANS. Plant Cell 12:2473–2484

Yao J, Wang L, Liu L, Zhao C, Zheng Y (2009) Association mapping of agronomic traits on chromosome 2A of wheat. Genetica 137:67–75

Yi N, George V, Allison DB (2003a) Stochastic search variable selection for identifying multiple quantitative trait loci. Genetics 164:1129–1138

Yi N, Xu S, Allison DB (2003b) Bayesian model choice and search strategy for mapping interacting quantitative trait loci. Genetics 165:867–883

Yi N, Banerjee S, Pomp D, Yandell BS (2007) Bayesian mapping of genomewide interacting quantitative trait loci for ordinal traits. Genetics 176:1855–1864

Yi N, Xu Z (2002) Linkage analysis of quantitative trait loci in multiple line crosses. Genetica 114:217–230

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

Young ND (1996) QTL mapping and quantitative disease resistance in plants. Ann Rev Phytopath 34:479–501

Youens-Clark K, Buckler E, Casstevens T, Chen C, DeClerck G, Derwent P, Dharmawardhana P, Jaiswal P, Kersey P, Karthikeyan AS, Lu J, McCouch SR, Ren L, Spooner W, Stein JC, Thomason J, Wei S, Ware D (2011) Gramene database in 2010: updates and extensions. Nucleic Acids Res 39:D1085–D1094

Yu J, Holland JB, McMullen MD, Buckler ES (2008) Genetic design and statistical power of nested association mapping in maize. Genetics 178:539–551

Yu J, Buckler ES (2006) Genetic association mapping and genome organization of maize. Curr Opin Biotechnol 17:155–160

Yu J, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2005) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203–208

Yu L-X, Lorenz A, Rutkoski J, Singh RP, Bhavani S, Huerta-Espino J, Sorrells ME (2011) Association mapping and gene–gene interaction for stem rust resistance in CIMMYT spring wheat germplasm. Theor Appl Genet 123:1257–1268

Yu L-X, Morgounov A, Wanyera R et al (2012) Identification of Ug99 stem rust resistance loci in winter wheat germplasm using genome-wide association analysis. Theor Appl Genet 125:749–758

Zeng Z-B (1994) Precision mapping of quantitative trait loci. Genetics 136:1457–1468

Zeng ZB, Kao CH, Basten CJ (1999) Estimating the genetic architecture of quantitative traits. Genet Res 74:279–289

Zhang Y-M, Gai J (2009) Methodologies for segregation analysis and QTL mapping in plants. Genetica 136:311–318

Zhang LY, Liu DC, Guo XL, Yang WL, Sun JZ, Wang DW, Zhang A (2010) Genomic distribution of quantitative trait loci for yield and yield-related traits in common wheat. J Integr Plant Biol 52:996–1007

Zhang YM, Mao Y, Xie C, Smith H, Luo L, Xu S (2005) Mapping quantitative trait loci using naturally occurring genetic variance among commercial inbred lines of maize (*Zea mays* L.). Genetics 169:2267–2275

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

Zheng S, Byrne PF, Bai G et al (2009) Association analysis reveals effects of wheat glutenin alleles and rye translocations on dough-mixing properties. J Cereal Sci 50:283–290

Zhou J, You A, Ma Z, Zhu L, He G (2012) Association analysis of important agronomic traits in japonica rice germplasm. Afr J Biotechnol 11:2957–2970

Zhu C, Gore M, Buckler ES, Yu J (2008) Status and prospects of association mapping in plants. Plant Genome 1:5–20

Chapter 12
Molecular Genetic Basis of the Domestication Syndrome in Cereals

Tao Sang and Jiayang Li

12.1 Introduction

Plant and animal domestication that was initiated approximately 10,000 years ago led to the dramatic evolution of human society and rapid speciation of plants and animals co-evolving with humans (Diamond 2002). The emergence of these new species and remarkable new traits fueled a series of scientific discoveries. The observation of rapid and drastic phenotypic changes under artificial selection stimulated at least partly Darwin's thinking of the origin of species under natural selection (Darwin 1859). Through experimental crosses and subsequent analyses of crop species, particularly pea plants, Mendel discovered the basic rules of genetics. With the arrival of the genomics era, recent studies yielded considerable new insights into the molecular basis of domestication traits and population genetic mechanisms underlying the domestication processes.

Cereal crops, including wheat, rice, maize, barley, sorghum, oats, and millets, provide the primary source of human calories. Of approximately 1.4 billion hectares or ~10 % of the terrestrial ecosystems converted to cropland, about half or ~0.7 billion hectares are currently used for producing cereals. The top three cereal crops, maize, rice, and wheat, grow on ~0.55 billion hectares (http://faostat.fao.org).

Cereals belong to the grass family, Poaceae, which consists of ~10,000 species worldwide. The domestication of cereals occurred independently in different

T. Sang (✉)
State Key Laboratory of Systematic and Evolutionary Botany, Key Laboratory of Plant Resources and Beijing Botanical Garden, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China
e-mail: sang@ibcas.ac.cn

J. Li
State Key Laboratory of Plant Genomics and National Center for Plant Gene Research, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China
e-mail: jyli@genetics.ac.cn

P. K. Gupta and R. K. Varshney (eds.), *Cereal Genomics II*,
DOI: 10.1007/978-94-007-6401-9_12,

continents, e.g., wheat and barley in Middle East ~10,000 years ago, rice and foxtail millet in China ~8,000 years ago, maize in Central America ~7,000–9,000 years ago, and sorghum and pearl millet in Africa ~4,000 years ago (Salamini et al. 2002; Doebley et al. 2006). With regard to the seemingly coincidental initiation of cereal domestications, an increasingly popular hypothesis is that they shared at least one common driving force, i.e., climate change following the last glacial maximum (Sage 1995; Richerson et al. 2001; Cunniff et al. 2008). The consequential change of global vegetation could have provided opportunities for humans to explore expanding grassland and developed a more reliable food source for feeding growing populations that faced increasing pressure of food shortage.

Despite the fact that the cereal crops were independently domesticated from distantly related grass species, the phenotypic modifications associated with the domestications were strikingly similar. The suite of phenotypic changes that transformed wild grasses into food crops are known as the domestication syndrome (Harlan 1992; Hancock 2004). This includes the following major trends: reduced grain shattering, improved threshing ability, weakened seed dormancy, reduced tiller number or shoot branches, synchronized grain maturation, lightened hull and seed colors, increased tiller erectness, enlarged panicles, and increased grain size and weight. Through these changes driven by artificial selection, plants became adapted to new environments and relied on humans for survival and reproduction.

In this chapter, we review the identification and molecular and population genetic analyses of genes involved in the development of the cereal domestication syndrome. With the literature review, we revisit the generalized theories and hypotheses concerning crop domestication. We then comment on the current status and future prospects of discovering domestication-related genes. From what has been learned about the molecular and population genetics of cereal domestication, we discuss how this information may be utilized to facilitate the domestication of new crops, especially perennial grasses potentially serving as lignocellulosic energy crops.

12.2 Domestication Syndrome and Genetic Analyses

Mature grains of wild grasses detach easily when disturbed by wind or animals so that their seeds can be dispersed promptly. The easy shattering of wild plants, however, makes grain harvest difficult, especially when they mature over a relatively long period of time during which substantial grain loss occurs as a result of strong wind or storms. Thus the reduction of grain shattering was necessary for effective harvest. Threshing was developed in association with the reduction in shattering and allowed grains or seeds to be separated and recovered from straw. There are two primary ways of threshing. In rice, mature grains are detached from harvested panicles by mechanistic forces, which is more or less equivalent to the partial reduction of shattering. The hulls are subsequently removed by milling. In wheat and barley, hulls open and separate from seeds at grain maturation so that seeds can be shaken out with moderate forces and collected.

Seed dormancy is an evolutionary safety strategy of wild grasses adapted to unstable environments. It allows seeds to remain in seed banks through the winter and germinate under favorable climatic conditions in the following spring. It also distributes seed germination over a prolonged period of time to avoid unpredicted harmful conditions such as flood and frost. However, this unwanted delay or uneven germination makes crops grow and mature at different rates, which in turn causes problems in crop management and harvest.

The loss of hulls and their appendices such as the stony fruitcases of maize kernels and awns of rice made food preparation easier but weakened protection against seed predators and ability of seed dispersal. The changes of hull colors, usually from dark to straw-white, might have been helpful for non-shattering grains to avoid predation at maturity while hull color becomes undistinguishable from that of withering straws (Zhu et al. 2011). Together, these phenotypic modifications changed the landscape of adaptation, resulting in well-adapted crop plants in the agricultural field, which could no longer survive in nature without human protection or assistance for reproduction.

The modification of plant architecture is another critical aspect of cereal domestication. This followed a common trend of reduction in tiller number or branches so that each tiller became stronger and capable of supporting a higher grain yield. Tillers also grew erect to allow a larger number of tillers to be compacted in a unit field. Panicles became more highly branched and capable of bearing a larger number of grains. These changes were clearly driven by selection for higher yield.

Domestication is evolution under artificial selection. The domestication syndrome highlights phenotypic changes that shifted the adaptive optima of wild grasses in their natural habitats to that of the new crops species in the agricultural field. To gain the mechanistic understanding of the process, we need to identify the genetic basis of these changes, which includes the number and chromosomal locations of loci/genes responsible for a domestication transition, the phenotypic effect of these loci/genes, and ultimately the casual mutations.

There are two main approaches for studying the genetic basis of domestication traits. A straightforward approach is to cross cultivars with their wild progenitors or relatives and subsequently conduct a quantitative trait locus (QTL) analysis. This so called “top-down” approach has been wildly adopted in crops because plants are relatively easy to hybridize. Numerous QTLs potentially underlying domestication transitions in almost all cereal crops have been reported. Another seemingly promising approach is to perform a genome-wide screening for signatures of artificial selection. Through a comparison of the distribution of nucleotide polymorphism between cultivars and their wild progenitors, loci/genes that presumably experienced selective sweeps are considered to be candidates involved in domestication. This is called a “bottom-up” approach (Ross-Ibarra et al. 2007).

In the first edition of *Cereal Genomics* published in 2004, domestication traits and corresponding QTLs were discussed (Pozzi et al. 2004). While the detected domestication-related QTLs continued to accumulate at a fast pace, a large portion of them have been deposited in user-friendly databases such as Gramene

(Youens-Clark et al. 2011). Over the past few years, the explosive development of grass genome resources and tools has considerably accelerated the molecular cloning of domestication-related QTLs, which have yielded markedly new insights into the genetic and evolutionary mechanisms of crop domestication. In addition, the advent of the next-generation sequencing technologies began to show a great potential of identifying candidate genes using the bottom-up approach (e.g., Xia et al. 2009). Thus here we choose not to update the growing list of identified QTLs, rather focus on the cloned QTLs/genes for a better understanding of cereal domestication.

12.3 Genes Underlying Domestication Syndrome

To date, more than a dozen of genes directly relevant to the development of the cereal domestication syndrome have been cloned (Table 12.1). Although many more genes identified in cereals were functionally related to the domestication traits, there was little evidence that they were directly responsible for the changes occurred during domestication. Domestication genes considered in this chapter share several features. First, the majority of them were cloned through the top-down approach, i.e., QTL analyses of crosses between cultivars and their wild relatives or between cultivars of independent origins, followed by fine mapping. Second, there was population genetic evidence indicating that the genes were under artificial selection. Finally, all but *tb1* were identified after 2005, which was apparently a result of recent advances in cereal genomics.

12.3.1 Shattering and Threshing

During cereal domestication, reduction in shattering was coupled with the maintenance or gain of threshing ability. The two traits together ensured grains to be effectively harvested and subsequently recovered. In rice, the balance of shattering and threshing was struck by weakening the function of the *sh4* gene that regulates the formation and function of the abscission zone from which a mature grain detaches from the pedicle (Li et al. 2006a). *Sh4* is the major shattering QTL in rice, explaining ~69 % of phenotypic variance between a traditional *indica* cultivar and the annual wild progenitor, *O. nivara* (Li et al. 2006b). The gene encodes a putative transcription factor. The causal mutation was a single nucleotide substitution leading to an amino acid substitution from lysine to asparagine in the predicted DNA binding domain. The substitution of the neutral for the positively charged amino acid, which presumably weakened but did not knock out the gene function, caused the incomplete development and partial function of the abscission zone (Li et al. 2006a). This disabled the natural detachment of grains necessary for seed dispersal in the wild species, but still allowed manual separation of grains from pedicels in cultivars during harvest.

Table 12.1 Genes controlling domestication traits of cereals

Domestication trait	Gene	Crop	Gene encoding	Causal mutation[a]	Origin	Fixation[b]	Year of publication
Shattering, threshing	*sh4*	Rice	Transcription factor	Trans-modification	Single	Fixed	2006
	qSH1	Rice	Transcription factor	Cis- modification	Single	Not fixed	2006
	Q	Wheat	Transcription factor	Trans-modification	Single	Fixed	2006
	nud	Barley	Transcription factor	Loss-of-function	Single	Fixed	2008
Plant architecture, inflorescence structure	*tb1*	Maize	Transcription factor	Cis-modification	Single	Fixed	1997
	prog1	Rice	Transcription factor	Tran- and Cis- modification	Single	Fixed	2008
	vrs1	Barley	Transcription factor	Loss-of-function	Multiple	Not fixed	2007
	Q	Wheat	Transcription factor	Trans-modification	Single	Fixed	2006
Grain/seed cover, size, and coloration	*tga1*	Maize	Transcription factor	Tran-modification	Single	Fixed	2005
	Bh4	Rice	Amino acid transporter	Loss-of-function	Multiple	Fixed	2011
	rc	Rice	Transcription factor	Loss-of-function	Multiple	Not fixed	2006
	rd	Rice	Enzyme	Loss-of-function	Multiple	Not fixed	2006
	phr1	Rice	Enzyme	Loss-of-function	Multiple	Not fixed	2008
	qSW5	Rice	Unknown	Loss-of-function	Single	Not fixed	2008
	GS3	Rice	Unknown	Loss-of-function	Single	Not fixed	2006

[a]The types of mutations considered here include loss-of-function and functional modification. For functional modification, a mutation occurring in the coding and regulatory region of a transcription factor is called trans- and cis-modification, respectively

[b]Domestication allele(s) that were found in all cultivars surveyed so for are considered to be fixed

Another shattering QTL, *qSH1*, accounting for ~69 % of phenotypic variance between *indica* and temperate *japonica* cultivars, was also cloned (Konishi et al. 2006). The causal mutation was a nucleotide substitution in the regulatory element located ~12 kb upstream of the coding region of a homeobox gene, which altered the level and pattern of the gene expression and disrupted the development of the abscission zone. The mutation was found in a portion of temperate *japonica* rice.

In barley, the derivation of non-shattering phenotype, also known as non-brittle rachis, was controlled primarily by two tightly linked loci, *btr1* and *btr2*. The homozygous recessive genotype at one of the loci, *btr1btr1/Btr2Btr2* or *Btr1Btr1/btr2btr2*, confers the non-brittle phenotype. Cultivars from the western parts of the world have predominantly the *btr1btr1/Btr2Btr2* genotype, while most of eastern cultivars have the *Btr1Btr1/btr2btr2* genotype. Although neither locus has been identified at the genic level, phylogenetic analysis of DNA sequences of the flanking regions showed that the eastern and western cultivars formed their own groups, indicating independent origins of non-brittle rachis from the eastern and western regions (Azhanguvel and Komatsusa 2007). The double homozygous recessive genotype *btr1btr1/btr2btr2*, however, has not been found in any barley cultivars and the linkage has never been broken up in experimental crosses (Komatsuda et al. 2004). It was hypothesized that *btr1* and *btr2* might be different mutations of the same gene (Sang 2009).

In tetraploid wheat with AABB genomic constitution, non-brittle rachis is controlled largely by two loci, Br_2 and Br_3, located in the homoeologous regions of group 3 chromosomes (Watanabe et al. 2002). They are potentially orthologous loci between the AA and BB genomes of the diploid parents. In hexaploid bread wheat, there is an additional brittle rachis locus, Br_1, also mapped to the orthologous location of group 3 chromosome, 3D, of the DD genome (Nalam et al. 2006; Watanabe et al. 2006). Furthermore, comparative mapping showed that this chromosomal region of wheat might be orthologous to that of barley containing *btr1* and *btr2* (Nalam et al. 2006; Pourkheirandish and Komatsuda 2007). The region is not orthologous to either of those harboring the rice shattering genes, *sh4* or *qSH1* (Li and Gill 2006; Sang 2009).

Unlike rice where grains are recovered from straw through threshing, seeds of free-threshing barley and wheat are directly removed from hulls that remain on straws. This required additional mutations that allowed easy release of seeds from hulls. The allele, *nud*, conferring free-threshing was cloned in barley (Taketa et al. 2008). In the wild progenitors of barley, the gene encodes an ethylene response factor that regulates lipid biosynthesis in the seed coat, which produces adhesive lipid between seed coats and hulls. A 17 kb deletion in the chromosome region containing *Nud* is responsible for the disruption of the lipid layer and consequently easy releasing of seeds from hulls in cultivated barley.

In wheat, free-threshing was achieved through the appearance of softened and easily separable hulls. Hulls of the free-threshing cultivars could open easily to release seeds under moderate forces such as beating or grinding. Genetic analysis between durum wheat and the wild progenitor of emmer wheat identified four QTLs for free-threshing (Simonetti et al. 1999). Of these, two with large

effect, each accounting for ~25 % of phenotypic variation, were *Tg* on the short arm of chromosome 2B and *Q* on the long arm of chromosome 5A. The free-threshing alleles, *tg* and *Q*, are partially recessive and partially dominant, respectively. The free-threshing tetraploid wheat has a genotype of $tgtg^{2B}QQ^{5A}$. In the hexaploid bread wheat, there is an additional recessive mutation at the *Tg* locus of the DD genome required for free-threshing, resulting in the genotype of $tgtg^{2B}tgtg^{2D}QQ^{5A}$ (Jantasuriyarat et al. 2004; Nalam et al. 2007).

Molecular cloning of *Q* showed that it is a gene belonging to the AP2 family of transcription factors (Simons et al. 2006). The *Q* allele had a higher level of transcription than the wild type allele, *q*, in spikes, leaves, and roots. The coding regions of the two alleles differed by an amino acid substitution, which was responsible for an increased abundance of homodimer of Q protein when tested in yeast. This mutation in the coding region, together with regulatory mutations potentially including a substitution at the microRNA binding site (Chuck et al. 2007), led to the gain-of-function mutation of *Q* that conferred the free-threshing phenotype. Interestingly, *Q* also contributes to the toughness of rachis or reduced shattering.

12.3.2 Plant Architecture and Inflorescence Structure

Maize has undergone the most drastic morphological modifications among all cereals. It involved the development of a single stalk from highly branched shoots of the wild progenitor, teosinte. A gene, *tb1*, controlling the difference was cloned using a maize mutant resembling the shoot branching pattern of teosinte (Doebley et al. 1997). It was confirmed that *tb1* was allelic to the major QTL underlying the architectural transition from teosinte to maize. The gene is a member of the TCP family of transcriptional regulators involved in the transcriptional regulation of cell cycle genes. In maize, *tb1* confers apical dominance by repressing the outgrowth of axillary meristems and branch elongation through its repressive effect on the cell cycle (Doebley et al. 2006). The causal mutations are located in the regulatory regions of the gene that alter the pattern and level of gene expression (Wang et al. 1999). The maize allele of *tb1* was highly expressed in the axillary buds whereas the teosinte allele showed no sign of expression (Hubbard et al. 2002).

In rice, a similar but less dramatic change occurred in plant architecture. In comparison to the wild progenitors, *O. nivara* and *O. rufipogon*, cultivated rice has fewer and more erect tillers. This architectural change has allowed cultivars to more effectively capture solar radiation and to be planted more densely in the field, both of which contribute to higher yield. *Prostrate growth 1* (*prog1*), responsible for this transition, was cloned using near-isogenic lines of rice cultivars that contained a small region of the short arm of chromosome 7 from *O. rufipogon* (Jin et al. 2008; Tan et al. 2008). The gene encodes a zinc-finger nuclear transcription factor and is predominantly expressed in the axillary meristems, from which tiller buds form. The causal mutation included primarily an amino acid substitution that weakens or disrupts the function of the gene and possibly those in its regulatory region as well.

Transgenic experiments showed that cultivated rice containing the wild-type allele from *O. rufipogon* had not only a larger number of more prostrate tillers but also shorter tillers with panicle bearing fewer primary and secondary branches and thus fewer grains. The pleiotropic effect of the gene matched almost perfectly the effect of a set of QTLs identified from a cross between an *indica* cultivar and *O. nivara* (Li et al. 2006b) and a cross between a *japonica* cultivar and *O. rufipogon* (Onishi et al. 2007). These QTLs, overlapped with *prog1* on the short arm of chromosome 7, had the largest effect on almost all morphological traits, including plant height, tiller number, tiller angle, and the number of primary and secondary branches of a panicle (Li et al. 2006b). If *prog1* is indeed allelic to the QTL for all of these traits, it represents one of the most important genes involved in the improvement of plant architecture and yield during rice domestication. It is amazing that artificial selection could have targeted a gene with such a wide range of pleiotropic effects rather than multiple genes each affecting various aspects of plant architecture, such as those individually controlling tiller number and angle or panicle branches (e.g., Li et al. 2003; Ashikari et al. 2005; Li et al. 2007; but see Jiao et al. 2010; Miura et al. 2010).

Another remarkable example of pleiotropic effect is the *Q* gene in wheat. While it is largely responsible for the development of free-threshing, it also contributes to non-brittle rachis, shorter culms, and shorter and denser spikes. The pleiotropic effect involves not only shattering and threshing necessary for effective harvest, but also plant architecture and inflorescence structure important for yield. Thus *Q* has been considered to be a super domestication gene (Faris et al. 2006).

During barley domestication, the appearance of six-rowed ears was a key innovation that substantially increased yield. On the ears of wild progenitors, each spike serving as a seed dispersal unit consists of three spikelets, of which the two lateral ones are reduced with only awns left to assist the dispersal of fully developed central spikelet. This trait remains the same in the domesticated two-rowed barley, while the awns are lost in the lateral spikeltes. In more advanced cultivars, the two lateral spikelets become fully developed so that the number of rows of grains is tripled.

Another barley gene, *Vrs1*, that controls the development of the lateral spikelets was cloned (Komatsuda et al. 2007). It encodes an HD-ZIP containing transcription factor expressed specifically in the lateral-spikelet primordia and suppresses the development of the lateral rows. The loss-of-function mutation of *Vrs1* allows further development of the lateral rows and gives rise to six-rowed barley. Three mutations in the coding region independently disrupt the function of the gene (Komatsuda et al. 2007).

12.3.3 Grain/Seed Cover, Size, and Coloration

The loss of grain/seed cover and coloration is another component of the domestication syndrome. The most remarkable is the loss of fruitcases during maize domestication. In teosinte, kernels are enclosed by stony fruitcases derived from modified

cupules and glumes. Teosinte ears disarticulate at maturity and the fruitcases become the units of seed dispersal. In maize, the fruitcases do not form so that naked kernels are readily edible. Cupules and glumes become a part of maize cobs on which kernels remain undetached at maturity. *Teosinte glume architecture1* (*tga1*), a major QTL controlling fruitcase formation, has been cloned (Wang et al. 2005). It is a member of the squamosa-promoter binding protein family of transcription regulators. In maize and teosinte, the gene is expressed in the inflorescence meristem of a developing ear, the spikelet primordia, and the adaxial junction of the spikelet and the inflorescence axis, the region where cupules and glumes develop. The functional difference between the maize and teosinte alleles of *tga1* appears to be due to a single amino acid substitution.

Changes in hull colors were widespread during cereal domestication. A common trend was the change from dark colors to the color that mimics withering straws, so-called straw-white. Two major QTLs were found to be responsible primarily for the change of hull color in rice (Gu et al. 2005). The one with larger effect, *bh4*, was recently cloned using a near-isogenic line of an *indica* cultivar containing a small region of chromosome 4 from *O. rufipogon* (Zhu et al. 2011). The wild-type allele in *O. rufipogon* encodes an amino acid transporter that is expressed specifically in the developing hulls. In different rice cultivars examined, it was found that two deletions and a nucleotide substitution in the coding region of the gene were independently responsible for the truncation of the BH4 protein and consequently for the loss of black hull color.

The color of seed coats or pericarps, although invisible during harvest, was also a target of artificial selection. The wild progenitors of rice have dark (brown to red) pericarps, whereas the pericarps of cultivated rice are predominantly white. Genetic analyses have shown that the color of rice pericarps is controlled primarily by two loci, *Rc* and *Rd*. Mutations at both loci have been identified. *Rc* encodes a bHLH protein that presumably regulates anthocyanin biosynthesis in the seed coat (Sweeney et al. 2006). Two mutations in exon 6 of the gene could independently be responsible for the loss of pigmentation in the pericarps. A 14-bp deletion was found in nearly 98 % of rice cultivars with white pericarps and a nucleotide substitution resulting in a premature stop codon was found in the remaining white rice (Sweeney et al. 2007).

Rd encodes dihydroflavonol-4-reductase, an enzyme involved in anthocyanin biosynthesis. The presence of premature stop codons in the first and second exons disrupts the function of the enzyme. When the lose-of-function alleles are denoted as *rc* and *rd*, the *Rc/Rd* and *Rc/rd* genotypes produce red and brown pericarps, respectively. The *rc/Rd* and *rc/rd* genotypes produce white pericarps (Furukawa et al. 2007).

Another color-related trait that experienced artificial selection is the darkening of hulls and pericarps of *indica* rice cultivars after prolonged storage. This, however, does not occur in *japonica* cultivars. The difference between the two types of cultivars is controlled by a single gene, *Phr1*, which encodes a polyphenol oxidase (Yu et al. 2008). The survey of *indica* cultivars identified three independent mutations, including two deletions and one insertion in the coding region that disrupted the function of the gene.

Larger and heavier grains are obviously favored by farmers. Numerous QTLs controlling grain size and weight have been identified. One of them, *qSW5*, with

the largest effect on the difference in grain width between a pair of *indica* and *japonica* cultivars appeared to have been targeted by artificial selection (Shomura et al. 2008). Molecular characterization of *qSW5* has indicated that in *indica* cultivars with narrower grains, the gene determines the number of cells in the outer glumes of rice flowers. In *japonica* cultivars with wider grains, a large deletion knocks out the gene function, which allows for the development of wider grains with additional rows of cells in the outer glumes.

The molecular cloning of a major QTL, *GS3*, controlling grain length provided additional insights into the changes of grain size and shape during rice domestication (Fan et al. 2006; Takano-Kai et al. 2009; Mao et al. 2010). The wild-type allele of *GS3* in the wild species and in the cultivars with grains of medium length is a negative regulator of grain size. Cultivars with long grains carry an allele with a loss-of-function mutation. It was targeted by artificial selection that favored long grains during rice cultivation.

It is interesting to note that none of the derived alleles responsible for an increase in rice grain width or length has been driven to fixation by artificial selection. This reflects the existence of highly diverse grain shapes and sizes among rice cultivars. It seems that while selection for larger grains was generally favored during rice cultivation, it was not one of the key driving forces of rice domestication. There are two reasons for this. One is that grain size and grain number on a panicle are often negatively correlated because the source for grain filling is limited (Wang et al. 2011). A trade-off between them may be beneficial for higher yield. The other reason is that in addition to grain size, variable grain shapes were probably selected by rice consumers and growers (Takano-Kai et al. 2009).

12.4 Gene Evolution and Domestication Processes

Because crops were derived from wild species under artificial selection, the cloning and evolutionary analyses of domestication genes should shed light on the history and processes of domestication. This type of information has been shown to be valuable for addressing questions such as how complex the genetic basis of a domestication trait could be? Whether a crop or a domestication trait originated once or multiple times? How long a domestication process could have lasted? In this section, we review the phylogenetic and population genetic analyses of the domestication-related genes identified in rice, maize, barley, and wheat, from which we attempt to gain a better understanding of domestication processes.

12.4.1 Rice

Rice, with a small genome and high-quality genome sequences (IRGSP 2005), has become a model system for studying plant function. It is thus not surprising that a much larger number of domestication genes have been cloned in rice than in other

cereals. Prior to the cloning of domestication genes, analyses of multiple neutral molecular markers suggested that rice was domesticated at least twice, with *indica* and *japonica* cultivars originated independently from wild species in different geographic locations (Cheng et al. 2003; Ma and Bennetzen 2004; Vitte et al. 2004; Zhu and Ge 2005; Londo et al. 2006; but see Molina et al. 2011a, b). The initial examination of the distribution of *sh4* alleles was apparently contradictory to the view of independent domestication. The finding that all cultivars shared the mutation responsible for the non-shattering phenotype indicated that there was a single origin of the domestication allele and possibly a single origin of cultivated rice (Lin et al. 2007).

Two models were proposed for the reconciliation (Sang and Ge 2007a). The snowballing model considers a single origin of a rice cultivar containing a set of essential domestication alleles including *sh4*. This original cultivar then spread and hybridized with wild populations, which gave rise to new cultivars with divergent genomic background. This process also led to an increase in genetic diversity of rice cultivars including their diversification into two groups of cultivars, *indica* and *japonica*. In the combination model, rice cultivars are believed to have been domesticated independently and cultivars of different origins possessed distinct sets of domestication genes. After these cultivars spread and hybridized, the best set of domestication genes were selected and combined into the modern cultivars while the genomic diversity was largely maintained.

A recent study based on the gene markers from three rice chromosomes claimed a strong support for the single origin of domesticated rice (Molina et al. 2011a). This result was challenged because the study seemed to have underestimated the likelihood of independent domestications (Ge and Sang 2011). While the models of rice domestication have yet to be tested, it is almost certain that in either case the gene flow between cultivars and wild progenitors was an important part of the process of rice domestication. The spreading of valuable domestication alleles through introgression was also documented at another domestication related locus, *Rc*, which primarily controls the pericarp color. It was shown that the common lost-of-function allele of the gene that accounted for 98 % of white-pericarp in rice originated in *japonica* rice and subsequently spread into *indica* cultivars through introgression (Sweeney et al. 2007).

Insight into the domestication of rice was also gained through population genetic analysis of *sh4*, which estimated the rate of fixation of the non-shattering allele in cultivated rice. A severe reduction of DNA sequence polymorphism in cultivated rice was observed at the *sh4* locus, suggesting that the allele experienced a selective sweep under strong artificial selection. Because the allele allowed for well balanced shattering and threshing, the selection could be strong enough to drive its fixation in a short period of ~100 years (Zhang et al. 2009).

However, it was recently found that in the archeological sites in China, the frequency of non-shattering phenotype increased relatively slowly, suggesting that rice domestication, as evaluated on the basis of the development of a non-shattering trait, might have lasted for two to three millennia (Fuller et al. 2009).

Nevertheless, this apparent discrepancy in the rate of rice domestication between genetic and archeological data can be reconciled under certain circumstances. One possibility is that *sh4* did not arise or did not have a chance to spread to the archeological sites where cultivars with inferior non-shattering alleles were present (Zhang et al. 2009). If this turns out to be the case, one can conclude that population genetic analyses of domestication genes coupled with archeological evidences would give a more complete picture of rice domestication. The fixation of a critical domestication allele could be rapid locally, but prolonged for the crop as a whole.

The similar situation was later found in another domestication gene, *prog1*, which was most likely allelic to the QTL controlling a suite of morphological changes leading to better architecture and higher yield of cultivated rice. Although there has not been a thorough phylogenetic and population genetic analysis of *prog1* in comparison to *sh4*, based on genetic analyses reported from two independent studies it seems clear that the domestication allele *prog1* with modified functions originated once and had been fixed in all rice cultivars examined (Jin et al. 2008; Tan et al. 2008).

The shared pattern between *sh4* and *prog1*, each underlying a key component of domestication syndrome in rice, reinforces the notion that introgression played an essential role in the rapid fixation of domestication alleles giving rise to superior phenotypes. It was strong artificial selection that led to the spreading and fixation of the most desirable set of domestication alleles. Meanwhile, natural selection on hybrids with distinct genomic backgrounds would help maintain local adaptation and genetic diversity of cultivars (Sang and Ge 2007b).

The two genes controlling hull and pericarp colors, *Bh4* and *Rc*, share another pattern. Cultivars have alleles with loss-of-function mutations, with one allele being predominant and others occurring at low frequencies and having independent origins (Sweeney et al. 2007; Zhu et al. 2011). These low-frequency alleles have been maintained in cultivars probably because they are functionally indistinguishable from the common alleles and have not been wiped out by selective sweep associated with the common alleles. Given such a low nucleotide polymorphism of the common white-hull allele, it seems to have originated recently, possibly even more recent than *sh4*. If the change of hull color from black to straw-white was indeed to mimic straw color and avoid bird predation, it makes sense that this change occurred after the non-shattering phenotype was widely established (Zhu et al. 2011).

Like *bh4* and *rc*, the derived alleles of *GS3* and *SW5* also had loss-of-function mutations. However, the frequencies of the loss-of-function alleles are much lower, consistent with the hypothesis that longer and/or wider grains were not universally favored in cultivated rice. The change in grain shape and increase in grain size must have met various limitations such as source availability and consumer preference. Although both alleles experienced artificial selection and increased in frequency during rice domestication, they were obviously different from those domestication alleles that carried clearly selective advantages in one direction.

12.4.2 Maize

Phylogenetic analyses of genome-wide neutral markers were in agreement with the archeological evidence, pointing to a single domestication of maize near the central Balsas river valley of southern Mexico approximately 7,000–9,000 years ago (Matsuoka et al. 2002; Doebley 2004). The identification of two major domestication genes, *tb1* and *tg1*, each regulating a key phenotypic transition lend additional support to this conclusion.

The story of *tg1* is probably the most straightforward of all domestication genes. A single nonsynonymous substitution was primarily responsible for the loss of stony fruitcases that protected teosinte kernels. The allele was derived ~10,000 years ago and was fixed in maize under strong artificial selection (Wang et al. 2005). The estimated early origin of this allele also suggests that an absence of fruitcase in kernels for easier human consumption was an important early step of maize domestication. The single domestication event allowed *tg1* to be fixed more easily than *sh4* in rice.

The gene *tb1* was the first cloned QTL for domestication of crops. The search for the causal mutation, however, took much longer than any other domestication genes later identified. After almost a decade of persistent effort, the mutation(s) were narrowed down to a region ~58–69 kb 5′ to the coding region of *tb1* (Clark et al. 2004, 2006). It was also confirmed that the mutation(s) were in the regulatory region, which changed *tb1* expression and plant architecture. Despite the early detection of strong artificial selection in the 5′ upstream region of *tb1*, the strikingly distant location of the causal mutation(s) from the gene itself complicated the fine mapping process.

Although the domestication history of maize seems to be the best characterized of all cereals, there are still many questions left to be addressed with continuing effort to identify and analyze domestication genes. It is still unclear whether a single mutation or several mutations of independent origins gave rise to *tb1* in maize. With genome sequences available, cloning additional important domestication genes of maize will be of great interest for comparison with rice domestication. The extent to which introgression between cultivars and teosinte might have influenced maize domestication has just begun to be revealed (van Heerwaarden et al. 2011).

12.4.3 Barley and Wheat

Barley and wheat, belonging to the same subfamily, Pooideae, were domesticated from the Fertile Crescent ~10,000 years ago. Map-based QTL cloning is relatively difficult in barley and wheat due to their large genome sizes. Despite this, essential domestication QTLs such as *btr* and *br* controlling brittle rachis were mapped on chromosomes, although their characterization at the molecular level is still awaited. The fact that these QTLs are mapped to the orthologous chromosomal

regions between barley and wheat suggests that the orthologous genes might have been targeted by artificial selection for non-brittle rachis in these crops (Sakuma et al. 2011). This will be an interesting hypothesis to test once the QTLs are cloned (Paterson et al. 1995).

For the free-threshing trait, the major QTL, *Nud*, was cloned from barley, and a secondary QTL, *Q*, was cloned from wheat. Each of these two domestication genes had a single origin and spread into cultivars of independent origins (Simons et al. 2006; Taketa et al. 2008). For barley, a growing body of phylogenetic evidences suggested that it contained cultivars with distinct genomic backgrounds derived from wild progenitors in two places, the Fertile Crescent and another area 1,500–3,000 km farther east (Morrell and Clegg 2007). This situation resembles the case of rice domestication where two types of cultivars (indica and japonica) with distinct genomic backgrounds share the domestication genes of single origins. It appears that the models of rice domestication could at least partly apply to other cereals.

Free-threshing wheat includes both tetraploid and hexaploid cultivars. With a gain-of-function mutation, one *Q* locus could confer the phenotype in the polyploid wheat. The free-threshing condition may have developed in the AABB-genome wheat with the genotype $tgtg^{2B}QQ^{5A}$, which then served as the tetraploid parent of hexaploid bread wheat. In this scenario, an additional recessive mutation at *Tg* of the DD genome would be sufficient. Alternatively, *Q* was selected initially in the hexaploid wheat and then spread into tetraploid wheat through introgression (Faris et al. 2006; Simons et al. 2006). Either scenario supports the importance of gene flow during cereal domestication.

12.4.4 Generalization

Above we reviewed and discussed recent advances in the identification and analyses of genes controlling the domestication syndrome of cereals. The rapid progress made along this line over the past few years substantially enhanced our understanding of crop domestication. This provides us with the opportunity to evaluate certain general conclusions drawn on domestication mechanisms and processes.

Several decades ago, Beadle proposed the "one-gene, one-trait" hypothesis as an explanation for the finding that one chromosome region often accounted for a major domestication transition from teosinte to maize (Beadle 1939; Doebley 2001). The cloning of domestication genes in cereals supports this hypothesis to a large extent (Sang 2009). A single gene or even a single mutation in many cases controlled a critical domestication transition. These include *tb1* and *tg1* in maize, *sh4* and *prog1* in rice, and *nud* in barley, each primarily responsible for the development of a key component of the domestication syndrome. It was remarkable that the one-gene, one-trait hypothesis derived from maize holds for rice and barley despite the fact that each of these crops might have had multiple origins. Models developed for rice provided a mechanistic explanation for the one-gene, one-trait

hypothesis (Sang and Ge 2007a, b). They suggested that a highly favorable domestication allele could be quickly fixed even for crops of multiple origins by strong artificial selection combined with introgression.

Whether a domestication allele is fixed in a crop or not probably reflects the strength of artificial selection. It is noticeable that the majority of alleles conferring reduction in coloration and increase in grain length or width are not fixed in rice. These traits are not essential for cultivation; actually in some cases, wild-type alleles were selected for the maintenance of phenotypic diversity as exemplified by with red-pericarp and variable grain shapes among rice cultivars. Furthermore, these alleles, including *rc*, *rd*, *phr1*, *qSW5*, and *GS3*, were derived from loss-of-function mutations that often had multiple origins.

It was previously suggested, based on the review of domestication-related genes in a broader range of crops, that key domestication genes tended to be regulatory genes with modified functions in crops, whereas genes controlling varietal variation were broader in functional categories and more frequently underwent loss-of-function mutations during domestication (Doebley et al. 2006). The growing number of domestication-related genes cloned in cereals supports the hypothesis and further suggests that essential domestication alleles tend to have single origins while alleles controlling varietal variation often have multiple origins (Table 12.1).

Taken together, a critical domestication transition necessary for cultivation tends to be controlled by a single mutation of a single gene. The domestication genes are often transcription factors with modified functions resulting from mutations in either regulatory or coding regions (Table 12.1). The functional modification could cause a cascade of downstream effects that substantially altered a trait but had relatively little deleterious effects on plants (Doebley and Lukens 1998). The selection on these alleles was so strong that could have driven them to fixation in cultivars of independent origins. Genes for further improvement of domestication traits were not absolutely necessary for cultivation, and therefore were more diverse in functional categories and more often had independent loss-of-function mutations. The domestication alleles of these genes are often not fixed in all cultivars due to either relatively weak selection or diversifying selection for varietal variation.

12.5 Recent Advances

The cloning and characterization of genes underlying the domestication syndrome of cereals yielded intriguing insights into the molecular and population genetics of crop domestication. It is expected that the pace of cloning domestication QTLs will increase with rapid advances in grass genomics. With additional cereal genomes, especially those of sorghum and foxtail millet, being recently sequenced, the effort to identify domestication-related genes from these two crops will begin to bear fruits and consequently allow a comparison of the genetic basis of domestication in a wider range of cereal crops.

Another major advance in genomics over the past few years was the development of the next-generation sequencing technologies. This will bring substantial changes to the molecular genetic studies of domestication traits. QTL cloning may be considerably accelerated with the use of genotyping methods relying on next-generation sequencing. It has been shown that resequencing the genomes of recombinant populations at low genome coverage could markedly speed up genotyping and improve the resolution of QTL mapping (Huang et al. 2009). QTL maps with higher precision and resolution will facilitate the process of gene cloning (Wang et al. 2011).

Rapid genome resequencing also opened great opportunities for identifying domestication-related genes through a direct analysis of genome-wide nucleotide polymorphisms between crops and their wild progenitors. Severe reduction of nucleotide polymorphism in cultivars is an indication of selective sweep and these regions potentially contain genes that experienced artificial selection. This "bottom-up" approach has been applied to maize using PCR-based markers (Vigouroux et al. 2002; Wright et al. 2005), although the full potential of SNPs identified from whole-genome resequencing has not been tested in crops. The resequencing data seem to be most accessible for rice whose small and homozygous genome offers an excellent system for exploring the potential of this approach (He et al. 2011).

Despite the seemingly promising potential, we have yet to see a body of publications emerging from this idea. Apparently, several technical issues need to be addressed. First, it requires the sequences of a reference genome for a crop system. In cereals, this requirement is met better than any other groups of crops. High-quality genome sequences are available for rice, maize, and sorghum (Paterson et al. 2009; Schnable et al. 2010), and will soon be available for foxtail millet (Doust et al. 2009). Sequencing the large genomes of barley and wheat is still challenging, but should become increasingly realistic with the emergence of the third-generation sequencing technologies.

Second, high-density SNPs generated by whole-genome resequencing provide great opportunities for narrowing down candidate genes anywhere in the genome, but it presents considerable challenges for the detection of regions that indeed experienced selective sweep. With the whole genome under examination, the possibility of detecting false positives increases dramatically. Various stochastic factors combined with population demography are very likely to cause difficulties in setting appropriate statistical threshold for maximizing the detection of domestication-related genes and meanwhile minimizing false positives. New population genomic models and analytical methods are needed in order to take full advantage of resequencing data to single out regions under reasonably strong artificial selection.

Finally, a few sampling strategies need to be considered. Appropriate genome coverage of resequencing should be determined for specific cases. There is a possibility that a relatively low coverage is sufficient for generating a high-density haplotype map through data imputation as long as the sample size is large enough (Huang et al. 2010). The difficulty remains however for the imputation in crops and their wild relatives whose genomes are heterozygous. The same is true for

taxonomic sampling strategy. A suitable number of cultivars and wild relatives that effectively cover the entire diversity of crops and wild relatives needs to be determined. There should be ways to identify individuals derived from introgression between cultivars and their wild relatives. Including the hybrids in the analysis may violate the assumptions of population genetic models. Furthermore, unlike the top-down approach that starts with traits known to be important for domestication, candidate genes identified from genome screening need to be tested for their phenotypic effect and for their role in artificial selection.

12.6 Summary and Outlook

The independent domestication of cereal crops was triggered at least partly by the climate change following the last glacial maxima. Today the world is facing an anthropogenic climate change that will presumably have serious consequences. This is primarily due to rapid consumption of fossil fuels since the Industrial Revolution. The quick depletion of fossil energy and the consequential climate changes has forced the world to develop renewable sources of energy. Of various possible sources, bioenergy is the one that holds the greatest potential for replacing fossil oil. The large-scale production of bioenergy will have to rely heavily on energy crops.

Crops that are currently used for producing bioethanol and biodiesel are those previously domesticated for food, sugar or vegetable oil, such as maize, sugarcane, soybean, and rapeseed. They are known as first-generation energy crops. However, it has been increasingly realized that these annual crops cannot be sustainable solutions to energy shortage or climate change because of their low net energy output, low potential for greenhouse gas mitigation, and inability to utilize marginal land (Fargione et al. 2008; Robertson et al. 2008; Searchinger et al. 2008). Under these circumstances, the concept of second-generation energy crops emerged. These are dedicated energy crops that can be grown on marginal land with relatively little energy input involved in plantation, irrigation, fertilization, harvest, and transportation. Meanwhile, they should yield high biomass and have a strong ability of carbon sequestration. With these features, second-generation energy crops are capable of providing a major sustainable source of renewable energy with little or even negative greenhouse gas emission (Heaton et al. 2008; Karp and Shield 2008; Oliver et al. 2009).

Grasses are likely to play a leading role as potential second-generation energy crops. Therefore, what we learned from cereal crop domestication can be used for guiding us in bring about the next round of crop domestication for energy and environment security. Two perennial grasses, switchgrass and *Miscanthus*, native to North America and Asia respectively, have already emerged as the top candidates for second-generation energy crops in the northern temperate regions of the world (Somerville et al. 2010; Sang and Zhu 2011). They share the features of being C4 perennials with high water and nutrient use efficiencies and are adapted to a wide range of climates. Similar to cereal crops, the domestication of these energy crops also tends to target at a small portion of grass species native to different continents.

Unlike cereal domestication, where artificial selection focused on planting, harvest efficiencies, grain yield, and harvest indices, the domestication of lignocellulosic energy crops will focus on sustainable production under unfavorable soil and climate conditions (Sang 2011). With regard to the domestication traits, they may share characteristics such as strong biotic and abiotic resistance, maximized length of vegetative growth as permitted by local climates, the highest possible water, nutrient, and radiation use efficiencies, at least partial sterility that minimizes seed production, and modified lignocellulosic properties for effective biorefining.

In addition to the domestication syndrome, the duration of domestication of dedicated energy crops will have to be much shorter than that of cereals. Instead of centuries to millennia taken for domesticating a food crop, the domestication of energy crops needs to be completed within decades in order to address problems of energy shortage and anthropogenic climate change facing the world. Fortunately, the knowledge gained from the study of cereal domestication provides us with encouraging clues to rapid domestication. Because a key domestication transition could result from selection for a single mutation, it is possible that a desirable trait develops quickly in a new crop, if a large population with ample natural and induced mutations is subjected to strong artificial selection. We can then shape up the domestication syndrome by combining a suite of desirable traits through carefully designed experimental crosses. Genetic and genomic studies that identify valuable loci and genes will considerably facilitate and speed up the process of domestication through molecular breeding and biotechnology.

References

Ashikari M, Sakakibara H, Lin S, Yamamoto T, Takashi T, Nishimura A, Angeles E, Qian Q, Kitano H, Matuoka M (2005) Cytokinin oxicase regulates rice grain production. Science 309:741–745

Azhanguvel P, Komatsusa T (2007) A phylogenetic analysis based on nucleotide sequence of a marker linked to the brittle rachis locus indicates a diphyletic origin of barley. Ann Bot 100:1009–1015

Beadle GW (1939) Teosinte and the origin of maize. J Hered 30:245–247

Cheng C, Motohashi R, Tsuchimoto S, Fukuta Y, Ohtsubo H, Ohstubo E (2003) Polyphyletic origin of cultivated rice: based on the interspersion pattern of SINEs. Mol Biol Evol 20:67–75

Chuck G, Meeley R, Irish E, Sakai H, Hake S (2007) The maize *tasselseed4* microRNA controls sex determination and meristem cell fate by targeting *Tasselseed6/indeterminate spkelet1*. Nat Genet 39:1517–1521

Clark RM, Linton E, Messing J, Doebley JF (2004) Pattern of diversity in the genomic region near the maize domestication gene *tb1*. Proc Natl Aced Sci USA 101:700–707

Clark RM, Wagler TN, Quijada P, Doebley JF (2006) A distant upstream enhance at the maize domestication gene *tb*1 has pleiotropic effects on plant and inflorescent architecture. Nat Genet 38:594–597

Cunniff J, Osborne CP, Ripley BS, Charles M, Jones G (2008) Response of wild C4 crop progenitors to subambient CO_2 highlights a possible role in the origin of agriculture. Glob Change Biol 14:576–587

Darwin CR (1859) On the origin of species by means of natural selection. Jone Murray, London

Diamond J (2002) Evolution, consequences and future of plant and animal domestication. Nature 418:700–707

Doebley JF (2001) George Beadle's other hypothesis: one-gene, one-trait. Genetics 158:487–493
Doebley JF (2004) The genetics of maize evolution. Ann Rev Genet 38:37–59
Doebley JF, Lukens L (1998) Transcriptional regulators and the evolution of plant form. Plant Cell 10:1075–1082
Doebley JF, Stec A, Hubbard L (1997) The evolution of apical dominance in maize. Nature 386:485–488
Doebley JF, Gaut BS, Smith BD (2006) The molecular genetics of crop domestication. Cell 127:1309–1321
Doust AN, Kellogg EA, Devos KM, Bennetzen JL (2009) Foxtail millet: a sequence-driven grass model system. Plant Physiol 149:137–141
Fan C, Xing Y, Mao H, Lu T, Han B, Xu C, Li X, Zhang Q (2006) *GS3*, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. Theor Appl Genet 112:1164–1171
Fargione J, Hill J, Tilman D, Polasky S, Hawthorne P (2008) Land clearing and the biofuel carbon debt. Science 319:1235–1238
Faris JD, Simons KJ, Zhang Z, Gill BS (2006) The wheat super domestication gene *Q*. Wheat Information Service—Frontiers of Wheat Bioscience 100:129–148. (http://www.shigen.nig.ac.jp/wheat/wis/No100/100.html)
Fuller DQ, Qin L, Zheng Y, Zhao Z, Chen X, Hosoya LA, Sun GP (2009) The domestication process and domestication rate in rice: spikelet bases from the lower Yangtze. Science 323:1607–1610
Furukawa T, Maekawa M, Oki T, Suda I, Lida S, Shimada H, Takamure I, Kadowaki K (2007) The Rc and Rd genes are involved in proanthocyanidin synthesis in rice pericarp. Plant Journal 49:91–102
Ge S, Sang T (2011) Inappropriate model rejects independent domestications of *indica* and *japonica* rice. Proc Natl Acad Sci USA 108:E75
Gu XY, Kianian SF, Hareland GA, Hoffer BL, Foley ME (2005) Genetic analysis of adaptive syndromes interrelated with seed dormancy in weedy rice (*Oryza sativa*). Theor Appl Genet 110:1108–1118
Hancock JF (2004) Plant evolution and the origin of crop species, 2nd edn. CABI Publishing, Cambridge
Harlan JR (1992) Crops and man, 2nd edn. American Society of Agronomy and Crop Science Society of America, Madison
He Z, Zhai W, Wen H, Tan T, Wang Y, Lu X, Greenburg AJ, Hudson RR, Wu C-I, Shi S (2011) Two evolutionary histories in the genome of rice: the roles of domestication genes. PLoS Genet 7:e1002100
Heaton EA, Flavell RB, Mascia PN, Thomas SR, Dohleman FG, Long SP (2008) Herbaceous energy crop development: recent progress and future prospects. Curr Opin Biotech 19:202–209
Huang XH, Feng Q, Qian Q, Zhao Q, Wang L, Wang AH, Guan JP, Fan DL, Weng QJ, Huang T, Dong GJ, Sang T, Han B (2009) High-throughput genotyping by whole-genome resequencing. Genome Res 19:1068–1076
Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z, Li M, Fan D, Guo Y, Wang A, Wang L, Deng L, Li W, Lu Y, Weng Q, Liu K, Huang T, Zhou T, Jing Y, Li W, Lin Z, Buckler ES, Qian Q, Zhang Q, Li J, Han B (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. Nat Genet 42:961–967
Hubbard L, McSteen P, Doebley J, Hake S (2002) Expression patterns and mutant phenotype of *teosinte branched1* correlate with growth suppression in maize and teosinte. Genetics 162:1927–1935
International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. Nature 436:793–800
Jantasuriyarat C, Vales MI, Watson CJW, Riera-Lizarazu O (2004) Identification and mapping of genetic loci affecting the free-threshing habit and spike compactness. Theor Appl Genet 108:261–273
Jiao Y, Wang Y, Xue D, Wang J, Yan M, Liu G, Dong G, Zeng D, Lu Z, Zhu X, Qian Q, Li J (2010) Regulation of *OsSPL*14 by OsmiR156 defines ideal plant architecture in rice. Nat Genet 42:541–544

Jin J, Huang W, Gao J-P, Yang J, Shi M, Zhu M-Z, Luo D, Lin H-X (2008) Genetic control of rice plant architecture under domestication. Nat Genet 40:1365–1369

Karp A, Shield I (2008) Bioenergy from plants and the sustainable yield challenge. New Phytol 179:15–32

Komatsuda T, Maxim P, Senthil N, Mano Y (2004) High-density AFLP map of nonbrittle rachis 1 (*btr*1) and (*btr*2) genes in barley (*Hordeum vulgare* L.). Theor Appl Genet 109:986–995

Komatsuda T, Pourkheirandish M, He C, Azhaguvel P, Kanamori H, Perovic D, Stein N, Graner A, Wicher T, Tagiri A et al (2007) Six-rowed barley originated from a mutation in a homeodomain-leucine zipper I-class homeobox gene. Proc Natl Acad Sci USA 104:1424–1429

Konishi S, Izawa T, Lin SY, Ebana K, Fukuta Y, Sasaki T, Yano M (2006) An SNP caused loss of seed shattering during rice domestication. Science 312:1392–1396

Li W, Gill BS (2006) Multiple genetic pathways for seed shattering in the grasses. Funct Integr Genomics 6:300–309

Li X, Qian Q, Fu Z, Wang Y, Xiong G, Zeng D, Wang X, Liu X, Teng S, Hiroshi F, Yuan M, Luo D, Han B, Li J (2003) Control of tillering in rice. Nature 422:618–621

Li C, Zhou A, Sang T (2006a) Rice domestication by reducing shattering. Science 311:1936–1939

Li C, Zhou A, Sang T (2006b) Genetic analysis of rice domestication syndrome with the wild annual species, *Oryza nivara*. New Phytol 170:185–194

Li P, Wang Y, Qian Q, Fu Z, Wang M, Zeng D, Li B, Wang X, Li J (2007) *LAZY1*controls rice shoot gravitropism through regulating polar auxin transport. Cell Res 17:402–410

Lin Z, Griffith ME, Li X, Zhu Z, Tan L, Fu Y, Zhang W, Wang X, Xie D, Sun C (2007) Origin of seed shattering in rice (*Oryza sativa* L.). Planta 226:11–20

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

Ma J, Bennetzen JL (2004) Rapid recent growth and divergence of rice nuclear genomes. Proc Natl Acad Sci USA 101:12404–12410

Mao H, Sun S, Yao J, Wang C, Yu S, Xu C, Li X, Zhang Q (2010) Linking differential domain functions of the GS3 protein to natural variation of grain size in rice. Proc Natl Acad Sci USA 107:19579–19584

Matsuoka Y, Vigouroux Y, Goodman MM, Sanchez J, Buckler E, Doebley J (2002) A single domestication for maize shown by multilocus microsatellite genotyping. Proc Natl Acad Sci USA 99:6080–6084

Miura K, Ikeda M, Matsubara A, Song X, Ito M, Asano K, Matsuoka M, Kitano H, Ashikari M (2010) OsSPL14 promotes paniclebranching and higher grain productivity in rice. Nat Genet 42:545–549

Molina J, Sikora M, Garud N, Flowers JM, Rubinstein S, Rynolds A, Huang P, Jackson SA, Schaal BA, Bustanante CD, Boybo AR, Purugganan MD (2011a) Molecular evidence for a single evolutionary origin of domesticated rice. Proc Natl Acad Sci USA 108:8351–8356

Molina J, Sikora M, Garud N, Flowers JM, Rubinstein S, Rynolds A, Huang P, Jackson SA, Schaal BA, Bustanante CD, Boybo AR, Purugganan MD (2011b) Reply to Ge and Sang: a single origin of domesticated rice. Proc Natl Acad Sci USA, doi/10.1073/pnas.1112466108

Morrell PL, Clegg MT (2007) Genetic evidence for a second domestication of barley (*Hordeum vulgare*) east of the Fertile Crescent. Proc Natl Acad Sci USA 104:3289–3294

Nalam VJ, Vales MI, Watson CJW, Kianian SF, Riera-Lizarazu O (2006) Map-based analysis of genes affecting the brittle rachis character in tetraploid wheat (*Triticum turgidum* L.). Theor Appl Genet 112:373–381

Nalam VJ, Vales MI, Watson CJW, Johnson EB, Riera-Lizarazu O (2007) Map-based analysis of genetic loci on chromosome 2D that affect glume tenacity and threshability, components of the free-threshing habit in common wheat (*Triticum aestivum* L). Theor Appl Genet 116:135–145

Oliver RJ, Finch JW, Taylor G (2009) Second generation bioenergy crops and climate change: a review of the effects of elevated atmospheric CO_2 and drought on water use and the implications for yield. GCB Bioenergy 1:97–114

Onishi K, Horiuchi Y, Ishigoh-Oka N, Takagi K, Ichikawa N, Maruoka M, Sano Y (2007) A QTL cluster for plant architecture and its ecological significance in Asian wild rice. Breeding Sci 57:7–16

Paterson AH, Lin YR, Li Z, Schertz KF, Doebley JF, Pinson SRM, Liu SC, Stansel JW, Irvine JE (1995) Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. Science 269:1714–1718

Paterson AH et al (2009) The Sorghum bicolor genome and the diversification of grasses. Nature 457:551–556

Pourkheirandish M, Komatsuda T (2007) The importance of barley genetics and domestication in a global perspective. Ann Bot 100:999–1008

Pozzi C, Rossini L, Vecchietti A, Salamini F (2004) Gene and genome changes during domestication of cereals. In Gupta PK, Varshney RK (eds) Cereal genomics, pp 165–198

Richerson PJ, Boyd R, Bettinger RL (2001) Was agriculture impossible during the Pleiostocene but mandatory during the Holocene? A climate change hypothesis. Amer Antiq 66:387–411

Robertson GP, Dale VH, Doering OC, Hamburg SP, Melillo JM, Wander MM, Parton WJ, Adler PR, Barney JN, Cruse RM, Duke CS, Fearnside PM, Follett RF, Gibbs HK, Goldember J, Dladenoff DJ, Ojima D, Palmer M, Sharpley A, Wallace L, Weathers KC, Wiens JA, Wilhelm WW (2008) Sustainable biofuels redux. Science 322:49–50

Ross-Ibarra J, Morrell PL, Gaut BS (2007) Plant domestication, a unique opportunity to identify the genetic basis of adaptation. Proc Natl Acad Sci USA 104:8641–8648

Sage R (1995) Was low atmospheric CO_2 during the Pleistocene a limiting factor for the origin of agriculture? Glob Change Biol 1:93–106

Sakuma S, Salomon B, Komatsuda T (2011) The domestication syndrome genes responsible for the major changes in plant form in the Triticeae crops. Plant Cell Physiol 52:738–749

Salamini F, Ozkan H, Brandolini A, Schafer-Pregl R, Marin W (2002) Genetics and geography of wild cereal domestication in the Near East. Nat Rev Genet 3:420–441

Sang T (2009) Genes and mutations underlying domestication transitions in grasses. Plant Physiol 149:63–70

Sang T (2011) Toward the domestication of lignocellulosic energy crops: learning from food crop domestication. J Integr Plant Biol 53:96–104

Sang T, Ge S (2007a) The puzzle of rice domestication. J Integr Plant Biol 49:760–768

Sang T, Ge S (2007b) Genetics and phylogenetics of rice domestication. Cur Opin Genet Dev 17:533–538

Sang T, Zhu W-X (2011) China's bioenergy potential. GCB Bioenergy 3:79–179

Schnable et al (2010) The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112–1115

Searchinger T, Heimlich R, Houghton RA, Dong F, Elobeid A, Fabiosa J, Tokgoz S, Hayes D, Yu TH (2008) Use of U.S. croplands for biofuels increases greenhouse gases through emissions from land use change. Science 319:1238–1244

Shomura A, Izawa T, Ebana K, Ebitani T, Kanegae H, Konishi S, Yano M (2008) Deletion in a gene associated with grain size increased yields during rice domestication. Nat Genet 40:1023–1028

Simonetti MC, Bellomo MP, Laghetti G, Perrino P, Simeone R, Blanco A (1999) Quantitative trait loci influencing free-threshing habit in tetraploid wheats. Genet Res Crop Evol 46:267–271

Simons KJ, Fellers JP, Trick HN, Zhang Z, Tai YS, Gill BS, Faris JD (2006) Molecular characterization of the major wheat domestication gene *Q*. Genetics 172:547–555

Somerville C, Yongs H, Taylor C, Davis SC, Long SP (2010) Feedstocks for lignocellulosic biofuels. Science 329:790–792

Sweeney MT, Thomson MJ, Pfeil BE, McCouch SR (2006) Caught red-handed: *Rc* encodes a basic helix-loop-helix protein conditioning red pericarp in rice. Plant Cell 18:283–294

Sweeney MT, Thomson MJ, Cho YG, Park YJ, Williamson SH, Bustamante CD, McCouch SR (2007) Global dissemination of a single mutation conferring white pericarp in rice. PLoS Genet 3:e133

Takano-Kai N, Jiang H, Kubo T, Sweeney M, Matsumoto T, Kanamori H, Padhukasahasram B, Bustamante C, Yoshimura A, Doi K, McCouch S (2009) Evolutionary history of GS3, a gene conferring grain length in rice. Genetics 182:1323–1334

Taketa S, Amano S, Tsujino Y, Sato T, Saisho D, Kakeda K, Nomura M, Suzuki T, Matsumoto T, Sato K et al (2008) Barley grain with adhering hulls is controlled by an ERF family transcription factor gene regulating a lipid biosynthesis pathway. Proc Natl Acad Sci USA 105:4062–4067

Tan L, Li X, Liu F, Sun X, Li C, Zhu Z, Fu Y, Cai H, Wang X, Xie D, Sun C (2008) Control of a key transition from prostrate to erect growth in rice domestication. Nat Genet 40:1360–1364

van Heerwaarden J, Doebley J, Briggs WH, Glaubitz JC, Goodman MM, Sánchez González JJ, Ross-Ibarra J (2011) Genetic signals of origin, spread and introgression in a large sample of maize landraces. Proc Natl Acad Sci USA 108:1088–1092

Vigouroux Y, McMullen M, Hittinger CT, Houchins K, Kresovich S, Matsuoka Y, Doebley J (2002) Identifying genes of agronomic importance in maize by screening microsatellites for evidence of selection during domestication. Proc Natl Acad Sci USA 99:9650–9655

Vitte C, Ishii T, Lamy F, Brar D, Panaud O (2004) Genomic paleontology provides evidence for two distinct origins of Asian rice (*Oryza sativa* L.). Mol Gen Genet 272:504–511

Wang RL, Stec A, Hey J, Lukens L, Doebley JF (1999) The limits of selection during maize domestication. Nature 398:236–239

Wang H, Nussbaum-Wagler T, Li BL, Zhao Q, Vigouroux Y, Faller M, Bomblies K, Lukens L, Doebley JF (2005) The origin of the naked grains of maize. Nature 436:714–719

Wang L, Wang AH, Huang XH, Zhao Q, Dong GJ, Qian Q, Sang T, Han B (2011) Mapping 49 quantitative trait loci at high resolution through sequencing-based genotyping of rice recombination inbred lines. Theor Appl Genet 122:327–340

Watanabe N, Sugiyama K, Yamagishi Y, Sakata Y (2002) Comparative telosomic mapping of homoeologous genes for brittle rachis in tetraploid and hexaploid wheats. Hereditas 137:180–185

Watanabe N, Fujii Y, Kato N, Ban T, Martinek P (2006) Microsatellite mapping of the genes for brittle rachis on homoeologous group 3 chromosomes in tetraploid and hexaploid wheats. J Appl Genet 47:93–98

Wright SI, Vroh Bi I, Schroeder SG, Yamasaki M, Doebley JF, McMullen MD, Gaut BS (2005) The effects of artificial selection on the maize genome. Science 308:1310–1314

Xia Q et al (2009) Complete resequencing of 40 genomes reveals domestication events and genes in silkworm (*Bombyx*). Science 326:433–436

Youens-Clark K, Buckler E, Casstevens T, Chen C, DeClerck G, Derwent P, Dharmawardhana P, Jaiswal P, Kersey P, Karthikeyan AS, Lu J, McCouch SR, Ren L, Spooner W, Stein JC, Thomason J, Wei S, Ware D (2011) Gramene database in 2010: updates and extensions. Nucleic Acids Res 39:D1085–D1094

Yu Y, Tang T, Qian Q, Wang Y, Yan M, Zeng D, Han B, Wu C-I, Shi S, Li J (2008) Independent losses of function in a polyphenol oxidase in rice: Differentiation in grain discoloration between subspecies and the role of positive selection under domestication. Plant Cell 20:2946–2959

Zhang L, Zhu Q, Wu Z, Ross-Ibarra J, Gaut BS, Ge S, Sang T (2009) Selection on grain shattering genes and rates of rice domestication. New Phytol 184:708–720

Zhu Q, Ge S (2005) Phylogenetic relationships among A-genome species of the genus *Oryza* revealed by intron sequences of four nuclear genes. New Phytol 167:249–265

Zhu B, Si L, Wang Z, Zhou Y, Zhu J, Shangguan Y, Lu D, Fan D, Li C, Lin H, Qian Q, Sang T, Zhou B, Minobe Y, Han B (2011) Genetic control of a transition from black to straw-white seed hull in rice domestication. Plant Physiol 155:1301–1311

Chapter 13
High-Throughput and Precision Phenotyping for Cereal Breeding Programs

Boddupalli M. Prasanna, Jose L. Araus, Jose Crossa, Jill E. Cairns, Natalia Palacios, Biswanath Das and Cosmos Magorokosho

13.1 Introduction

Cereals hold unique position in world agriculture as a source of food, feed and diverse products of industrial importance. For several million farmers and consumers in countries with low- and middle-income, cereals (especially rice, wheat and maize) are the preferred staple food crops. The future of cereal production, and consequently, the livelihoods of several million small farmers worldwide, is therefore, dependent to a great extent on developing improved high yielding varieties of cereals. Over the past several decades, conventional breeding, complemented by an array of disciplines, largely contributed to the development of a number of improved cereal varieties adapted to different agro-ecologies and for meeting the diverse demands of the stakeholders.

The challenges to food security are indeed urgent and real due to global climate change, and sharply increasing demands of food and feed, coupled with degradation and scarcity of natural resources (Shiferaw et al. 2011; Cairns et al. 2012). Cereal-growing regions in the world, especially in sub-Saharan Africa and South

B. M. Prasanna (✉) · B. Das
International Maize and Wheat Improvement Center (CIMMYT), ICRAF House, United Nations Avenue, Gigiri, Nairobi 00621, Kenya
e-mail: b.m.prasanna@cgiar.org

J. L. Araus
Unitat de Fisiologia Vegetal, Facultat de Biologia, Universitat de Barcelona, Barcelona, Spain

J. Crossa · N. Palacios
CIMMYT, Km 45 Carretera Mexico-Veracruz, Texcoco 56130, Mexico

J. E. Cairns · C. Magorokosho
CIMMYT, P.O. Box MP 163, Mount Pleasant, Harare, Zimbabwe

P. K. Gupta and R. K. Varshney (eds.), *Cereal Genomics II*,
DOI: 10.1007/978-94-007-6401-9_13,

Asia, are now experiencing rising temperatures, more frequent droughts, excess rainfall/flooding, as well as new and evolving pathogens and insect-pests. Cereal harvests at the current levels of productivity growth will still fall short of demand, unless cutting-edge technologies, including high throughput and precision phenotyping are introduced for accelerating germplasm screening and development of improved cultivars.

13.2 Modern Breeding and the Need for High Throughput and Precision in Phenotyping

In recent years, the technological opportunities for crop improvement have increased significantly. "Molecular breeding" is a general term used to describe modern breeding strategies where genotypic markers are used as a substitute for phenotypic selection to accelerate the development and release of improved germplasm (Ribaut et al. 2010). Molecular marker-assisted breeding largely relies on the identification of DNA markers that have significant association with expression of specific target traits (except in case of genomic selection, also described as genome-wide selection). The main molecular breeding schemes currently being employed are marker-assisted selection (MAS), marker-assisted backcrossing (MABC), marker-assisted recurrent selection (MARS) and genome-wide selection (GWS) (Ribaut et al. 2010). The development and availability of an array of molecular markers, ultra-high-throughput and reduced cost of genotyping assays, and above all, the recent availability of the complete genome sequences of important food crops (e.g., rice, Yu et al. 2002; maize, Schnable et al. 2009; sorghum, Paterson et al. 2009; soybean, Schmutz et al. 2010) within the public domain are providing researchers access to unprecedented genomic information for improving these important crop plants. These developments have significant implications not only to our understanding of the cereal genome organization and evolution, but also in designing and implementing strategies that can utilize the rapidly expanding genomic information for crop improvement.

The use of molecular techniques within breeding pipelines is widely, and successfully, employed within the private sector (Eathington et al. 2007) and with greater emphasis in the public sector (Dwivedi et al. 2010; Whitford et al. 2010). In addition to advances in molecular biology, great strides have been made with regard to technologies such as doubled haploids (DH), which reduce the time taken for developing completely homozygous lines (Röber et al. 2005; Phillips 2009). For example, CIMMYT Global Maize Program, in collaboration with University of Hohenheim, Germany, is making intensive efforts for development of tropically adapted haploid inducer lines in maize with an induction rate of $\geq$10 %. This will allow the production of completely homozygous lines in 2–3 seasons, compared to 8–9 seasons required for inbred line development using conventional breeding (Prigge et al. 2011; Prasanna et al. 2012).

It is now well-recognized by many research institutions (in both public and private sectors) that high throughput genotyping is no longer a major limiting factor, but high throughput and precision in phenotyping is really demanding. Historically large gains have been made through conventional breeding, where the Green Revolution led to large increases in cereal production (Evenson and Gollin 2003). Conventional breeding has resulted in yield gains at a rate of 73 kg ha^{-1} yr^{-1} under mild stress in temperate maize (Duvick 1997) and 144 kg ha^{-1} yr^{-1} under drought stress in tropical maize (Edmeades et al. 1999). In rice, direct selection for grain yield under drought stress has resulted in gains of 25 % relative to unselected lines in both lowland and upland conditions (Venuprasad et al. 2008). Such successes have been partly attributed to the application of proven breeding methodologies in managed stress screening within the target environment (Bänziger et al. 2006). However, in comparison to advances in modern breeding methodologies, advances in phenotyping have been much slower. Phenotyping still remains largely laborious, expensive and rather empirical in assumptions. Therefore, there is a distinct need to invest in improved phenotyping platforms to capitalize on breeding gains (Araus et al. 2008; Cabrera-Bosquet et al. 2012).

Improved phenotyping platforms will provide the foundation for further success of conventional breeding, and for implementation of molecular and transgenic breeding for complex quantitative traits, such as yield and adaptation to abiotic stresses like drought. In fact, the need for precise and high throughput phenotypic data in plant breeding is not new. Long before the genomic era, improved techniques for assessing plant phenotypes were constantly explored in germplasm screening and cultivar development for various traits. But, what has certainly changed is the biological understanding and the technological advances (especially in bioinstrumentation and associated data handling and analysis tools) that makes high throughput and precision phenotyping possible for complementing the tidal wave of genotypic information generated by next-generation genotyping/sequencing technologies. The level of improvement in phenotyping platforms will partly determine the amount of information leveraged from molecular breeding into crop improvement programs. Thus, there is an urgent need to increase the throughput, scale and precision of phenotyping to meet the needs of molecular biologists and breeders.

13.3 High Throughput Phenotyping Platforms

High throughput phenotyping platforms (HTPP) could be particularly useful for obtaining detailed measurements of plant characteristics that collectively provide reliable estimates of trait phenotypes. These platforms are also useful in modelling (especially taking into account ‘hidden variables’) for predicting genotypic performance in different climate scenarios (under controlled experimental conditions). HTTP operations are based on three key criteria: data recording/scoring, speed of data collection, and automation (either partially or fully), and the platform relies

mostly on non-destructive, non-invasive and remote sensing phenotyping tools, together with robotics and automatic data gathering and image processing algorithms. Digital images are primarily gathered from 3D colour imaging (for plant biomass, structure, phenology and health, including chlorosis and necrosis), infrared imaging (for tissue water content and transpiration), and fluorescence imaging (photosynthetic status). Thus, information is correlated with experimental records (e.g., genetic data). Watering and precision stress management as well as pot randomization are also taken into account.

In recent years, there has been an increasing interest in establishing HTPPs not only by the major private sector institutions, which have been pioneering such endeavor, but also by some of the public research institutions worldwide. This is the case for example of the "The Australian Plant Phenomics Facility" which includes, up to now, the "High Resolution Plant Phenomics Centre", placed in Canberra, and "The Plant Accelerator" at the University of Adelaide (http://www.plantphenomics.org.au) (Finkle 2009).

There are companies that are actively engaged in developing such facilities (e.g. LemnaTec, http://www.lemnatec.com) at both hard- and software levels for plant phenomics and high-throughput phenotyping. Following are some examples of HTPPs.

- The **ScanAlyzer** HTPP platform (developed by LemnaTec) focuses on the mature stage of the plant, and has the capability to image plants in a greenhouse by automatically moving plants, placing them on beltways, and positioning them in front of a stereoscopic camera. Proprietary software analyzes the images to extract phenotypic-related information (http://www.lemnatec.com/product/scanalyzer).
- **PHENOPSIS**, a custom growth chamber phenotyping system, was developed by Optimalog, on contract by the Laboratory of Plant Ecophysiological responses to Environmental Stresses, in Montpellier, France (Granier et al. 2005).
- **TraitMill™**, designed and constructed by a Belgian plant biotechnology company CropDesign, is perhaps the world's largest corporate phenotyping facility for applied genomics research and development in cereals (http://www.cropdesign.com/tech_traitmill.php).
- **PhenoFab™** facility, located in Wageningen in the heart of the Dutch Agro Food Center of Excellence: Food Valley, is another HTPP that combines the proprietary second generation automated imaging and conveyor belt system developed and marketed by LemnaTec with the plant breeding and phenotyping algorithm expertise of the Dutch Ag-Biotech company KeyGene (http://www.phenofab.com).
- DuPont-Pioneer's **FAST Corn™** (Functional Analysis System for Traits—Corn) facility at Johnston, Iowa (USA) is a proprietary HTPP that not only enables growing first generation maize plants to maturity in approximately two months (compared to more than 100 days in field conditions), but also has fully-automated digital imaging system to precisely measure the growth of a maize plant throughout its lifecycle.

13.4 The Need for High Throughput and Precision in Field-Based Phenotyping

Although HTPPs are powerful and time-effective, the proprietary platforms require a large investment in the appropriate infrastructure; therefore, their easy deployment and maintenance, especially in the developing world, are of major concern. Beside the huge cost of such facilities, the platforms have the intrinsic limitation of not growing the plants under real field conditions. This limits the application for phenotyping to specific stages of the crop (for example, early vigour).

Currently, there is an increasing interest to develop relatively low-cost, field phenotyping platforms to overcome the limitation of greenhouse or growth chamber phenotyping like that of the High Resolution Plant Phenomics, Centre (http://www.plantphenomics.org/HRPPC). Such field platforms include placement of the remote sensing cameras in aerial platforms such as balloons, zeppelins or remote-controlled airplanes or "polycopters" (e.g., http://www.mikrokopter.de/ucwiki/en/MikroKopter) or using light curtains and spectral reflectance sensors mounted on a tractor for evaluating crop performance under field conditions (Montes et al. 2011). Again, one of the main limitations of such platforms is cost.

13.5 Generating High-Quality Data from Field Phenotyping

As pointed out above, field phenotyping is a time consuming and labour-intensive exercise, which still keeps a high degree of empiricism on the traits to choose and the way to assess them. This is particularly true, when the targets for breeding is improving yield potential, or enhancing adaptation to abiotic stresses or any other complex quantitative trait. For a secondary trait (i.e., other than yield itself) to be useful in a breeding program, it must comply with several requirements. Among them are: (1) genetic variability for the trait of interest (2) high genetic correlation with grain yield in the target environment, (3) less affected by environment than grain yield, and (4) rapid and reliable measurement, which is less expensive than measuring yield itself. These requirements are detailed elsewhere (e.g., Bänziger et al. 2000; Araus et al. 2008 and references herein). Moreover, the phenotyping strategy, which includes experimental design, traits and tools, will also depend on the target environment for selection. We outline here some ideas concerning (1) how to tackle the limitations inherent to field site variation, (2) the selection strategies to adopt, and (3) the traits and tools to implement. These important aspects shall be discussed with examples from maize, although the principles will also apply to a large extent to other cereal crops.

13.5.1 Characterizing and Controlling Site Variation in Precision Field Phenotyping

The phenotyping environment plays a vital role in the quality of phenotypic data generated through experiments, and consequently, the efficiency of breeding. Highly variable field sites produce highly variable data, thereby masking important genetic variation for key traits and reducing repeatability, regardless of the cost and precision of a specific phenotyping protocol.

Phenotypic variation among individuals is a result of both genetic and environmental factors. Heritability is specific to a specific population within in a specific environment and can be reduced due to increased environmental variation without any genetic change occurring. Broad-sense heritability can be defined as the proportion of phenotypic variation that is due to genetic variation (Falconer and Mackay 1996). As the phenotypic variation of a population is caused by both genetic ("signal") and environmental factors ("noise"), broad sense heritability estimate provides a useful understanding of the proportion of phenotypic variance that can be attributed to genetic effects. Broad-sense heritability is population specific within a particular environment and typically decreases with increased site (environmental) variability. Without an increase in the "signal-to-noise" ratio within experimental sites, the breeding process cannot be properly optimized, and the power of genomics cannot be fully exploited.

Soils are highly heterogeneous in terms of physical, chemical and biological properties. While lack of uniform experimental fields could significantly reduce genetic gains, the importance of site uniformity and the measures to understand and deal with soil heterogeneity for field phenotyping have often been overlooked. The most common causes of field variability are related to inherent soil variability and agronomic practices (Blum 2011). Additional factors that influence within-site variability include topography, and site history (Kaspar et al. 2003, 2004). Knowledge of this variability is essential to eliminate the use of highly variable sites prior to the initiation of expensive phenotyping efforts, or to develop management strategies and experimental designs to reduce variability. For example, in phenotyping for abiotic stress the main sources of spatial variability are variation in soil texture (drought and low N phenotyping), topography (water-logging phenotyping) and residual soil nitrogen (low N phenotyping).

The water holding capacity and water release characteristics of soils are largely determined by soil texture, with sandy soils having much lower water holding capacity and releasing more of their water at low suctions than clayey soils (Marshall et al. 1996; Table 13.1). Soil texture also plays an important role in determining soil penetration resistance during drought stress (Bengough and Mullins 1990). In soils with high sand content, the increase in penetration resistance with concomitant drying is small compared with soils with a lower sand content (Cairns et al. 2009, 2011). Variability in soil strength variability is likely to lead to variability in the depth of soil roots could exploit thereby influencing response to drought stress potentially bias results. For phenotyping under low N, sandy soils are adequate since generally they have low levels of mineral N and organic matter. Since

Table 13.1 Average water content as a volume fraction of sand, loam and clay soils (adapted from Marshall et al. 1996)

Texture	Clay content (%)	Water content (%)		
		Field capacity	Permanent wilting point	Plant available water
Sand	3	0.06	0.02	0.04
Loam	22	0.29	0.05	0.24
Clay	47	0.41	0.20	0.21

mineral N is derived from soil organic matter, the higher the soil organic matter level the greater the amount of N will be formed by mineralization each season.

Causes of Site Variability: Site Characterization

While improved statistical designs and tools can reduce the influence of environmental noise results from soil variability within field trials, knowledge of the causes of field variability can be used as complementary approach. Knowledge of field site variability prior to experimentation can be used (1) to exclude sites, where large experimental error is likely to be introduced through highly variable soil properties, and (2) reduce experiment or within replicate environmental error, by avoiding areas of high spatial variability and/or blocking individual trials within gradients (Cairns et al. 2004, 2009). Destructive soil sampling for key soil physical and chemical properties can provide information on the unsuitability of a site but lack detailed information for spatial delineation of variability. The quantification of soil spatial variability is commonly used within precision agriculture to identify within-field management options but is rarely used within public breeding programs to improve phenotyping precision. Many techniques are available for mapping variability within field sites based on soil sampling, soil sensors and measurements of plant growth as surrogates of variability. Geostatistical methods (e.g., kriging) play an ever increasing role in precision agriculture and over the years have become an integral component of statistical tools for spatial analyses and a very valuable tool for site characterization under drought and low N.

Given the influence of soil texture within drought and nitrogen experiments, the ability to identify gradients of texture within fields is important. Soil conductivity is closely related to texture and electromagnetic surveys can be used as to determine high resolution variability in soil texture (Anderson-Cook et al. 2002; Johnson et al. 2003; Sudduth et al. 1997). In phenotyping under low N conditions, variability in plant growth is likely to be related to variation in mineral N content.

An example on the use of geostatistical kriging methods by the CIMMYT Global Maize Program for measuring the variation in soil electrical conductivity (EC) in maize drought phenotyping sites in Chirdezi (Zimbabwe) is presented in the Fig. 13.1. EC was measured on a grid basis of 2 and 2 m separation using an EM38 sensor (Geonics Ltd, Canada) in dipolar mode. This site had previously

been used for one season; however, plant growth was found to be highly variable. Large variation in EC was identified within this field site, in agreement with the high variability observed by breeders and the potential use of EC to eliminate potential drought phenotyping sites prior to use.

In low N trials, a single genotype is normally planted across the entire field to deplete residual soil N. The use of a single genotype during this depletion cycle can be exploited for site characterization by dividing the field into small plots and rapidly measuring above ground biomass visually or indirectly using the

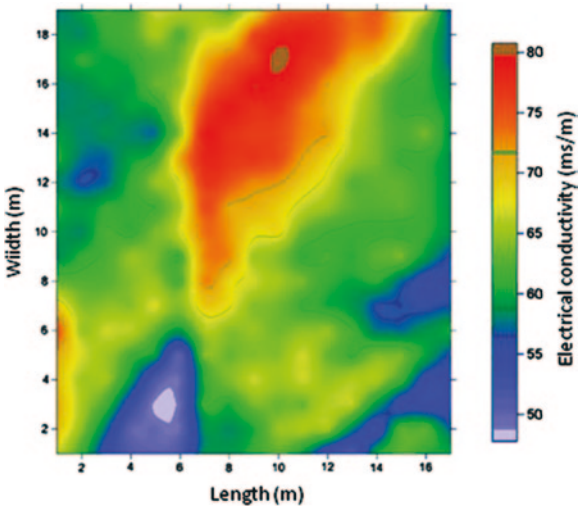

Fig. 13.1 Variation in electrical conductivity as measured by an EM38 soil sensor within a field site in Chiredzi, Zimbabwe

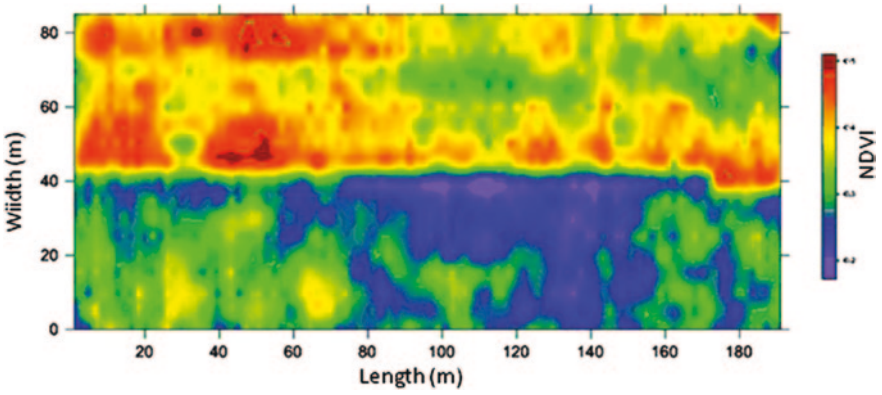

Fig. 13.2 Variation in above-ground biomass of maize seedlings, estimated indirectly through normalized differential vegetation index (*NDVI*), during a N depletion cycle in a low N phenotyping field block in Tlatizapan, Mexico

Normalized Difference Vegetation Index (NDVI) as explained elsewhere in this Chapter. In the CIMMYT Maize Experimental Station at Tlatizapan, Mexico a single commercial maize hybrid was planted in a field for N depletion and measurements of NDVI were taken 21 days after sowing. Variation in biomass, estimated through NDVI, is presented in Fig. 13.2. A large radient in biomass was identified across the field, and subsequent experiments were planted within each block to minimize variation due to field gradients.

A Priori Control of Site Variability: Experimental Design

Spatial variability can be partly controlled by using appropriate experimental design. Block designs (complete or incomplete) attempt an a priori reduction of the experimental error considering spatial heterogeneity among blocks. Control of local variability can be achieved by blocking. Blocking is the arrangement of experimental units into groups (referred to as blocks) that are similar to one another and is used to reduce or eliminate the contribution of noise factors within the experimental error. The basic concept is to create homogeneous blocks in which the "noise" factors are held constant while the factor of interest ("signal") is allowed to vary. The complete block should be as homogeneous as possible in terms of soil and other environmental factors. Often with experience, researchers know the direction of large gradients within a field; this knowledge can be used in appropriate design of experiments. However, it is clear that smaller block sizes will reduce soil heterogeneity within complete blocks, with larger block sizes increasing the chances of soil heterogeneity within blocks. In drought and low nitrogen experiments strong gradients can be observed across the field due to topography. An example of a soil gradient is presented in Fig. 13.3 with a gradient across the field, from top to bottom. Three different options for laying out an experiment with 6 treatments (T_1–T_6) in one replicate of a randomized complete block design are illustrated (Fig. 13.3). Treatments are sown perpendicular

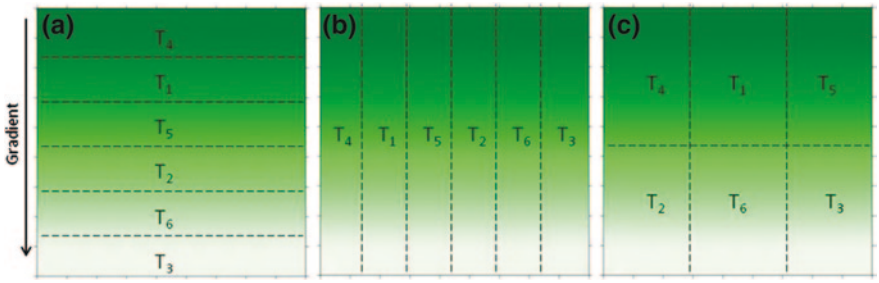

Fig. 13.3 Experimental design options for an experiment with 6 treatments (T) within a field with a known gradient across the field, (**a**) treatments in a complete block are sown perpendicular to the gradient, (**b**) treatments in a complete block sown parallel to the gradient, and (**c**) treatments in a complete block sown by reducing the length of treatments along the gradient and increasing the width of treatments across the gradient

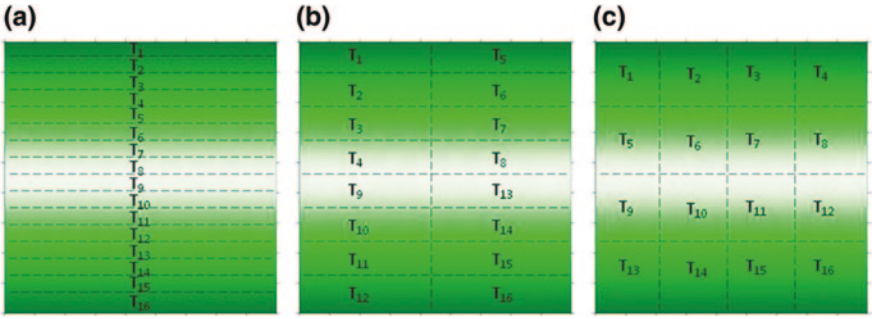

Fig. 13.4 Experimental design options for an experiment with 16 treatments (T) within a field with a known gradient across the field using (**a**) a randomised complete block design, (**b**) four incomplete blocks of 4 treatments each, and (**c**) four incomplete blocks of 4 treatments each While the examples presented above can be used to reduce the influence of spatial variability within trials, relatively spatial variability within trials can arise from multiple factors across the field which complete and incomplete block designs cannot remove. In the 1990s mixed linear models theory was developed incorporating nearest neighbour (spatial) analysis as well as geostatistical tools such as kriging to investigate the variance structure of field trials and to use the appropriate structure to estimate the effects of different factors that may affect the genotype means and their contrasts

(Fig. 13.3a) and parallel (Fig. 13.3b) to the gradient, and split across the gradient (Fig. 13.3c). Figure 13.3b and c represent optimal designs for this gradient as the noise within each treatment is similar; in contrast the design used in Fig. 13.3a will increase unwanted differences between treatments due to field noise.

However, as the number of treatments increase, the ability to find a relatively homogeneous complete block reduces and the use of another design is required. In this situation blocks (or replicates) should be as compact as possible. Figure 13.4 shows the possible layout in the field of a randomized complete block design (RCBD) and the layout of an incomplete block design of four incomplete blocks of size 4. The layout of the RCBD covers a band of the field that accounts for soil gradient on the upper and lower parts of the field. The two possible layouts of the 4×4 incomplete blocks will control local variability in a much more efficient manner.

A Posteriori Control of Site Variability: Spatial Analyses

A posteriori control of the residual effect using a model that provides a good fit to the data can effectively complement the control of local variability provided by the experiment design. Recently, efficient experiment designs (both unreplicated and replicated) have been developed based on the assumption that observations are not independent in contiguous plots in the field and may be spatially correlated.

The concept of adjusting plots for spatial variability using information from neighboring plots was conceived over 70 years ago (Papadakis 1937); however, the statistical theory was not fully developed until 1990s (Cullis and Gleeson 1991).

One robust method for studying patterns of soil heterogeneity is spatial autocorrelation of neighboring plots within both rows and columns. This method uses autocorrelation between residuals at various distances apart; if there is no spatial pattern most of the correlations will be low, whereas if there is pattern within the residuals, the neighboring residuals will be more similar and, therefore, will exhibit a higher correlation. The two directions (row and column) model assumes that within a field, rows and columns are regularly spaced. The practice of assigning every plot a grid coordinate based on row and column positions and performing spatial analysis considering the correlation among residuals between plots has increased precision of trials under drought and low N (Lafitte and Edmeades 1994).

13.5.2 Selection Strategies

Yield Potential Versus Stress Adaptation: Are They Mutually Exclusive?

The breeding value of any trait depends on the growing conditions (i.e., soil and climate) of the target region (Araus et al. 2008, 2011; Tardieu and Tuberosa 2010). Improvement in stress tolerance in the recently developed over the old maize hybrids have been demonstrated under various conditions, including high plant population density, weed interference, low night temperatures during the grain-filling period, low soil N and low soil moisture (Tollenaar and Wu 1999; Bänzinger and Araus 2007), and under abiotic stresses, such as heat and drought, excessively cool and wet weather, low soil fertility, and high density planting (Duvick 1997, 2005). Selection of maize for high yield potential in itself has led to consistent increases in yield in both stress and non-stress conditions (Castleberry et al. 1984; Blum 2011). Recent maize hybrids had relatively better capacity than older ones to access soil water under drought stress (Hammer et al. 2009). Therefore, improved yield potential may also translate to better performance under stress. However, the genetic progress can be achieved by taking different strategies (Araus et al. 2008; Collins et al. 2008). For example, under moderate water deficit stress (up to 60–70 % of reduction in potential yield) there are complementary strategies which optimize total yield, sometimes at expenses of yield stability:

1. Increased yield potential (i.e., in absence of stress) through for example an increased transpiration rate or through a stay-green pattern. This may translate to a higher biomass accumulation and grain yield under mild to moderate stresses, but at the risk of crop failure following excessive soil dehydration.
2. Maintaining growth, and thus, biomass accumulation under decreasing water status. Again, this strategy may also expose plants to increased risk of excessive soil dehydration.

For more severe water stress scenarios there are other strategies which focus on reducing the risk of total yield lost by decreasing cumulative transpiration, which imply a penalty in yield potential and performance under mild to moderate water

stress. In other words, such strategies increase yield stability in exchange of a lower yield potential. These strategies includes, for example:

3. Shortening the duration of the crop cycle (i.e., phenological adjustment) to escape from the drought period. This strategy has proved most successful for C3 cereals such as wheat or barley in environments with terminal drought such as those of the Mediterranean basin (Araus et al. 2002).
4. Reducing leaf area or further stomatal conductance, which frequently increases the water-use efficiency (i.e., the amount of harvested biomass per unit transpired water).

Summarizing any trait can have different impact (positive, negative or neutral) on yield, depending on the drought scenario. Water-conservation traits have beneficial effects in most severe scenarios of drought or in terminal drought scenarios, in which it is useful to save water for the end of the crop cycle and to decrease the amounts of sinks for obtaining larger seeds. These 'conservative' traits are, in particular, low stomatal conductance, low leaf growth rate, high water-use efficiency or deep but sparse root systems. Conversely, growth 'maintenance' traits have beneficial effects in milder drought scenarios in which the soil profile is periodically re-watered. Nevertheless, this trait-by-environment interaction is not universal in the sense that there are cases where constitutive traits (i.e. expressed in absence any apparent stress) confers a better adaptation at any water stress level (Tambussi et al. 2005; Slafer and Araus 2007). In other words, there are examples of very low existence of a genotype-by-environment interaction. This is also the case of heterosis in maize, where hybrids show better water status and higher yield than inbreds no matter the water regime during grown (Araus et al. 2010).

Increasing Water Use Efficiency or Water Capture?

Photosynthetic-driven plant biomass accumulation is intrinsically linked to transpiration because stomatal aperture and leaf area determine the rate of both processes. There is therefore an inherent conflict between biomass accumulation and stress avoidance via reduction of transpiration (Araus et al. 2002, 2008, 2011; Blum 2005, 2009; Lopes et al. 2011). Moreover, yield stability and general stress tolerance in new temperate maize hybrids are highly associated and yield stability does not appear to have declined with increasing yield potential (Duvick et al. 2004; Tollenaar and Lee 2002).

Biomass production is tightly linked to transpiration and nitrogen accumulation (Bänziger et al. 1999). Improved Water Use Efficiency (WUE) defined as the ratio of biomass accumulated to water transpired, is often considered to be a key determinant for improved drought tolerance (Tambussi et al. 2007). However, Blum (2005, 2009) argued that breeding for high WUE under drought conditions will result in low yielding genotypes with reduced drought tolerance because WUE is increased by reduced transpiration and water use. Therefore, biomass production under most drought conditions (i.e., except for marginal drought-prone areas) can

only be enhanced by maximizing soil moisture capture for transpiration, which also involves minimized water loss by soil evaporation This has been defined by Blum (2009, 2011) as Effective Use of Water (EUW).

Since biomass production is tightly linked to transpiration, breeding for maximized soil moisture capture for transpiration is a most important target for yield improvement under drought stress. In addition, EUW by way of improving plant water status helps sustain assimilate partition and reproductive success resulting in increased harvest indices. This is particularly important for maize, since the reproductive stage is the most affected by drought. Herein we will focus mostly to these traits which improving EUW.

Some important traits and mechanisms that confer enhanced performance of maize genotypes under mild to moderate water stress are included in the following section. More details on specific traits for selection may be found elsewhere (Araus et al. 2008, 2011; Blum 2009, 2011).

13.6 Traits and Tools

13.6.1 Increased Transpiration: Root System Architecture and Efficiency

Plant root system architecture is plastic and dynamic, allowing plants to respond to their environment in order to optimize acquisition of important soil resources. Root system architecture (RSA) is dependent on a number of genetic and environmental factors, including soil temperature, soil mechanical resistance, soil water content, C and N status of the plant, and nutrient availability. Root characteristics may play a pivotal role as stress adaptive traits enhancing soil moisture capture and transpiration. It may seem obvious that increasing soil exploration by the root system is a positive feature for improving drought tolerance. Indeed, genomic regions controlling root system architecture in controlled conditions also determine yield in a drought stress field experiment (Tuberosa et al. 2002). However, several breeding programmes for drought tolerance have resulted in a decrease in the weight of the root system (Bruce et al. 2002; Campos et al. 2004). Bolaños et al. (1993) found a significant association between reduced root mass and increased ear growth under drought in one tropical population improved for drought tolerance. Edmeades et al. (1999) demonstrated that root mass can be increased by recurrent selection but were also unable to reveal a positive relationship between root ramification and yield under drought.

Nevertheless, it seems that the spatial distribution of the roots in the soil is the one which defines the ability of a root system to take up water, and not root mass per se (Draye et al. 2010). An ‘ideal’ root system, with an even distribution of ca. 1 cm root per cm^3 of soil (root length density—RLD) over the rooting depth, is suitable for most field conditions (Tardieu et al. 1992). Higher RLD are adequate when evaporative demand increases and soil water reserve decreases, but too high RLD values ultimately result in a waste of photosynthates without

appreciable increase in water uptake. In defining target types for root architecture and efficiency, the water retention capacity and conductivity properties of the soil, as well as the hydraulic properties of the roots, are crucial elements to be considered (Javaux et al. 2008). Root angles and branching may also have a basic role in improving water acquisition by plants (Trachsel et al. 2011).

Many screening techniques have been used to allow root traits to be quickly and reproducibly assessed under controlled conditions including aeroponics, hydroponics, wax-petrolatum layer, soil filled chambers or pipes (for reviews see Tuberosa et al. 2003; Gowda et al. 2011). While these systems offer fast, reproducible screens for root traits, the soil environment is highly heterogeneous. It is important to consider the degree to which root phenotypes will be expressed in the target environment, particularly when considering the improvement of drought tolerance via modification of the root system architecture. Excavation techniques, including soil cores and profiles, have been used to study root systems in the field (Samson and Sinclair 1994; Cairns et al. 2004). While these techniques have been used to screen mapping populations for the identification of QTL associated with root traits in the field (Yue et al. 2005; Cairns et al. 2009), they are very laborious and heritability is low.

Trachsel et al. (2011) developed and presented a novel approach called "Shovelomics" that enables visual scoring of 10 root architectural traits of the root crown of an adult maize plant in the field in a few minutes. The authors reported a sample processing time of three (sand) to 8 min (silt-loam). Initial unpublished results from CIMMYT suggest steeper root angles and less branching are associated with traits associated with drought tolerance in the field. This is in agreement with previous studies which found drought tolerance was associated with reduced root growth in the resource-poor zones (Hund et al. 2009). Iyer-Pascuzzi et al. (2010) described a nondestructive imaging and analysis system for automated phenotyping and trait ranking of root system architecture.

Improvements in phenotyping, including plant imaging under real field conditions, will facilitate the genetic analysis of root architecture and aid in the identification of the genetic loci underlying useful agronomic traits (Zhu et al. 2011). "Rhizotrons" consisting of long tubes placed outdoors (with rain shelter) is a proper alternative for root studies, allowing to assess simultaneously the water balance of plants (Fig. 13.5).

Yazdanbakhsh and Fisahn (2009) described a range of applications of a recently developed plant root monitoring platform (PlaRoM). PlaRoM consists of an imaging platform and a root extension profiling software application. PlaRoM can investigate root extension profiles of different genotypes in various growth conditions (e.g., light protocol, temperature, growth media). Simultaneous with these developments, several software packages have been developed to automate the analysis of root traits in minirhizotron images, including WinRhizoTRON (www.regentinstruments.com), RootView (www.mv.helsinki.fi/aphalo/RootView.html), RooTracker (www.biology.duke.edu/rootracker), and MR-RIPL (http://rootimage.msu.edu). To quantify root growth kinetics, commercial software such as WINRHIZO (Arsenault et al. 1995), OPTIMAS analysis software (Media Cybernetics, Bethesda, MD, USA) or IMAGE J (Abramoff et al. 2004) were also introduced.

Fig. 13.5 Rhizotronic facility with maize genotypes placed outdoors (at ICRISAT, Hyderabad, India) that allows assessing water balance in plants using a crane

Searching for root characteristics associated to heterosis could be another promising research avenue to detect traits conferring a better EWU. Heterosis in maize, while associated with higher yield potential, also confers adaptation under a full range of growing conditions. A CIMMYT study has shown that hybrids have better water use than inbreds, regardless of the water conditions during cultivation (Araus et al. 2010). Heterosis for growth and yield in maize may (at least in part) be mediated by accumulated differences in water use and status, such differences being present even under well-watered conditions. Hoecker et al. (2006) demonstrated in maize that heterosis can already be seen during the very early stages of root development, a few days after germination. In that study lateral root density showed the highest degree of heterosis. Chun et al. (2005) reported heterosis for total length of lateral roots together with a higher ratio of lateral to axial root length in 20 days maize seedlings growing under different N conditions. Later in development, lateral roots become dominant and are responsible, together with the post-embryonic shoot-borne root system, for the major portion of water and nutrient uptake (Hochholdinger et al. 2004). In addition, Li et al. (2008) concluded that seedlings of hybrids develop a greater number of fine roots than their parents. Furthermore, heterosis appears to exist for water uptake ability at the root cell level under well-watered conditions (Liu et al. 2009). However, all the above (and other) studies were performed on seedlings, usually under controlled conditions, while no systematic studies on root traits associated to heterosis have been performed on adult plants in the field.

13.6.2 Growth Maintenance

A decrease in leaf growth is the first process occurring under water deficit, before reduction in stomatal closure (Saab and Sharp 1989; Araus et al. 2008). A large genetic variability for sensitivity to leaf expansion has been detected in temperate and tropical maize (Welcker et al. 2007). A high sensitivity of leaf grown to water deficit is a ‘stress avoidance’ mechanism which makes plants more tolerant to severe

drought scenarios by saving soil water and making it available for the end of the crop cycle. Leaf rolling or epinasty has essentially the same role by decreasing the active leaf area, but on short-term and in a revertible way. However, high sensitivity to these processes has two drawbacks. First, leaves are also the site of photosynthesis, so a drought-induced decrease in leaf area causes a reduced rate of biomass accumulation. Second, the genetic determinism of leaf growth is partly shared with that of reproductive growth (Welcker et al. 2007), thereby reducing the sink source.

A large number of mechanisms underlying growth maintenance have been proposed (e.g., cell cycle, hormonal regulation or hydraulics or cell wall mechanical properties), without compelling evidence for any of them. However, some candidate mechanisms may be considered as best choices, in particular plant hydraulic properties and their control by aquaporins. Aquaporins are proteins which facilitate water transport through membranes, thereby increasing the hydraulic conductivity of tissues when their channel is open. Both the amount of aquaporins and their gating contribute to the control of hydraulic properties when plants are challenged by the plant hormone abscisic acid, by osmotic stress or by anoxia (Boursiac et al. 2008; Tournaire-Roux et al. 2003). The functional analysis of the aquaporin family has been quarried out in several species including maize (Hachez et al. 2008). Osmotic adjustment is another mechanism, which maintains turgor under drought or osmotic stress, thereby maintaining growth (Zhang et al. 1999) but there are reports indicating that this mechanism is not so relevant in maize (Tardieu 2006). Some of the above mechanisms may apply too for increasing transpiration and yield potential in absence of stress.

13.6.3 Stay-Green

Stay green i.e., the ability of genotypes to maintain a green leaf area at the end of the crop cycle, has been shown in sorghum to be most often a consequence of earlier traits, namely a slow leaf growth rate which saves water for the end of the crop cycle and a deep root system (Harris et al. 2007). However, the strongest evidence in favour or stay-green in maize as a factor involved in adaptation to abiotic stresses comes from retrospective studies comparing temperate maize hybrids produced during the last 70 years in the USA (Tollenaar and Lee 2006). The higher dry matter accumulation during grain filling in the new versus the old hybrids can be attributed, in part, to a longer duration of the grain-filling period in the former. In addition, the newer hybrids are better adapted to higher planting densities (Tollenaar and Lee 2006).

Therefore, the higher yield potential of the more recent hybrids is related with a higher stay-green expression together with the capability to growth under higher planting density. This suggest that stay-green, beside its positive role in increasing the amount of radiation used and then of assimilates produced by photosynthesis, is a consequence of the higher capability of these hybrids to cope with limited nutrient and water resources (Tollenaar and Lee 2006). Of course, stay-green without water available to keep stomata open and photosynthesis active does not

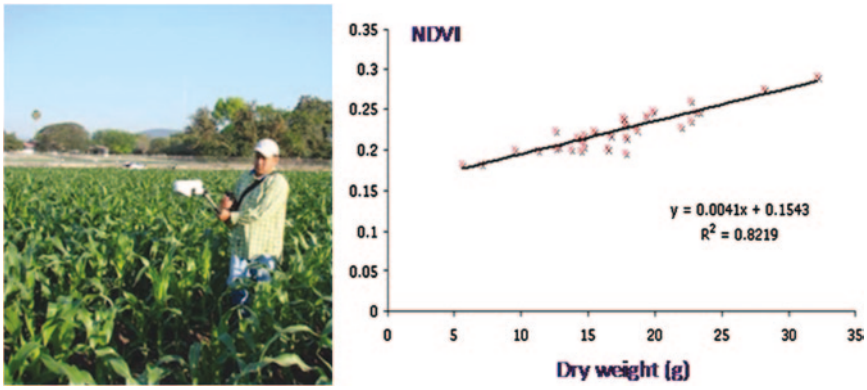

Fig. 13.6 Use of a portable spectroradiometer (GreenSeeker^{TM}) to assess in a fast and non-destructive manner aerial biomass in maize using the normalized differential vegetation index (*NDVI*)

provide any advantage to the plant; rather, it could have an opposite effect in terms of grain quality (e.g., lower N content).

There is evidence that supports higher tolerance of low resource availability in newer maize hybrids associated with better tolerance to high plant density (Tollenaar and Lee 2006). In fact, plant water deficit will occur more readily at high rather than at low density, and resistance to high plant density involves resistance to drought stress when moisture becomes limiting (Tollenaar and Wu 1999). The constitutive nature of stay-green together with its expression in a wide range of growing conditions is also placed in evidence in heterosis, where hybrids remain green longer during grain filling no matter the water conditions considered (Araus et al. 2010).

Stay green and delayed senescence may be easily and fast assessed using a visual ranking or being quantified through remote sensing spectroradiometrical approaches, either at the leaf level (e.g., using a portable leaf chlorophyll meter such as the SPAD^{tm}) or at the whole canopy level using portable spectroradiometers (Araus et al. 2008), such as, for example, the GreenSeeker^{tm} (Fig. 13.6). Maintained growth under drought may be also easily assessed with a portable spectroradiometer, particularly when evaluating inbred lines (Lu et al. 2011).

13.6.4 Seed Abortion and Early Seed Growth

Effective seed number is a major component, mainly determined at flowering and slightly after it. In most species, the number of ovules largely exceeds the number of seeds, and water deficit reduces even more the ratio seed/ovule via abortion (Barnabas et al. 2008). This is again an adaptive mechanism which allows the remaining seeds to be appropriately filled in spite of reduced photosynthate supply, with a positive or negative effect on yield depending on the drought scenario.

However, seed abortion under stressed or non-stressed conditions is a negative trait for both tropical and temperate maize.

Carbon metabolism and carbon transport make a significant contribution to early seed growth and abortion, as shown by experiments with sucrose feeding and metabolite measurements (Zinselmeier et al. 1995; McLaughlin and Boyer 2004). However, this effect is not due to a straightforward "sugar hunger" as sugar content is usually increased in leaves and maintained in ovules of droughted plants. It is probably linked to a disruption of metabolic pathways, or of the distribution of sugars in the plant or of combined effect of these metabolic disruptions with changes in ear development (Carcova and Otegui 2007). Other processes as a delayed silk growth (anthesis-silking interval; ASI) in maize also have a major effect on seed abortion and seed number (Bolaños and Edmeades 1996), and have been improved genetically in the last 50 years (Duvick et al. 2005) which apparently makes difficult a further substantial improvement in maize drought adaption based in this trait (Monneveux et al. 2008). It is also not easy to elucidate if a larger ASI rather than the cause of a decrease in grain number and yield is just a consequence of drought stress diminishing availability of assimilates for growing silks (discussed above) and/or affecting water status and thus the cell expansion of this tissue. Thus, Welcker et al. (2007) concluded that the genetic control of ASI under water deficit is partly shared with the sensitivity of leaf elongation rate, thereby contributing to the interest of alleles maintaining leaf growth under water deficit.

13.6.5 Measuring Water Use: Stable Isotopes and Low Cost Surrogates

As pointed out above, except under very severe drought stress conditions, EWU is a more important adaptive trait than WUE (Araus et al. 2008; Blum 2005, 2009; Slafer and Araus 2007). This would be related to the genotypic capacity to use available water and, therefore, to sustain transpiration under unfavorable conditions (Slafer and Araus 2007). Oxygen isotope composition ($\delta^{18}O$) measured in plant matter has been proposed as a time-integrative indicator of transpiration and EWU in different species (Barbour, 2007; Farquhar et al. 2007; Cernusak et al. 2007, 2009; Cabrera-Bosquet et al. 2009a) including maize (Cabrera-Bosquet et al. 2009b; Araus et al. 2010). Moreover, associated genotypic differences in $\delta^{18}O$ and grain yield have reported in maize (Cabrera-Bosquet et al. 2009b). Cabrera-Bosquet et al. (2009b) also reported significant negative relationships between $\delta^{18}O$ and grain yield under both well-watered and moderated water stressed conditions, meaning that these genotypes transpiring more were the most productive.

Given the cost, technical skills, and facilities involved in oxygen isotope analysis, its large-scale application to breeding programs may be unfeasible for many breeding programs. The accumulation of mineral or ash content in vegetative tissues has been proposed as a cost-effective and easier way to predict yield and genotypic adaptation to drought in different cereals, including maize (Cabrera-Bosquet et al. 2009c).

The mechanism of mineral accumulation in vegetative tissues seems to be explained through the passive transport of minerals via xylem driven by transpiration (Tanner and Beevers 1990; Masle et al. 1992; Mayland et al. 1993; Araus et al. 1998, 2002). In such a way it has been reported in tropical maize a close relationship between ash content in leaves and isotope enrichment in leaves and grains (Cabrera-Bosquet et al. 2009c). Based on these results, it could be concluded that ash content in leaves could be used as an alternative criterion to $\delta^{18}O$ for assessing yield performance in maize grown under drought conditions. Ash (or total mineral content) analysis is not costly and do not require strong facilities and a high qualified technical staff, which make this trait ideal for breeding programs in developing countries.

Heterosis in maize seems to confer a higher transpiration in the hybrids compared with the inbred lines regardless the growing conditions, as shown by the lower $\delta^{18}O$ and higher ash content of hybrids compared with inbreeds (Araus et al. 2010) and the linear negative relationship between both traits (Fig. 13.7). Combining data from both hybrids and inbreds, $\delta^{18}O$ of kernels is strongly negatively correlated and leaf ash content strongly positively correlated with total biomass and grain yield (Figs. 13.8 and 13.9).

The above studies on maize conducted by CIMMYT (Cabrera-Bosquet et al. 2009b,c; Araus et al. 2010) and other unpublished results, using CIMMYT germplasm support the potential usefulness of $\delta^{18}O$ in kernels and ash content in leaves as secondary traits to indirectly select for maize genotypes with improved performance under drought through a higher EWU. Moreover near infrared reflectance spectroscopy (NIRS) technique may allow a fast, cheap and non-destructive analysis of ash content and $\delta^{18}O$ in maize samples (Cabrera-Bosquet et al. 2011) which may speed the phenotyping process However, further studies are needed to investigate the extent of genotypic variance, heritability and genetic correlation of both

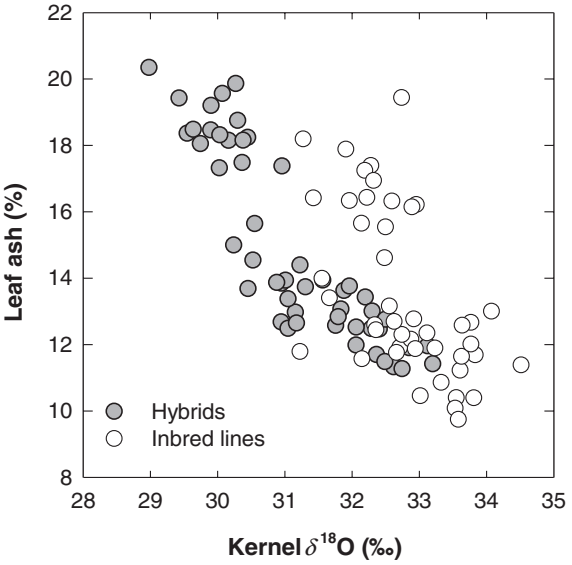

Fig. 13.7 Relationship between oxygen isotope composition in mature kernels ($\delta^{18}O$) and ash concentration in leaves about two weeks after anthesis. Data from a set of maize inbred lines and derived hybrids grown under three different water regimes were plotted together ($n = 96$). Each point represents a mean value for three plots of a single genotype grown under a particular water regime (redrawn from Cabrera-Bosquet et al. 2009c; Araus et al. 2010)

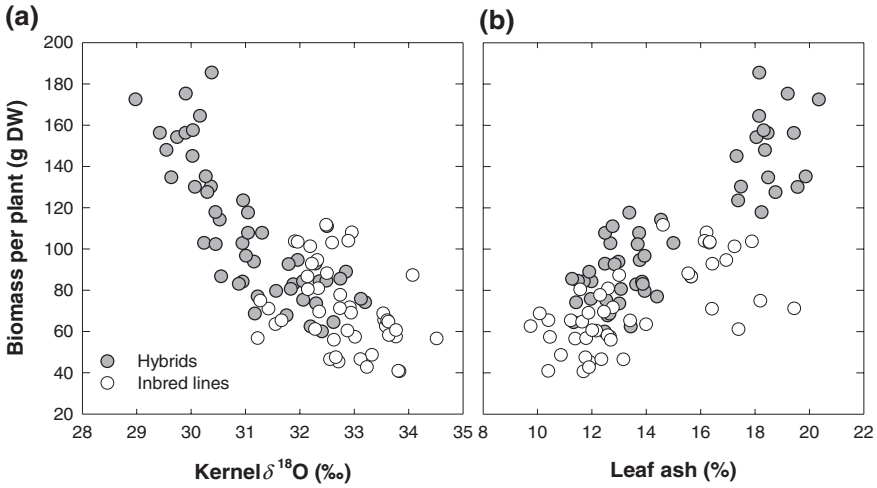

Fig. 13.8 Relationship between biomass per plant about two weeks after anthesis with **a** oxygen isotope composition in mature kernels (δ^{18}O) and **b** ash concentration in leaves about two weeks after anthesis. Data from a set of maize inbred lines and derived hybrids grown under three different water regimes were plotted together ($n = 96$). Each point represents a mean value for three plots of a single genotype grown under a particular water regime (redrawn from Araus et al. 2010)

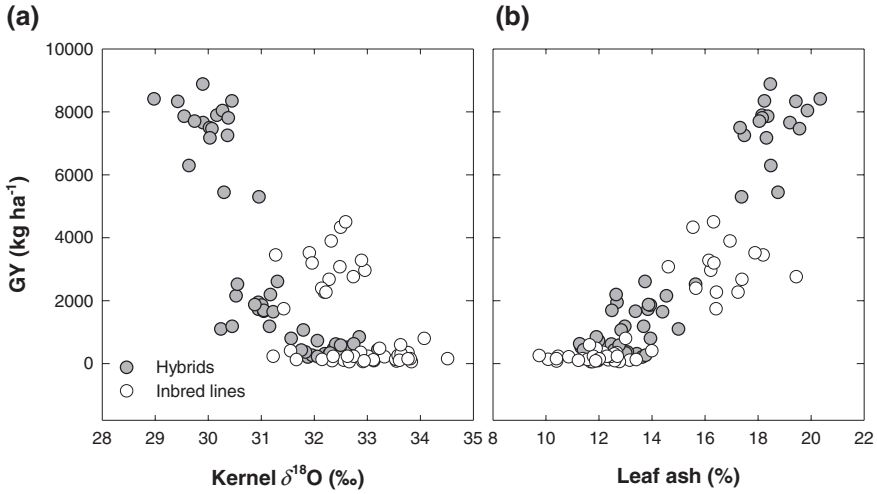

Fig. 13.9 Relationship between grain yield with **a** oxygen isotope composition in mature kernels (δ^{18}O) and **b** ash concentration in leaves about two weeks after anthesis. Data from a set of maize inbred lines and derived hybrids grown under three different water regimes were plotted together ($n = 96$). Each point represents a mean value for three plots of a single genotype grown under a particular water regime (redrawn from Araus et al. 2010)

$\delta^{18}O$ and ash content with grain yield. Additionally, it is important to elucidate the root traits (if any) responsible for the genotypic variation in $\delta^{18}O$ and mineral content. A field application of NIRS deals with the direct empirical assessment of grain yield based in the use of the entire array of reflected wavelengths captured by the spectroradiometer.

An additional tool to assess water uptake capacity of the root system is the use of the isotope composition of oxygen ($\delta^{18}O$) and hydrogen ($\delta^{2}H$) in stem water (Ehleringer and Dawson 1992). This technique is based on the existence of evaporative gradients in the isotope composition of soil water, which, due to evaporation, tends to be more enriched in the heavier isotope in the soil surface than in deeper layers. Comparing the isotopic composition of stem water (i.e., the water taken by the plant) with that of soil at different depths, it is possible to determine which part of the root system is contributing more to the water uptake (see e.g., Ferrio et al. 2005; Holst et al. 2010; Patz 2009). In addition, previous works suggest that genetic differences in water uptake are significant and may be partly responsible of observed genetic variability in $\delta^{18}O$ of plant tissues (Barnard et al. 2006; Ferrio et al. 2007; Retzlaff et al. 2001; Voltas et al. 2008). Traditionally, stable isotopes have been determined by isotope ratio mass spectrometry (IRMS), which is an expensive technique requiring costly equipment and a complex laboratory infrastructure. Besides, in standard IRMS devices it is not possible to analyze $\delta^{18}O$ and $\delta^{2}H$ simultaneously, which doubles the time and cost required for analyses. Recently, commercial alternatives based on high precision infrared laser spectroscopy have emerged, providing cheap, fast and accurate measures of both $\delta^{18}O$ and $\delta^{2}H$ in a unique measurement (Aggarwal et al. 2006; Lis et al. 2008).

Other techniques related with the measurement of water use deals with the infrared thermography or thermal imaging (Fig. 13.10). In spite of the high cost of thermal cameras (beyond USD 6000) and the fact the thermal imaging has not yet reached fully maturity (Blum 2011) even when is a promising alternative in maize (Romano et al. 2011; Masuka et al. 2012) this option may be basically more suited for maize than infra-red thermometry. Thus, for large leaf species, such as maize, there is a basic problem in recording a truly representative ground-level canopy temperature where dark (cooler) spaces between the large leaves might bias the reading (Jones et al. 2009). Because the comparative low cost of infra-red thermometers (USD 150 and up) and the relatively easy protocol of use, this technique has been also applied to measure temperature of individual leaves (Araus et al. 2011); however, in this case, a large number of leaves need to be measured per plot, which implies time and the interacting problem of a daily pattern of radiation and temperature.

13.6.6 Novel Tools for Disease and Insect-Pest Phenotyping

Accurate, precise, and reproducible assessment of biotic stress symptoms is needed for estimating disease incidence and severity, and is of particular importance for monitoring epidemics, resistance screening, and assessing the effect of

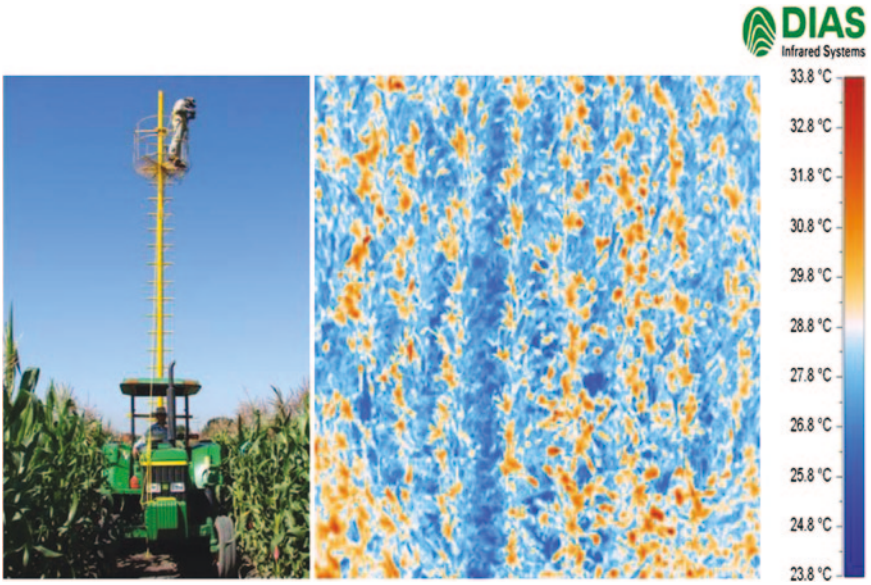

Fig. 13.10 Thermal images of tropical maize on rows taken at the CIMMYT station of Tlaltizapan (Mexico). Emitted long-wave infrared radiation is processed into color digital image display of the different temperatures in the target area

disease on yield. Disease severity has often been quantified by visual estimation aided by illustrated diagrammatic scales, such as the drawings of various foliar diseases by James (1971). However, visual estimates may be subjective and of limited accuracy, as it depends on the experience of the evaluators. Several technologies have been developed aimed at improving the precision, accuracy and reproducibility of assessing diseases and insect damage (Nilsson 1995). Disease severity and insect damage have been quantified using leaf area meters (Sutton 1985), percent sunlight reflectance and infrared radiation (Nutter et al. 1993). To improve the throughput, the SPAD-502 chlorophyll meter was used as a non-destructive means for measuring chlorophyll content in several plant species, including assessment of damage to the leaf tissue caused by the Russian wheat aphid (Belefant-Miller et al. 1994) and Sorghum greenbug (Deol et al. 1997), and also for following the foliar disease symptoms (Scholes and Rolfe 2009).

The use of digital imaging systems offers several advantages including availability of a nondestructive and noninvasive method that can capture, process, and analyze information from images (Richardson et al. 2001; D′ıaz-Lago et al. 2003; Karcher and Richardson 2003; Masuka et al. 2012). Computer automated digital image analysis appears to provide a consistent, unbiased, precise and highly reproducible way to assess insect damage, and disease intensity and severity (Mirik et al. 2006). An alternative method for high throughput phenotyping of disease and insect damage in cereal crops is to measure the reflectance from the vegetation surface. Reflectance data has also been found to provide accurate and precise quantification of disease

or insect-pest damage of plants (Nilsson and Johnsson 1996; Riedell and Blackmer 1999). Remote sensing using spectral reflectance has the potential to improve the speed and accuracy of biotic stress assessment (González-Pérez et al. 2011; Nilsson 1995; Patil and Kumar 2011). This approach, coupled with digital image analysis, was successfully used to determine damage by greenbugs in wheat with repeatable accuracy and precision (Mirik et al. 2006). Aerial photography and photogrammetry using infrared film or color filter combinations to enhance the differentiation between healthy and diseased tissue, represent a separate approach to insect-pest and disease assessment (González-Pérez et al. 2011; Patil and Kumar 2011).

The establishment of fast and non-destructive methods for precise evaluation of quality and safety of raw grains is another important demand nowadays, to avoid toxic contamination of food and feed. Several analytical methods including Thin-layer chromatography (TLC), Gas chromatography (GC), High performance liquid chromatography (HPLC), methodologies based on immunosorbent assay (ELISA), immuno-affinity and fluorescence, have been tested for their feasibility and applicability to detect and quantify mycotoxins in cereals (Zheng et al. 2005). All these methods are generally difficult, time-consuming, expensive and not suitable for real time control measures. These methodologies are also not amenable to large-scale breeding programs or to developing country conditions, where these technologies are urgently needed. A high throughput methodology based on Near-Infrared Spectroscopy (NIRS) is being developed for detecting and quantifying mycotoxins in cereals. Some mycotoxins are potent carcinogens that have been associated with high rates of liver cancer, and human death, especially in Asia and Africa. Contamination of cereal grains by mycotoxigenic fungi is a perennial problem (e.g., for maize in some African countries) that negatively impacts grain quality and food safety. NIRS is an excellent candidate for a rapid and low-cost method for the detection of various mycotoxins, such as aflatoxins, fumonisins and deoxnivelanol (DON) in cereals (Berardo et al. 2005; Fernandez-Ibanez et al. 2009; Bolduan et al. 2009). Several institutions, including CIMMYT, are developing calibration curves as a way of adapting this tool for high throughput detection and quantification of mycotoxins and mycotoxin causing fungi in grain commodities.

13.6.7 Nutritional Trait Phenotyping and Metabolite Profiling

Grain quality can be defined as the net sum of the underlying genetics determining the functional components of the grain, combined with the influence of the environment in which the grain was produced (O'Brien and Cracknell 2009). To properly asses and utilize the quality traits in crop plants, approved analytical or testing methods are required, coupled sometimes with a range of processing technologies that can convert the grains into food, feed and industrially useful end-products.

Cereals contain a wide range of macronutrients (carbohydrates, lipids, protein, and fiber) and micronutrients (vitamins, minerals). Some of the micronutrients

are subject of biofortification programs worldwide (Bouis and Welch 2010). For example, maize grains can be source of minerals like Fe and Zn, provitamins A, vitamin E, essential aminoacids (lysine, tryptohan, methionine), anthocyanins, etc. (Nuss and Tanumihardjo 2010). A major step forward in the application of testing methods with special implications for early generation testing of such traits in the breeding programs is the use of Near-Infrared Reflectance Spectroscopy (NIRS). NIRS is a chemometric technique that combines spectroscopy and mathematics to rapidly produce indirect, quantitative estimates of concentration of OH-, NH-, CH- or SH-containing compounds. Compared to wet chemistry procedures, NIRS requires simple sample preparation methods and is rapid, non-destructive, and relatively inexpensive, facilitating analysis of several traits simultaneously (Montes et al. 2007). Spectral data are correlated with biochemical components obtained by standard methods. However, NIRS is an indirect method that requires development and validation of calibrations by analysis of a large number of samples covering the range of variability for each trait and with more or less uniform distribution between extreme values (Cabrera-Bosquet et al. 2011).

NIR is being used to asses a wide range of compounds with different degrees of precision. Protein, starch, oil and grain moisture are commonly measured by NIR. However, its potential is also clearly demonstrated to predict more complex quality traits like total gluten content, mix time, loaf volume, flour particle size, fiber, digestibility, dry matter, energy-related traits not only in laboratory infrared spectroscopy but also with field-based NIR spectroscopy (Montes et al. 2007; Dowell et al. 2006). NIR transmittance and reflectance spectroscopy have been also used for classifying seeds for particular attributes like color, insect infestation, hardness, starch composition or vitreousness of wheat grain and starch, hardness and mycotoxin levels in maize (Spielbauer et al. 2009). Tallada et al. (2009) and Spielbauer et al. (2009) have recently developed reliable predictions for percentage starch, protein and oil levels in single maize kernels using a glass tube NIR instrument. This tool integrates seed weight measurements with spectral acquisition, overcoming technical challenges for single maize kernel analysis and demonstrating its potential as an effective seed phenomics technology.

Microplate-based colorimetric assays also have high-throughput potential for analysis of amylose, lysine, tryptophan, starch, total anthocyanins, among other nutrients in maize (Nurit et al. 2009; Galicia et al. 2010). Phenotyping for carotenoids, amino acid, anthocyanin profiles and mycotoxin quantification in maize can be achieved by High Performance Liquid Chromatography (HPLC). Although HPLC is very precise and allows identification of specific compounds, it is costly and time consuming to support analysis of large number of samples of a breeding program. Ultra-Performance Liquid Chromatography (UPLC) is a promising alternative to HPLC, due to the lower cost of reagents, and the throughput can be increased to more than 6 times that of HPLC for carotenoid profiles.

Breeding programs in cereals for quality traits normally focus on the selection of appropriate parental lines or starting material, and deploy testing methods in early generations. Early breeding generations implies a large number of samples to be screened; thus, a few key quality descriptors are monitored over a large array

of germplasm or breeding materials; therefore, the phenotyping methods must be high-throughput and low-cost (Nurit et al. 2009). However, in some cases, those methods explain end-use quality only partially, as it is also the case with the DNA markers available for quality traits (Pena et al. 2002). Therefore, during advanced breeding stages and variety release process, the type and extent of quality testing could expand, and consequently, the degree of sophistication and predictiveness of end-use functionality (O'Brien and Cracknell, 2009).

Metabolomics approaches enable the assessment of the levels of a broad range of analytes from different chemical classes (polar to lipohilic). Metabolomics have been documented to have great value in phenotyping, diagnostic analyses in plants as well as in bioprospecting of novel pharmaceuticals (Fernie and Schauer 2009; Hall et al. 2008). Applications of metabolomic technologies in both applied and fundamental science strategies are growing rapidly. There is now increase use of Liquid Chromatography-Mass Spectroscopy (LC–MS), Gas Chromatography-Mass Spectroscopy (GC–MS), Capillary Electrophoresis-Mass Spectroscopy (CE-MS) and Nuclear Magnetic Resonance (NMR) in phenotyping for quality and nutritional traits of crop plants. These include most of the primary and secondary plant metabolites such as soluble sugars, sugar phosphates, complex carbohydrates, amino acids, organic acids, alcohols, lipids, sterols, phenylpropanoids, lignings and triterpene saponins (Hall et al. 2011). However, effective application of metabolomics is possible only if the analytical methods are combined with appropriate statistical tools, databases and software that allow not only proper analysis but also integration of metabolite data with genomic data (Dixon et al. 2006).

In maize, several studies have been conducted with regard to metabolite changes during growth and development (Seebauer et al. 2004), including the influence of environment (Harrigan et al. 2007) and the impact of genetic background and growing season (Rohling et al. 2009). Food composition databases are now being established that can serve as valuable repositories of information on the natural variation in nutrients available in crop germplasm, nutrition education, nutritional labeling, dietary recommendations and community nutrition (Toledo and Burlingame 2006; Burlingame et al. 2009). Technological advances in metabolite profiling, including: (a) high-throughput platforms and tools that enable analysis of hundreds of important metabolites and significantly decrease the costs of such analysis, and (b) development of better models that describe the links both within metabolism itself and between metabolism and yield-associated traits, offer great potential to implement metabolite-assisted breeding for crop improvement (Shauer and Fernie 2006).

13.7 Organization and Analysis of High-Throughput Phenotypic Data

The organization and analysis of large, heterogeneous phenotypic datasets could be a major challenge. The present efforts with regard to organization and analysis of phenotypic data generated from high-throughput phenomics studies utilize

both manual and computational methods, with their own distinctive advantages and constraints. The manual data integration and analytical methods may provide more accurate gene–phenotype associations, but are time- and labour-consuming. In contrast, automated computational data integration and analysis will be cost- and time-effective, but may lack accuracy unless properly optimized.

Vankadavath et al. (2009) described the utility of a software titled "PHENOME" that allows researchers to accumulate, categorize, and manage large volume of phenotypic data (from both lab and the field). The "PHENOME" application uses a Personal Digital Assistant (PDA) with built-in barcode scanner in concert with customized database specific for handling large populations, and aids collection and analysis of data obtained in large-scale mutagenesis, assessing quantitative trait loci (QTLs), raising mapping population, sampling of several individuals in one or more ecological niches, etc.

13.8 Summary and Outlook

It is impossible to imagine how economically viable crop production could be sustained and continuously improved without the development of climate change resilient cultivars (Blum 2011; Cairns et al. 2012). A large part of this progress has to happen through genetic improvement and deployment of high-yielding and stress resilient cultivars, in addition to improvements in crop management. While plant breeding has been successful in continuously improving complex traits like drought, future advances in breeding will be only possible by combining genetic analyses, high-throughput genotyping and phenotyping (preferably in the field) together with prediction models (Araus et al. 2008, 2011; Collins et al. 2008; Tardieu and Tuberosa 2010), and molecular breeding. However, the capacity of several institutions both in the developing and developing world for undertaking precision phenotyping, particularly under repeatable and representative levels of stress in the field, is lagging far behind the capacity to generate genomic information. This might limit not only the progress in generating gene–phenotype associations for traits, but also, consequently, accelerating genetic gains and breeding progress.

Field phenotyping of the right traits, using low cost, easy-to-handle tools, should become an integral and key component in the breeding pipeline, particularly for national agricultural research systems and small- and medium-sized seed companies from developing countries, which may not be able to afford establishment and maintenance of expensive high throughout phenotyping platforms. Utilizing technological advances with regard to phenotyping instrumentation must also go hand-in-hand with methods to characterize and control field site variation (for improving repeatability), adopting appropriate experimental designs, selection of right traits, and finally, proper integration of heterogeneous datasets, analysis, and application.

Acknowledgments The support received from the *'Drought Tolerant Maize for Africa'* (DTMA) project (funded by the Bill and Melinda Gates Foundation), and the *'Precision phenotyping for improving drought stress tolerant maize in southern Asia and eastern Africa'* project (funded by BMZ, Germany) for implementing some of the work reported in this article is gratefully acknowledged.

References

Abramoff MD, Magelhaes PJ, Ram SJ (2004) Image processing with ImageJ. Biophotonics Int 11:36-42

Aggarwal PK, Ahmad T, Groening M, Gupta M, Owano T, Baer D (2006) Laser spectroscopic analysis of stable isotopes in natural waters: a low-cost, robust technique for the use of environmental isotopes in hydrological and climate studies. In: American Geophysical Union, fall meeting 2006, Abstract #H51D-0504

Anderson-Cook CM, Alley MM, Roygard JKF, Khosla R, Nobel RB, Doolittle JA (2002) Differentiating soil types using electromagnetic conductivity and crop yield maps. Soil Sci Soc Am J 66:1562–1570

Araus JL, Amaro T, Casadesus J, Asbati A, Nachit MM (1998) Relationships between ash content, carbon isotope discrimination and yield in durum wheat. Aust J Plant Physiol 25:835–842

Araus JL, Sanchez C, Edmeades GO (2011) Phenotyping maize for adaptation to drought. In: Monneveux P, Ribaut J-M (eds.) Drought phenotyping in crops: from theory to practice. Generation Challenge Programme, pp 259–282

Araus JL, Slafer GA, Reynolds MP, Royo C (2002) Plant breeding and water stress in C3 cereals: what to Breed for? Ann Bot 89:925–940

Araus JL, Slafer GA, Royo C, Serret MD (2008) Breeding for yield potential and stress adaptation in cereals. Crit Rev Plant Sci 27:1–36

Araus JL, Cabrera-Bosquet LL, Sánchez C (2010) Is heterosis in maize mediated through better water use? New Phytol 187:392–406

Arsenault JL, Poulcur S, Messier C, Guay R (1995) WinRHIZO, a root-measuring system with a unique overlap correction method. Hortic Sci 30:906

Bänziger M, Edmeades GO, Lafitte HR (1999) Selection for drought tolerance increases maize yields over a range of N levels. Crop Sci 39:1035–1040

Bänziger M, Edmeades GO, Beck D, Bellon M (2000) Breeding for drought and nitrogen stress tolerance in maize: from theory to practice. CIMMYT, Mexico

Bänziger M, Setimela PS, Hodson D, Vivek B (2006) Breeding for improved abiotic stress tolerance in Africa in maize adapted to southern Africa. Agric Water Manag 80:212–214

Bänzinger M, Araus JL (2007) Recent advances in breeding maize for drought and salinity stress tolerance. In: Jenks MA, Hasenawa PM, Jain SM (eds) Advances in molecular-breeding toward drought and salt tolerant crops. Springer, Berlin, pp 587–601

Barbour MM (2007) Stable oxygen isotope composition of plant tissue: a review. Funct Plant Biol 34:83–94

Barnabas B, Jager K, Feher A (2008) The effect of drought and heat stress on reproductive processes in cereal. Plant, Cell Environ 31:11–38

Barnard R, de-Bello F, Gilgen AK, Buchmann N (2006) The δ18O of root crown water best reflects source water δ18O in different types of herbaceous species. Rapid Commun Mass Spectrom 20:3799–3802

Belefant-Miller H, Porter DR, Pierce ML, Mort AJ (1994) An early indicator of resistance in barley to Russian wheat aphid. Plant Physiol 105:1289–1294

Bengough AG, Mullins CE (1990) Mechanical impedance to root growth: a review of experimental techniques and root growth responses. J Soil Sci 41:341–358

Berardo N, Pisacane V, Battiliani P, Scandolara A, Pietro A, Marocco A (2005) Rapid detection of kernel rots and mycotoxins in maize by near-infrearred reflectance spectrosocpy. J Agric Food Chem 53:8128–8134

Blum A (2005) Drought resistance, water-use efficiency, and yield potential—are they compatible, dissonant, or mutually exclusive? Aust J Agric Res 56:1159–1168

Blum A (2009) Effective use of water (EUW) and not water-use efficiency (WUE) is the target of crop yield improvement under drought stress. Field Crop Res 112:119–123

Blum A (2011) Plant breeding for water-limited environments. Springer, New York, p 255

Bolaños J, Edmeades GO (1996) The importance of the anthesis-silking interval in breeding for drought tolerance in tropical maize. Field Crop Res 48:65–80

Bolaños J, Edmeades GO, Martinez L (1993) Eight cycles of selection for drought tolerance in lowland tropical maize. III. Responses in drought-adaptive physiological and morphological traits. Field Crop Res 31:269–286

Bolduan C, Montes JM, Dhillon BS, Mirdita V, Melchinger AE (2009) Determination of mycotoxin Concentration by ELISA and Near-Infrared Spectroscopy in *Fusarium*-inoculated maize. Cereal Res Commun 37:521–529

Bouis HE, Welch RM (2010) Biofortification—a sustainable agricultural strategy for reducing micronutrient malnutrition in the global South. Crop Sci 50:S20–S32

Boursiac Y, Boudet J, Postaire O, Luu DT, Tournaire-Roux C, Maurel C (2008) Stimulus induced downregulation of root water transport involves reactive oxygen species-activated cell signalling and plasma membrane intrinsic protein internalization. Plant J 56:207–218

Bruce WB, Edmeades GO, Barker TC (2002) Molecular and physiological approaches to maize improvement for drought tolerance. J Exp Bot 53:13–25

Burlingame B, Charrondiere R, Mouille B (2009) Food composition is fundamental to the cross-cutting initiative on biodiversity for food and nutrition. J Food Compos Anal 22:361–365

Cabrera-Bosquet L, Molero G, Nogués S, Araus JL (2009a) Water and nitrogen conditions affect the relationships of $\Delta^{13}C$ and $\Delta^{18}O$ with gas exchange and growth in durum wheat. J Exp Bot 60:1633–1644

Cabrera-Bosquet L, Sánchez C, Araus JL (2009b) How yield relates to ash content, $\Delta^{13}C$ and $\Delta^{18}O$ in maize grown under different water regimes. Ann Bot 104:1207–1216

Cabrera-Bosquet L, Sanchez C, Araus JL (2009c) Oxygen isotope enrichment ($\Delta^{18}O$) reflects yield potential and drought resistance in maize. Plant Cell Environ 32:1487–1499

Cabrera-Bosquet L, Sanchez C, Rosales A, Palacios-Rojas N, Araus JL (2011a) Near-infrared reflectance spectroscopy (NIRS) assessment of $^{\delta}18O$ and nitrogen and ash contents for improved yield potential and drought adaptation in maize. J Agric Food Chem 59:467–474

Cabrera-Bosquet L, Sánchez C, Rosales A, Palacios-Rojas N, Araus JL (2011b) NIRS-assessment of $\delta^{18}O$, nitrogen and ash content for improved yield potential and drought adaptation in maize. J Agric Food Chem 59:467–474

Cabrera-Bosquet L, Crossa J, Zitzewitz JV, Serrte MD, Araus JL (2012) High throughput phenotyping and genomic selection: the frontiers of crop breeding converge. J Integr Plant Biol 54:312–320

Cairns JE, Aubebert A, Townend J, Price AH, Mullins CE (2004) Effect of soil mechanical impedance on root growth of two rice varieties under field drought stress. Plant Soil 267:309–318

Cairns JE, Audebert A, Mullins CE, Price AH (2009) Mapping quantitative loci associated with root growth in upland rice (*Oryza sativa* L.) exposed to soil water-deficit in fields with contrasting soil properties. Field Crop Res 114:108–118

Cairns JE, Impa SM, O'Toole JC, Jagadish SVK, Price AH (2011) Influence of the soil physical environment on rice (*Oryza sativa* L.) response to drought stress and its implications for drought research. Field Crop Res 121:303–310

Cairns JE, Sonder K, Zaidi PH, Verhulst N, Mahuku G, Babu R, Nair SK, Das B, Govaerts B, Vinayan MT, Rashid Z, Noor JJ, Devi P, San Vicente F, Prasanna BM (2012) Maize production in a changing climate: Impacts, adaptation and mitigation strategies. Adv Agron 114:1–58

Campos H, Cooper M, Habben JE, Edmeades GO, Schussler JR (2004) Improving drought tolerance in maize: a view from industry. Field Crop Res 90:19–34

Carcova J, Otegui ME (2007) Ovary growth and maize kernel set. Crop Sci 47:1104–1110

Castleberry CW, Crum CW, Krull CF (1984) Genetic yield improvement of US maize cultivars under varying fertility and climatic environments. Crop Sci 24:33–37

Cernusak LA, Winter K, Turner BL (2009) Physiological and isotopic ($\delta^{13}C$ and $\delta^{18}O$) responses of three tropical tree species to water and nutrient availability. Plant Cell Environ 32:1441–1455

Cernusak LA, Winter K, Aranda J, Turner BL, Marshall JD (2007) Transpiration efficiency of a tropical pioneer tree (*Ficus insipida*) in relation to soil fertility. J Exp Bot 58:3549–3566

Chun L, Mi G, Li J, Chen F, Zhang F (2005) Genetic analysis of maize root characteristics in response to low nitrogen stress. Plant Soil 276:369–382

Collins NC, Tardieu F, Tuberosa R (2008) QTL approaches for improving crop performance under abiotic stress conditions: where do we stand? Plant Physiol 147:469–486

Cullis BR, Gleeson AC (1991) Spatial analysis of experimental fields—an extension to two dimension. Biometrics 47:1449–1460

D′ıaz-Lago JE, Stuthman KJ, Leonard DD (2003) Evaluation of components of partial resistance to oat crown rust using digital image analysis. Plant Dis 87:667–674

Deol GS, Reese JC, Gill BS (1997) A rapid, nondestructive technique for assessing chlorophyll loss from Greenbug (Homoptera: Aphididae) feeding damage on sorghum leaves. J Kansa Entomol Soc 70:305–312

Dixon R, Gang D, Charlton A, Fiehn O, Kuiper H, Reynolds T, Tjeerdema R, Jeffery E, German B, Ridley W, Seiber J (2006) Applications of metabolomics in agriculture. J Agric Food Chem 54:8984–8994

Dowell FE, Maghirang F, Xie F, Lookhart GL, Pierce RO, Seabourn B, Bean S, Wilson J, Chung O (2006) Predicting wheat quality characteristics and functionality using near-infrared spectroscopy. Cereal Chem 85:529–536

Draye X, Kim Y, Lobet G, Javaux M (2010) Model-assisted integration of physiological and environmental constraints affecting the dynamic and spatial patterns of root water uptake from soils. J Exp Bot 61:2145–2155

Duvick DN (1997) What is yield. In: Edmeades GO, Bänziger M, Mickelson HR et al. (eds) Developing drought and low-N tolerant maize. Proceedings of a symposium, 25–29 Mar 1996, CIMMYT, El Batan, pp 332–335

Duvick DN (2005) The contribution of breeding to field advances in maize (*Zea mays* L.). Adv Agron 86:83–145

Duvick DN, Smith JCS, Cooper M (2004) Changes in performance, parentage, and genetic diversity of successful corn hybrids, 1930 to 2000. In: Smith CW, Betrán J, Runge ECA (eds) Corn: origin, history, technology and production. Wiley, New Jersey, pp 65–97

Dwivedi SL, Crouch JH, Mackill DJ, Xu Y, Blair MW, Ragot M, Upadhyaya HD, Ortiz R (2010) The molecular characterization of public sector crop breeding progress, problems and prospects. Adv Agron 95:163–318

Eathington SR, Crosbie TM, Edwards MD, Reiter RS, Bull JK (2007) Molecular markers in commercial breeding. Crop Sci 47:154–163

Edmeades GO, Bolaños J, Chapman SC, Lafitte HR, Bänziger M (1999) Selection improves drought tolerance in tropical maize populations. I. Gains in biomass, grain yield, and harvest index. Crop Sci 39:1306–1315

Ehleringer JR, Dawson TE (1992) Water uptake by plants: perspectives from stable isotope composition. Plant Cell Environ 15:1073–1082

Evenson R, Gollin D (2003) Assessing the impact of the green revolution, 1960–2000. Science 300:578–672

Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics, 4th edn. Longman, Essex

Farquhar GD, Cernusak LA, Barnes B (2007) Heavy water fractionation during transpiration. Plant Physiol 143:11–18

Fernandez-Ibanez V, Soldado A, Martinez-Fernandez B, De la Rosa-Delgado B (2009) Application of near infrared spectroscopy for rapid detection of aflatoxin in maize and barley as analytical quality assessment. Food Chem 113:629–634

Fernie AR, Schauer N (2009) Metabolomics-assisted breeding: a viable option for crop improvement? Trends Genet 25:39–48

Ferrio JP, Mateo MA, Bort J, Voltas J, Abdahla O, ArausM JL (2007) Relationships of grain $\delta^{13}C$ and $\delta^{18}O$ with wheat phenology and yield under water-limited conditions. Ann Appl Biol 150:207–215

Ferrio JP, Resco V, Williams DG, Serrano L, Voltas J (2005) Stable isotopes in arid and semi-arid forest systems. Investigación Agraria, Sistemas y Recursos Forestales 14:371–382

Finkle E (2009) With 'Phenomics" plant scientists hope to shift breeding into overdrive. Science 325:380–381

Galicia L, Nurit E, Rosales A, Palacios-Rojas N (2010) Laboratory protocols: maize nutrition quality and plant tissue analysis. CIMMYT, Mexico

González-Pérez JL, Espino-Gudiño MC, Torres-Pacheco I, Guevara-González RG, Herrera-Ruiz G, Rodríguez-Hernández V (2011) Quantification of virus syndrome in chili peppers. Afr J Biotechnol 10:5236–5250

Gowda VRP, Henry A, Yamauchi A, Shashidhar HE, Serraj R (2011) Root biology and genetic improvement for drought avoidance in rice. Field Crop Res 122:1–13

Granier C, Aguirrezabal L, Chenu K, Cookson SJ, Dauzat M, Hamard P, Thioux JJ, Rolland G, Bouchier-Combaud S, Lebaudy A, Muller B, Simonneau T, Tardieu F (2005) PHENOPSIS, an automated platform for reproducible phenotyping of plant responses to soil water deficit in Arabidopsis thaliana permitted the identification of an accession with low sensitivity to soil water deficit. New Phytol 169:623–635

Hachez C, Heinen RB, Draye X, Chaumont F (2008) The expression pattern of plasma membrane aquaporins in maize highlights their role in hydraulic regulation. Plant Mol Biol 68:337–353

Hall R, Brouwer I, Fitzgerald M (2008) Plant metabolomics and its potential application for human nutrition. Physiol Plant 132:162–175

Hall R, Fernie AR, Keurentjes J (2011) Genetics, genomics and metabolomics. Annu Plant Rev 43 (Biology of Plant Metabolomics)

Hammer G, Dong Z, McLean G, Doherty A, Messina C, Schussler J, Zinselmeier C, Paszkiewicz S, Cooper M (2009) Can changes in canopy and/or root system architecture explain historical maize yield trends in the U.S. corn belt? Crop Sci 49:299–312

Harrigan GG, Stork LG, Riordan S, Ridley WP, MacIsaac S, Halls S (2007) Metabolite analysis of grain from maize hybrids grown in the United States under drought and water conditions during the 2002 field season. J Agric Food Chem 55:6169–6176

Harris K, Subudhi PK, Borrell A, Jordan D, Rosenow D, Nguyen H, Klein P, Klein R, Mullet J (2007) Sorghum stay-green QTL individually reduce post-flowering drought-induced leaf senescence. J Exp Bot 58:327–338

Hochholdinger F, Woll K, Sauer M, Dembinsky D (2004) Genetic dissection of root formation in maize (*Zea mays*) reveals root-type specific developmental programmes. Ann Bot 93:359–368

Hoecker N, Keller B, Piepho HP, Hochholdinger F (2006) Manifestation of heterosis during early maize (*Zea mays* L.) root development. Theore Appl Genet 112:421–429

Holst J, Grote R, Offermann C, Ferrio JP, Gessler A, Mayer H, Rennenberg H (2010) Water fluxes within beech stands in complex terrain. Int J Biometeorol 54:23–36

Hund A, Ruta N, Liedgens M (2009) Rooting depth and water use efficiency of tropical maize. Plant Soil 318:311–325

Iyer-Pascuzzi AS, Symonova O, Mileyko Y, Hao Y, Belcher H, Harer J, Weitz JS, Benfey PN (2010) Imaging and analysis platform for automatic phenotyping and trait ranking of plant root systems. Plant Physiol 152:1148–1157

James WC (1971) An illustrated series of assessment keys for plant diseases, their preparation and usage. Can Plant Dis Surv 51:39–65

Javaux M, Schröder T, Vanderborght J, Vereecken H (2008) Use of a three-dimensional detailed modeling approach for predicting root water uptake. Vadose Zone J 7:1079–1088

Johnson CK, Mortensen DA, Wienhold DA, Shanahan JF, Doran DW (2003) Site-specific management zones based on soil electrical conductivity in semiarid cropping systems. Agronomy J 95:303–315

Jones HG, Serraj R, Loveys BR, Xiong L, Wheaton A, Price AH (2009) Thermal infrared imaging of crop canopies for the remote diagnosis and quantification of plant responses to water stress in the field. Funct Plant Biol 36:978–989

Karcher DE, Richardson MD (2003) Quantifying turfgrass color using digital image analysis. Crop Sci 43:943–951

Kaspar TC, Colvin TS, Jaynes DB, Karlen DL, James DE, Meek DW (2003) Relationship between six years of corn yields and terrain attributes. Precision Agric 4:87–101

Kaspar TC, Pulido DJ, Fenton TE, Colvin TS, Karlen DL, Jaynes DB, Meek DW (2004) Relationship of corn and soybean yield to soil and terrain properties. Agron J 96:700–709

Lafitte HR, Edmeades GO (1994) Improvement for tolerance to low soil nitrogen in tropical maize. 1. Selection criteria. Field Crop Res 39:1–14

Li B, Tian X-L, Wang G-W, Pan F, Li Z-H (2008) Heterosis of root growth in maize (*Zea mays* L.) seedlings under water stress. Acta Agron Sinica 34:662–668

Lis GP, Wassenaar LI, Hendry MJ (2008) High-precision laser spectroscopy D/H and $^{18}O/^{16}O$ measurements of microliter natural water samples. Anal Chem 80:287–293

Liu X-F, Zhang S-Q, Yang X-Q, Shan L (2009) Heterosis of water uptake ability by roots of maize at cell level. Acta Agron Sinica 35:1546–1551

Lopes MS, Araus JL, Van Heerden PDR, Foyer CH (2011) Enhancing drought tolerance in C4 crops. J Exp Bot 62:3135–3153

Lu Y, Hao Z, Xie C, Crossa J, Araus JL, Gao S, Vivek BS, Magorokosho C, Mugo S, Makumbi D, Taba S, Pan G, Li X, Rong T, Zhang S, Xua Y (2011) Large-scale screening for maize drought resistance using multiple selection criteria evaluated under water-stressed and well-watered environments. Field Crop Res 124:37–45

Marshall TJ, Holmes JW, Rose CW (1996) Soil physics. Cambridge University Press, UK

Masle J, Farquhar GD, Wong SC (1992) Transpiration ratio and plant mineral content are related among genotypes of a range of species. Aust J Plant Physiol 19:709–721

Masuka B, Araus JL, Das B, Sonder K, Cairns JL (2012) Phenotyping for abiotic stress tolerance in maize. J Integr Plant Biol 54:238–249

Mayland HF, Johnson DA, Asay KH, Read JJ (1993) Ash, carbon isotope discrimination and silicon as estimators of transpiration efficiency in crested wheatgrass. Aust J Plant Physiol 20:361–369

McLaughlin JE, Boyer JS (2004) Glucose localization in maize ovaries when kernel number decreases at low water potential and sucrose is fed to the stems. Ann Bot 94:75–86

Mirik M, Michels GJ Jr, Kassymzhanova-Mirik S, Elliott NC, Catana V, Jones DB, Bowling R (2006) Using digital image analysis and spectral reflectance data to quantify damage by greenbug (Hemitera: Aphididae) in winter wheat. Comput Electron Agric 51:86–98

Monneveux P, Sánchez C, Tiessen A (2008) Future progress in drought tolerance in maize needs new secondary traits and cross combinations. J Agric Sci 146:1–14

Montes J, Melchinger A, Reif J (2007) Novel throughput phenotyping platforms in plant genetic studies. Trends Plant Sci 12:433–436

Montes JM, Technow F, Dhillon BS, Mauch F, Melchinger AE (2011) High-throughput non-destructive biomass determination during early plant development in maize under field conditions. Field Crop Res 121:268–273

Nilsson H-E (1995) Remote sensing and image analysis in plant pathology. Annu Rev Phytopathol 15:489–527

Nilsson H-E, Johnsson L (1996) Hand-held radiometry of barley infected by barley stripe disease in a field experiment. J Plant Dis Prot 103:517–526

Nurit E, Tiessen A, Pixley K, Palacios-Rojas N (2009) A reliable and inexpensive colorimetric method for determining protein-bound tryptophan in maize kernels. J Agric Food Chem 57:7233–7238

Nuss ET, Tanumihardjo SA (2010) Maize: a paramount staple crop in the context of global nutrition. Compr Rev Food Sci Food Saf 9:417–436

Nutter FW Jr, Gleason ML, Jenco JH, Christians NC (1993) Assessing the accuracy, inter-rater repeatability and inter-rater reliability of disease assessment systems. Phytopathology 83:806–812

O'Brien L, Cracknell R (2009) The application of testing methods in cereals breeding programs. In: Cauvain S, Young L (eds) The ICC handbook of cereals, flour, dough and product testing. DE Stech Publications, Pennsylvania, pp 9–32

Papadakis JS (1937) Méthode statistique pour des expériences sur champ. Bulletin Institut d'Amélioration des Plantes á Salonique 23:13–21

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H et al (2009) The *Sorghum bicolor* genome and the diversification of grasses. Nature 457:551–556

Patil JK, Kumar R (2011) Advances in image processing for detection of plant diseases. J Adv Bioinf Appl Res 2:135–141

Patz N (2009) Effect of roots of different tree species on preferential flowpaths. Master's Thesis, University of Freiburg

Pena RJ, Trethowan R, Pfeiffer W, Ginkel M (2002) Quality (end-use) improvement in wheat. Compositional, genetic and environmental factors. In: Basra AS, Randhawa LS (eds) Quality improvement in field crops. Haworth Press, New York, pp 1–37

Phillips RL (2009) Mobilizing science to break yield barriers. Crop Sci 50:99–108

Prasanna BM, Chaikam V, Mahuku G (eds) (2012) Doubled haploid technology in maize breeding: theory and practice. CIMMYT, Mexico D.F., 50 pages

Prigge V, Sanchez C, Dhillon BS, Schipprack W, Araus JL, Bänziger M, Melchinger AE (2011) Doubled haploids in tropical maize: 1. Effects of inducers and source germplasm on in vivo haploid induction rate. Crop Sci 51:1498–1506

Retzlaff WA, Blaisdell GK, Topa MA (2001) Seasonal changes in water source of four families of loblolly pine (*Pinus taeda* L.). Trees Struct Funct 15:154–162

Ribaut JM, de Vicente MC, Delannay X (2010) Molecular breeding in developing countries: challenges and perspectives. Curr Opin Plant Biol 13:1–6

Richardson MD, Karcher DE, Purcell LC (2001) Quantifying turfgrass cover using digital image analysis. Crop Sci 41:1884–1888

Riedell WE, Blackmer TM (1999) Leaf reflectance spectra of cereal aphid-damaged wheat. Crop Sci 39:1835–1840

Röber FK, Gordillo GA, Geiger HH (2005) *In vivo* haploid induction in maize—performance of new inducers and significance of doubled haploid lines in hybrid breeding. Maydica 50:275–283

Rohling R, Eder J, Engel K (2009) Metabolite profiling of maize grain: differentiation due to genetics and environment. Metabolomis 5:459–477

Romano G, Zia S, Spreer W, Sanchez S, Cairns J, Araus JL, Müller J (2011) Use of thermography for screening genotypic water stress adaptation in tropical maize. Comput Electron Agric 79:61–74

Saab IN, Sharp RE (1989) Non-hydraulic signals from maize roots in drying soil: inhibition of leaf elongation but not stomatal conductance. Planta 179:466–474

Samson BK, Sinclair TR (1994) Soil core and minirhizotron comparison for the determination of root length density. Plant Soil 161:225–232

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W et al (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178–183

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F et al (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112–1115

Scholes JD, Rolfe SA (2009) Chlorophyll fluorescence imaging as tool for understanding the impact of fungal diseases on plant performance; a phenomics perspective. Funct Plant Biol 36:880–892

Seebauer JR, Moose SP, Fabbri BJ, Crossland LD, Below FE (2004) Amino acid metabolism in maize earshoots. Implications for assimilate preconditioning and nitrogen signaling. Plant Physiol 136:4326–4334

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 Secur 3:307–327

Shauer N, Fernie AR (2006) Plant metabolomics: towards biological function and mechanism. Trends Plant Sci 11:508–516

Slafer GA, Araus JL (2007) Physiological traits for improving wheat yield under a wide range of conditions. In: Spiertz JHJ, Struik PC, van Laar HH (eds) Scale and complexity in plant systems research: gene-plant-crop relations. Springer, Dordrecht, pp 145–154

Spielbauer G, Armstrong P, Baier J, Allen W, Richardson K, Shen B, Settles M (2009) High-throughput near-infrered reflectance spectroscopy for predicting quantitative and qualitative composition phenotypes of individual maize kernels. Cereal Chem 86:556–564

Sudduth KA, Drummond ST, Birrell SJ, Kitchen NR (1997) Spatial modeling of crop yields using soil and topographic data. In: Stafford JV (ed) Precision agriculture. Proceedings of the 1st European conference on precision agriculture, BIOS Scientific Publishers, Oxford, pp 439–447

Sutton JC (1985) Effectiveness of fungicides for managing foliar diseases and promoting yields of Ontario winter wheat. Phytoprotection 66:141–152

Tallada JG, Palacios-Rojas N, Armstrong PR (2009) Prediction of maize seed attributes using a rapid single kernel near infrared instrument. J Cereal Sci 50:381–387

Tambussi EA, Bort J, Araus JL (2007) Water use efficiency in C3 cereals under Mediterranean conditions: a review of physiological aspects. Ann Appl Biol 150:307–321

Tambussi EA, Nogués S, Ferrio JP, Voltas J, Araus JL (2005) Does a higher yield potential improve barley performance under Mediterranean conditions? A case study. Field Crop Res 91:149–160

Tanner W, Beevers H (1990) Does transpiration have an essential function in long distance ion transport in plants? Plant, Cell Environ 13:745–750

Tardieu F (2006) Leaf growth under water-limited conditions. In: Ribaut JM (ed) Drought adaptation in cereals. The Harworth Press, New York, pp 145–169

Tardieu F, Bruckler L, Lafolie F (1992) Root clumping may affect the root water potential and the resistance to soil-root water transport. Plant Soil 140:291–301

Tardieu F, Tuberosa R (2010) Dissection and modelling of abiotic stress tolerance in plants. Curr Opin Plant Biol 13:206–212

Toledo A, Burlingame B (2006) Biodiversity and nutrition: a common path toward global food security and sustainable development. J Food Comp Anal 19:477–483

Tollenaar M, Lee EA (2002) Yield potential, yield stability and stress tolerance in maize. Field Crop Res 75:161–169

Tollenaar M, Lee EA (2006) Dissection of physiological processes underlying grain yield in maize by examining genetic improvement and heterosis. Maydica 51:399–408

Tollenaar M, Wu J (1999) Yield improvement in temperate maize is attributable to greater water stress tolerance. Crop Sci 39:1597–1604

Tournaire-Roux C, Sutka M, Javot H, Gout E, Gerbeau P, Luu D-T, Bligny R, Maurel C (2003) Cytosolic pH regulates root water transport during anoxic stress through gating of aquaporins. Nature 425:393–397

Trachsel S, Kaeppler SM, Brown KM, Lynch JP (2011) Shovelomics: high throughput phenotyping of maize (*Zea mays* L.) root architecture in the field. Plant Soil 341:75–87

Tuberosa R, Salvi S, Sanguineti MC, Landi P, Maccaferri M, Conti S (2002) Mapping QTLs regulating morphophysiological traits and yield: case studies, shortcomings and perspectives in drought-stressed maize. Ann Bot 89:941–963

Tuberosa R, Salvi S, Sanguineti MC, Maccaferri M, Giuliani S, Landi P (2003) Searching for quantitative loci controlling root traits in maize: a critical appraisal. Plant Soil 255:35–54

Vankadavath RN, Hussain AJ, Bodanapu R, Kharshiing E, Basha PO, Gupta S, Sreelakshmi Y, Sharma R (2009) Computer aided data acquisition tool for high-throughput phenotyping of plant populations. Plant Methods 5:18. doi:10.1186/1746-4811-5-18

Venuprasad R, Sta Cruz MT, Amante M, Magbanua R, Kumar A, Atlin GN (2008) Response to two cycles of divergent selection for grain yield under drought stress in four rice breeding populations. Field Crop Res 107:232–244

Voltas J, Chambel MR, Prada MA, Ferrio JP (2008) Climate-related variability in carbon and oxygen stable isotopes among populations of Aleppo pine grown in common-garden tests. Trees Struct Funct 22:759–769

Welcker C, Boussuge B, Bencivenni C, Ribaut JM, Tardieu F (2007) Are source and sink strengths genetically linked in maize plants subjected to water deficit? A QTL study of the responses of leaf growth and of anthesis-silking interval to water deficit. J Exp Bot 58:339–349

Whitford R, Gilbert M, Langridge P (2010) Biotechnology in agriculture. In: Reynolds MP (ed) Climate change and crop production. CABI series in climate change, vol 1. CABI, UK, pp 219–244

Yazdanbakhsh N, Fisahn J (2009) High throughput phenotyping of root growth dynamics, lateral root formation, root architecture and root hair development enabled by PlaRoM. Funct Plant Biol 36:938–946

Yu J, Hu S, Wang J, Wong GK-S, Li S, Liu B et al. (2002). A draft sequence of the rice genome (*Oryza sativa* L. spp. *indica*). Science 296:79–92

Yue B, Xiong L, Xue W, Xing Y, Luo L, Xu C (2005) Genetic analysis for drought resistance of rice at reproductive stage in field with different types of soil. Theoret Appl Genet 11:1127–1136

Zhang J, Nguyen H, Blum A (1999) Genetic analysis of osmotic adjustment in crop plants. J Exp Bot 50:291–302

Zheng Z, Humphrey C, King R, Richard J (2005) A review of rapid methods for the analysis of mycotoxins. Mycopatologia 159:255–263

Zhu J, Ingram PA, Benfey PN, Elich T (2011) From lab to field, new approaches to phenotyping root system architecture. Curr Opin Plant Biol 14:310–317

Zinselmeier C, Lauer MJ, Boyer JS (1995) Reversing drought-induced losses in grain yield: sucrose maintains embryo growth in maize. Crop Sci 35:1390–1400

Chapter 14
Marker-Assisted Selection in Cereals: Platforms, Strategies and Examples

Yunbi Xu, Chuanxiao Xie, Jianmin Wan, Zhonghu He and Boddupalli M. Prasanna

14.1 Introduction

Cereals are the world's most important sources of food, both for direct human consumption and indirectly, as inputs to livestock production. Millions of farmers and consumers in both the developed and the developing world depend on cereals as their preferred staple food. The future of cereal production, affects not only the global food security, but also the livelihoods of several million small farmers worldwide.

Great strides have been made over the past several decades for the development of a large number of improved cereal varieties adapted to different agro-ecologies and meeting the diverse demands of the stakeholders, especially through conventional breeding. However, rising populations, increasing biotic and abiotic stresses, widespread malnutrition, and sharply depleting natural resources (especially water for agricultural purposes), warrant relentless efforts from the scientific community in developing and deploying improved cereal varieties with higher yield potential, input use efficiency, stress resilience and nutritional quality. This, in turn, requires introduction of time- and cost-effective, cutting-edge technologies, including molecular marker-assisted breeding coupled with high-throughput and precision phenotyping.

Mining of novel genetic variation and rapid improvement of genetic gains in breeding programs are among the strategies that can be used to lift the yield

Y. Xu (✉) · Z. He
Institute of Crop Sciences/International Maize and Wheat Improvement Center (CIMMYT), Chinese Academy of Agricultural Sciences, 12 South Zhongguancun Street, Beijing 100081, China
e-mail: y.xu@cgiar.org

C. Xie · J. Wan
Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, 12 South Zhongguancun Street, Beijing 100081, China

B. M. Prasanna
International Maize and Wheat Improvement Center (CIMMYT), ICRAF House, United Nations Avenue, Nairobi, Gigiri, Kenya

P. K. Gupta and R. K. Varshney (eds.), *Cereal Genomics II*,
DOI: 10.1007/978-94-007-6401-9_14,

ceiling. At the same time, realization of such as yield potential in farmer's fields through combination of selected varieties and crop management is also critical for enhancing yield stability and environmental adaptability. Marker-assisted selection, popularly referred to as MAS, is one of major approaches in molecular breeding, aided by advances in molecular biology, genomics, and statistics, that could form the core for genetic enhancement of crop plants, including cereals.

In this chapter, we discuss the strategies of MAS for trait improvement in cereals, including marker-assisted backcrossing (MABC) to transfer favorable alleles at a few loci from one genetic background to another, marker-assisted recurrent selection (MARS) to accumulate alleles from different sources, and marker-assisted genome-wide selection or genomic selection (GS) that uses genome-wide markers along with phenotypic data to predict the performance of the progeny for rapid-cycling. Current status and future prospects of MAS in cereals is provided.

14.2 Marker-Assisted Breeding Platform: Key Components

MAS offers great potential for increasing the genetic gain per crop cycle, by stacking favorable alleles at target loci and reducing the number of selection cycles. In the last decade, the private sector, especially the multinational seed companies, has benefitted immensely from integrating MAS in breeding strategies (Eathington et al. 2007). In contrast, efficient adoption of MAS in the cereal breeding programs in the public sector or in the small- and medium-sized seed companies in the developing world is still limited or hardly evident. Major bottlenecks for this include shortage of well-trained personnel, inadequate high-throughput capacity, poor phenotyping infrastructure, lack of information systems or adapted analysis tools, or simply put, resource-limitations at various levels (Xu and Crouch 2008; Ribaut et al. 2010; Delannay et al. 2012; Xu et al. 2012c). Shortage of validated functional markers (diagnostic markers) for important traits and poor linkage between molecular marker program and conventional breeding program are also the major reasons for slow adaptation of marker-assisted introgression, especially in China (He et al. 2011).

The emerging virtual platforms (Fig. 14.1) aided by the information and communication technology revolution will help to overcome some of these limitations by providing breeders with better access to genomic resources, advanced laboratory services, and robust analytical and data management tools. Apart from some advanced national agricultural research systems (NARS), the implementation of large-scale molecular breeding programs in developing countries will take time. However, the exponential development of genomic resources, the ever-decreasing cost of marker technologies and the **emergence of platforms for** accessing molecular breeding tools and support services are all grounds to predict that molecular breeding will have a significant impact on crop breeding even in the public sector institutions of the developing countries. The sequencing of major cereals will also greatly enhance the development of functional markers. Those predictions are supported by some preliminary successful examples summarized by Delannay et al. (2012).

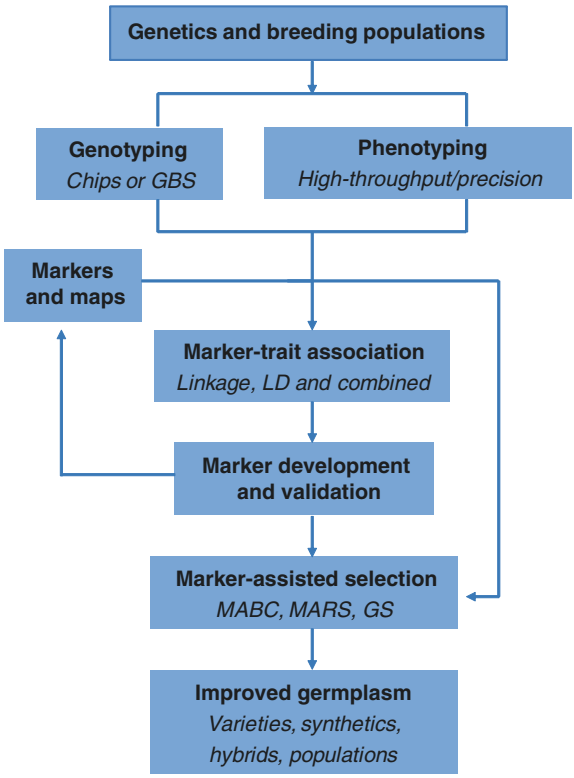

Fig. 14.1 Components of a marker-assisted breeding platform

14.2.1 Trait Genetics and Breeding Populations

Many types of populations have been used in genetic mapping, and identification/development of molecular markers for a range of traits in cereals. The most frequently used ones are biparental mapping populations, with the two parental lines either true-breeding (homozygous) or heterozygous. Such populations include segregating populations such as $F_{2:3}$ and backcross populations, whose genetic constitution will change with selfing, as well as "immortal" populations, such as recombinant inbred lines (RILs), doubled haploids (DHs), backcross introgression lines (BILs) or near-isogenic lines (NILs), whose genetic constitution will not change with selfing. NILs can be produced by various approaches (Xu 2010), including continued backcrossing or mutation. In addition, single chromosome segment substitution lines (CSSLs) can be produced for different chromosomes. A large number of mutations and BILs can be produced so that a library can be constructed to cover every gene in the genome each represented by a line different at that gene from the donor or wild type.

Some multi-parent populations are also increasingly used derived from three-way and four-way crosses or from different genetic mating schemes such as diallel and NC designs by using multiple crosses individually or combined. Recently,

two types of populations have been proposed and used for genetics and breeding, especially in cereals. The first is the nested association mapping (NAM) population which can be defined as a series of RIL populations developed by crossing of one common parent to a set of others. In maize, a NAM population was developed using 25 maize inbreds (testers) crossed to the same inbred (founder) to develop around 250 lines for each of 25 RIL populations (Yu et al. 2008). The second is the multiparent advanced generation intercross (MAGIC) population (Cavanagh et al. 2008), which is produced by starting with single crosses followed by continuous crossing to develop one terminal population that contain genes segregating from different parental lines. The segregating unit can be either homozygous as fixed RILs or heterozygous (segregating within each unit).

Mapping populations, including RILs, DHs and NILs, have their merits and demerits in terms of generation, maintenance, and utilization (Xu 2010). In some cases, joint use of multiple populations may be required for a target trait, particularly for mining multiple favorable alleles. Phenotyping under well-controlled environments, working with populations of large size, and using high-density molecular markers have all contributed to what would have been impossible years ago with the same type of population.

Using a panel of germplasm accessions collected and maintained in worldwide germplasm banks as "natural population" for linkage disequilibrium (LD) or association mapping (Flint-Garcia et al. 2003), is another important development in crop plants. Two important issues, population structure and extreme diversity within the population, should be taken into account when such an association mapping panel is used for genetic mapping. Extreme differences for some morphological traits, particularly plant height and flowering time, will significantly affect the accuracy of phenotyping, particularly when the same population is used for genetic mapping of different traits. In some cases, subpopulations should be formed to minimize the variation in non-target traits that could have significant effects on target trait phenotyping.

14.2.2 Markers and Maps

Molecular markers and their positions on the genome are two basic elements for designing an appropriate MAS strategy (Fig. 14.1). Evolution of molecular markers from Southern-blot based RFLPs (Restriction Fragment Length Polymorphisms) to PCR-based SSRs (Simple Sequence Repeats) is the first big step in MAS, which contributed to significant reduction of the cost and time required for genotyping. Diversity Arrays Technology (DArT) with much less cost (Jaccoud et al. 2001) has been successfully developed in wheat (Akbari et al. 2006), barley (Wenzl et al. 2006) and sorghum (Mace et al. 2008, 2009). The recent shift from SSRs to SNPs (Single Nucleotide Polymorphisms) provides further opportunities for automation, genome coverage and functional allele discovery.

SNPs are the most common polymorphisms among individuals of any species. Resequencing projects have led to the discovery of thousands to millions of SNPs

in various crop plants, enabling construction of high density molecular maps in many crops, including cereals. It has also been recognized that it may be desirable to identify and utilize favorable haplotypes from a set of SNPs rather than individual SNPs within a specific genomic region for genetic diversity analysis or for association mapping of a target trait. "Haplotype" is usually defined as a combination of multiple markers that come from a specific genomic region or a single chromosome. However, the concept has been also used for markers from different chromosomes that might be associated with the same target trait. The statistic association in a small genomic region and identification of multiple alleles within a haplotype block can be used to unambiguously to identify all polymorphic sites in the region. Resequencing of multiple genomes can be used to construct a haplotype map (HapMap) to cover the whole genome. As the thirst example for all organisms, International HapMap Consortium developed three versions of human HapMap. The latest version, HapMap3, contains about 1.6 million common SNPs genotyped in 1,184 individuals from 11 global populations (International HapMap 3 Consortium 2010). The first plant haplotype map was developed in the model plant species *Arabidopsis thaliana*, providing the genome-wide pattern of LD in a sample of 19 accessions of this species using 341,602 non-singleton SNPs, with an Affymetrix genotyping array containing 250,000 SNPs (Kim et al. 2007). Among crops, a HapMap has been developed in maize. The first HapMap consists of 3.3 million SNPs discovered using 27 diverse maize inbred lines, with a frequency of 1 SNP every 44 bp (Gore et al. 2009). Through an international collaboration, over 55 million SNPs were discovered recently, which have been used to develop the second generation of HapMap in maize (Chia et al. 2012). In rice, next-generation sequencing at 1 × coverage across 517 rice varieties led to identification of over 3.6 million SNPs that were used for the construction of a haplotype map (Huang et al. 2010).

Recently large-scale resequencing activities in maize revealed that more reference genomes, at least one tropical inbred line, one landrace and one wild relative in maize should be sequenced *de novo*, to have a HapMap developed for a better representation of maize genetic diversity. A higher quality reference genome sequence is also required in rice, particularly for indica rice, although there is one *de novo* sequence with good quality available for japonica rice. On the other hand, resequencing rice segregating populations has indicated that at least one large segregating population derived from a cross between diverse germplasm should be resequenced deeply for each crop to utilize the linkage information to fill gaps existing in the reference genomes and understand better the genome-wide recombination and segregation.

14.2.3 Genotyping Platforms

As genotyping system has evolved from gels to chips and sequencing, genotyping throughput has increased from singles to millions of markers per assay (or single to thousands of DNA samples per marker), while the cost per data point has decreased from several US dollars to 1/1,000 cent or less. SNP chips have been developed for

a number of cereal crops: rice (McCouch et al. 2010; Yamamoto et al. 2010), maize (Yan et al. 2010a), barley (Close et al. 2009) and durum wheat (Trebbi et al. 2011). In maize, developing chip-based genotyping through Cornell-CIMMYT collaboration brought up three Illumina 1536-SNP chips (Yan et al. 2009, 2010a), which were soon replaced by Illumina MaizeSNP50 Beadchip consisting of 56,110 SNPs, 1 SNP/40 kb, covering 19,540 genes with 2 SNPs/gene. These SNPs were functionally tested with over 30 diverse maize lines. A recent study demonstrated that despite the availability of millions of discovered SNPs in maize, only a very small portion of those polymorphisms could be utilized for the development of robust, versatile assays, and has real practical value in MAS (Mammadov et al. 2012). In rice, a SNP discovery with the OryzaSNP project identified approximately 160,000 high quality SNPs that are informative across 20 diverse rice varieties (McNally et al. 2009). Likewise, resequencing of over 100 varieties at 3 ~ 55 × coverage through the Rice SNP Consortium (www.ricesnp.org) provided an even larger SNP discovery pool from which a one million-SNP chip has been developed (McCouch et al. 2010).

In wheat, 768 SNPs were chosen irrespective of their genomic repetitiveness from 2,659 SNPs identified on 1,206 consensus sequences. When these SNPs were assayed on the Illumina Bead Express genotyping system, 275 (35.8 %) SNPs matched the expected genotypes observed in the SNP discovery phase. Of these SNPs, 157 were mapped in one of the two mapping populations used and integrated into a common genetic map. Despite the relatively low genotyping efficiency of the Golden Gate assay, the validated Complexity Reduction of Polymorphic Sequences (CRoPS)-derived SNPs showed valuable features for genomics and breeding applications; these features include a uniform distribution across the wheat genome, a prevailing single-locus codominant nature and a high level of polymorphism (Trebbi et al. 2011).

An alternative approach for large-scale genotyping is genotyping-by-sequencing (GBS). A simple and highly-multiplexed system for constructing reduced representation libraries was developed for the Illumina next-generation sequencing platform (Elshire et al. 2011). Constructing GBS libraries was based on reducing genome complexity with restriction enzymes, which may reach important regions of the genome that are inaccessible to sequence capture approaches. The procedure has been demonstrated with maize (IBM) and barley (Oregon Wolfe Barley) RIL populations where roughly 200,000 and 25,000 sequence tags were mapped, respectively. With this method, species that lack a complete genome sequence can have a reference map developed around the restriction sites, which can be done in the process of sample genotyping. This system is being optimized for reducing missing data points and improved SNP calls. In the foreseeable future, however, choosing of chip or GBS for genotyping will depend on their cost for genotyping and related data management, analysis and delivery systems.

As sequencing becomes increasingly cheap and high-throughput, resequencing has become an alternative for genotyping. Large-scale resequencing has been done in rice (Xu et al. 2012b) and maize (Chia et al. 2012) for diversity, evolutionary and genetic studies. In rice, resequencing has been used to genotype segregating populations (Huang et al. 2009, 2010) that are usually genotyped by SSR markers or SNP chips.

14.2.4 Precision and High-Throughput Phenotyping

Precision and high-throughput phenotyping is critical for genetic analysis using molecular markers (Fig. 14.1), and for time- and cost-effective implementation of MAS in breeding. Precision phenotyping would allow the researcher to obtain detailed measurements of plant characteristics that collectively provide reliable estimates of trait phenotypes. Our ability to characterize phenomes—the full set of phenotypes of an individual—lags much behind our ability to characterize genotypes. Phenomics should be recognized and pursued to enable the development and adoption of high-throughput phenotyping (Houle et al. 2010).

Among ongoing phenomics projects for plants, International Plant Phenomics Network (IPPN) focuses on development and implementation of phenotyping, including high-throughput, robotic, non-invasive imaging across the life cycle of small, short-lived model and crop plants, including analysis of metabolomes, and quantitative phenotyping. The phenomics project for *Arabidopsis* has resulted in an initial GWA (genome-wide association) study on an overlapping set of 191 inbred lines for 107 mostly quantitative phenotypes (Atwell et al. 2010). Development of the strategies for large plants with longer life cycles, such as maize and sorghum, is much more challenging. The basic requirements of an ideal phenomics effort are easy to state but difficult to achieve: genomic information on a large sample of genotypes, which are each exposed to a range of environments; extensive and intensive phenotyping across the full range of spatial and temporal scales; and low cost. None of the pioneering phenome projects comes close to realizing the full phenomic vision, largely because of the costs and limited phenotyping capabilities (Houle et al. 2010).

While high-throughput and precision under controlled experimental conditions could be important for certain objectives, more important from the breeding perspective is the ability to improve the throughput and precision of field-based phenotyping. For field-based phenotyping, it is important to reduce "signal-to-noise" ratio, including selection of research plots with low spatial variability in soil properties, uniform application of inputs with good weed, pest and disease control, use of adequate plot borders, selecting experimental designs that control within-replicate variability, and analyzing data to reduce or remove spatial trends (Xu et al. 2012c). It also depends on utilization of new field-based techniques, such as precision application of nutrients and water and remote sensing to detect secondary traits, and correct selection, calibration and application of instruments, such as neutron probes, radiation sensors, and chlorophyll and photosynthesis meters.

In many cases, phenotyping needs to be done simultaneously in two contrasting environments. The concept of near iso-environments (NIEs) was proposed for dissecting environmental factors when two contrasting environments are involved (Xu 2002, 2010). The first environment imposes much less stress on plants than the second. The effect of the stress environment can be measured using the much-less-stress or normal environment as a control. A relative trait value is then derived from two direct trait values measured in each environment to ascertain the sensitivity of plants to the stress. If different plants have an identical phenotype under the much-less-stress

environment, the direct trait value in the stress environment can be used to measure sensitivity. When both environments impose little stress on plants, however, one should use relative trait values instead. A typical example is the photoperiod sensitivity that can only be measured in NIEs, one under short day and the other under long day. A relative measure for this type of trait (sensitivity) should be: sensitivity = the difference of measures in the NIEs, divided by the measure in one of the NIEs or in the normal environment when the other is stressful. Traits suitable for measurement under NIEs include all abiotic/biotic stresses and plant responses to different environmental factors and agronomic practices.

14.2.5 Marker-Trait Association Analysis

Marker-trait association analysis or trait mapping can be generally defined as identifying genes/alleles/genomic regions that are significantly associated with specific traits. The association can be established in several ways. Two important approaches are linkage analysis using biparental or multi-parental populations and LD analysis using natural populations (Fig. 14.1). The latter has also been called association mapping, a concept that can be used to cover all types of marker-trait association analysis. Other approaches include comparative analysis using mutated populations and near-isogenic (introgression) lines and selective analysis using sub-populations based on selective sweeps. This article is not intended to review all such reported marker-trait associations, as there are several articles and reviews on this for a range of crop plants, including cereals. For example, Roy et al. (2011) reviewed the QTL (quantitative trait loci) that have been identified in numerous wheat, barley and rice populations for abiotic stress tolerance (drought, cold, heat, mineral toxicity, salinity and nutrient deficiencies). Also, the reviews by Langridge et al. (2006), Collins et al. (2008), Genc et al. (2010) and Fleury et al. (2010) provide a more comprehensive coverage of QTL linked to abiotic stress tolerance in cereals and other plant species.

Various statistical methods have been developed, particularly for linkage mapping using bi-parent populations (see Xu 2010 for a comprehensive review). These methods consider various genotypic and environmental effects and their interactions, including epistasis, genotype-by-environment (G × E) interaction, or different types of traits that need some special treatments, such as developmental traits (dynamic and conditional QTL mapping using the same traits scored at different stages), endosperm traits (ploidy with a different generation from their mother plants), and expression traits (eQTL). Integrated analyzes using all information publicly available, such as meta-QTL analysis and in silico mapping, have been also applied.

With more markers available to cover each gene, LD or association mapping has become a choice of mapping strategy. LD mapping has several advantages compared to linkage mapping, including the time and resources saved from generating segregating or immortal mapping populations, presence of multiple alleles in the population, and higher resolution than linkage mapping. However,

there are several factors that could result in false positives in association detection (Jannink and Walsh 2002; Flint-Garcia et al. 2003; Mackay and Powell 2007; Xu 2010), of which the most important is the population structure that can be removed through some statistical approaches. Another constraint is that traits controlled by the genes with rare alleles cannot be mapped effectively and in some cases, novel alleles we are looking for do not exist in the population at all, which can be only mapped using biparental populations with the target allele segregating.

To effectively combine the advantages offered by both linkage and LD mapping approaches, a joint linkage-LD mapping strategy has been proposed (Manenti et al. 2009; Myles et al. 2009; Lu et al. 2010). The joint mapping can be done through parallel mapping, which run linkage and LD mapping using biparental and natural populations separately, or integrated mapping using a single mapping procedure combining the information from both biparental and natural populations. The first joint mapping has been reported in maize using both parallel and integrated mapping approaches (Lu et al. 2010), which involved using three RIL populations and one natural population with 305 inbred lines, genotyped by 2053 SNP markers. Joint mapping for anthesis-silking interval (a trait for drought tolerance) identified 18 additional QTL that could not be identified by linkage or LD mapping alone. For the 277 SNPs that were excluded from LD analysis due to minor allele frequency of <5 %, 93 were polymorphic in the one of the RIL populations with normal allele frequencies recovered and three of these markers were associated with the target trait. Considering NAM populations as an alternative for joint linkage-LD mapping, three strategies can be developed by using biparental and natural populations together: several biparentals + one natural population (as shown by Lu et al. 2010); combined use of multiple populations such as NAM (Yu et al. 2008); more biparental populations contained in NAM plus one big natural population. The last option may be considered as the best and is achievable particularly when high-density genotyping is becoming cheaper and precision phenotyping becoming practicable for a large number of samples. In addition, joint mapping can be extended to multi-parent populations such as MAGIC.

Haplotype-based mapping can be used to replace individual marker-based mapping to improve the mapping power and depending on how a haplotype is constructed, it can be used to identify specific alleles within a gene or allele combinations at different loci that contribute to the same target trait. In maize, the use of haplotypes constructed using all SNPs within 10 kb-windows improved mapping efficiency for anthesis-silking interval, with the sum of phenotypic variation explained (PVE) increasing from 5.4 to 23.3 % for single SNP-based analysis (Lu et al. 2010). A similar result in maize was also obtained through comparative SNP and haplotype analysis for plant height and biomass as secondary traits of drought tolerance (Lu et al. 2012).

Selective genotyping of individuals from the two tails of the phenotypic distribution of a population provides a cost-effective alternative to analysis of the entire population for trait mapping (Lebowitz et al. 1987; Lander and Botstein 1989). Past applications of this approach have been confounded by the small size of entire and tail populations, and insufficient marker density, which result in a high probability of false positives in QTL detection (Xu and Crouch 2008). The effect of these

factors on the power of QTL detection was investigated by simulation of mapping experiments using population sizes of up to 3,000 individuals and tail population sizes of various proportions, and marker densities up to one marker per cM (Sun et al. 2010). The results indicate that selective genotyping can be used, combined with pooled DNA analysis, to replace genotyping the entire population, for mapping QTL with relatively small effects as well as linked and interacting QTL. Using phenotypic extremes from diverse germplasm including all available genetic and breeding materials, it is theoretically possible to develop an "all-in-one plate" approach where one 384-well plate could be designed to map 192 traits a time, which would include almost all agronomic traits of importance in a crop species. This "all-in-one plate" concept has been examined using extreme phenotypes collected in maize from over ten segregating populations genotyped with a 1536-SNP chip. Actually pooled DNA analysis can be used with any genotyping platforms, because for the tightly linked markers two DNA pools will have different alleles and they can be scored correctly although non-linked markers may not be scorable at all.

As many crops, including rice, maize and sorghum, have been sequenced de novo, and many more crops including wheat are expected, millions of SNP can be developed for each crop as discussed above. High-density genetic maps or GBS has significantly improved linkage mapping using biparental populations. A recent example involved a QTL analysis of 150 rice RILs derived from a cross between two varieties, *Oryza sativa* ssp. *indica* cv. 93–11 and *Oryza sativa* ssp. *japonica* cv. Nipponbare (Wang et al. 2011a, b). The RILs were genotyped through resequencing, which accurately determined the recombination breakpoints and provided a new type of genetic markers, recombination bins, for QTL analysis. A total 49 QTL were identified with phenotypic effect ranging from 3.2 to 46.0 % for 14 agronomics traits. Five QTL of relatively large effect (14.6–46.0 %) were located on small genomic regions, where strong candidate genes were found. The analysis using GBS thus offers a powerful solution to map QTL with high resolution. Although the number of markers is fast becoming unlimited for a given crop species, the size of the population used for genetic and breeding purposes is becoming a constraint from the phenotyping viewpoint.

High-density SNP data enables GWA to test all the genes in the genome for their association with target traits. In rice, one GWAS was performed for 14 agronomic traits (Huang et al. 2010). In maize, the NAM population has been used for analysis of leaf architecture (Tian et al. 2011) and quantitative resistance to southern corn leaf blight (Kump et al. 2011). On the other hand, numerous markers generated by resequencing are also useful in high-resolution linkage mapping using biparental populations (Huang et al. 2009; Xie et al. 2010). Thus, next-generation sequencing technology (Varshney et al. 2009), in combination with GWAS strategy, offers powerful tools for dissecting complex traits. Such new technologies have the potential to accelerate the detection and cloning of QTL, which enable pyramiding favorable QTL alleles, and to make desired changes in agronomically and nutritionally important traits.

As more and more QTL information for a range of important traits in crop plants becomes publicly available, approaches for integrated analysis of all such data have

received greater attention. Two basic strategies of exploiting the currently available information are meta and in silico analyzes, which have been described in detail elsewhere (Xu 2010). Meta-analysis was done in maize for ADL/NDF-related traits (Barrière et al. 2008). By projecting QTL on a reference map, the IBM2 Neighbors map, the QTL confidence intervals (CIs) were compared and 43 different loci based on 58 individually observed QTL were determined with hot-spots of three or more QTL identified on bins 2.08, 5.03, 6.04 and 9.06. In another example, meta-analysis has been performed for 59 QTL for traits associated with digestibility and 150 QTL for traits associated with cell wall composition from 11 different mapping experiments (Truntzler et al. 2010). A total of 26 and 42 meta-QTL were identified for digestibility and cell wall composition traits, respectively. To reveal the genetic architecture for flowering time and photoperiod sensitivity, a comprehensive evaluation of literatures was performed recently, followed by a meta-analysis of photoperiod sensitivity (Xu et al. 2012a). A total of 25 synthetic consensus loci and four hot-spot genomic regions were identified for photoperiod sensitivity. These regions include13 genes with known functions, which relate to photoperiod response or flower morphogenesis and development in maize.

Recently, integration of genetic maps and detected QTL has been done across three different generations of a specific mapping population ($F_{2:3}$, RIL, and BC_2F_2) derived from a cross between dent corn inbred Dan232 and popcorn inbred N04 (Li et al. 2011). In total, 103 QTL, 42 pairs of epistatic interactions and 16 meta-QTL (mQTL) were detected. Twelve out of 13 QTL with contributions (R^2) over 15 % were consistently detected in 3 ~ 4 environments (or in combined analysis) and integrated in meta-QTL. In wheat, the genetic architecture of Fusarium Head Blight (FHB) resistance in hexaploid wheat was revealed by QTL meta-analysis (Löffler et al. 2009). As more QTL will be fine mapped and a reference genome with high density markers becomes available, it can be expected that meta-analysis and in silico mapping will play an important role to identify consistent major gene loci and important genomic regions for MAS.

14.2.6 Marker Development and Validation

The purpose of establishing marker-trait association is to utilize the information in MAS-based breeding programs. As the recombination between markers and genes for the target trait is proportional to the power of MAS for major gene-controlled traits, development of genic and functional markers becomes increasingly important for marker-assisted gene introgression. Completion of de novo sequencing has facilitated map-based gene cloning with many genes cloned, particularly in rice (Qiu et al. 2011; Miura et al. 2011). Cloned rice genes include grain number, grain size and weight, heading date, disease resistance, salt tolerance, cold tolerance, submergence tolerance, and yield-related domestication genes (Miura et al. 2011). However, genes have been identified and cloned using intersubspecific species by using the large diversity. Another gene cloning strategy is based on comparative genomics.

With complete reference genome sequences available for many species that are closely related to each other, comparative genomics approach has returned to their positions for gene cloning and marker development based on sequence homology.

Marker-trait association established for major genes/QTL using one specific population needs to be validated before it can be used for MAS with target populations (Fig. 14.1). There are some excellent examples for marker validation. In barley, SSR markers linked to net-form net blotch resistance QTL (QRpt6) and spot-form net blotch resistance QTL (QRpts4) were validated in two barley populations unrelated to the original mapping population. The lines homozygous for the resistant-parent alleles at both marker loci in two other populations had significantly lower infection than lines homozygous for the susceptible-parent alleles at both loci (Grewal et al. 2010).

In rice, the C to A mutation in the second exon of *GS3* was reported to be functionally associated with enhanced grain length in rice. Besides the C-A mutation, three novel polymorphic sites, SR17, RGS1, and RGS2, were discovered in the second intron, the last intron and the final exon of *GS3*, respectively (Wang et al. 2011a). A number of alleles at these four polymorphic loci were observed in a total of 287 accessions collected worldwide including wild relatives. The result indicated that A allele at SF28 was highly associated with long rice grain while various motifs of (AT)n at RGS1 and (TCC)n at RGS2 were mainly associated with medium to short grain in Chinese rice. The C-A mutation at SF28 explained 33.4 % of the grain length variation in the whole rice population tested, whereas (AT)n at RGS1 and (TCC)n at RGS2 explained 26.4 and 26.2 % of the variation, respectively. The genic marker RGS1 based on the motifs (AT)n was further validated as a functional marker using two sets of backcross recombinant inbred lines (Wang et al. 2011a). In another report, Ghd7 and Ghd8 were found to have pleiotropic effects on grain number per panicle, plant height and heading date. Comparative sequencing of some landraces and modern varieties showed the wider diversity in the promoter than other regions in Ghd7 and Ghd8. A series of near isogenic lines confirmed that Ghd7 and Ghd8 each have several functional alleles with strong or weak effects on target traits (Yuqing He, Huazhong Agric. Univ., China, personal communications).

In maize, sequence-tagged, PCR-based markers were developed and demonstrated for use in selecting favorable alleles of LCYE (*lycopene epsilon cyclase*), a crucial gene in the carotenoid pathway. Markers for favorable alleles of *LCYE* (Harjes et al. 2008) and for another critical gene in the pathway, *CrtR-B1* (carotene beta-hydroxylase 1), were developed (Yan et al. 2010b). Allele mining and marker development are also underway for other genes of the carotenoid biosynthetic pathway, including *PSY* (phytoene synthase) and *CCD* (carotenoid cleavage dioxygenases), giving hope that MAS will soon be possible for several genes which together explain much of the variation for provitamin A in maize.

In wheat, more than 30 genes have been cloned in common wheat and its relatives, from which about 100 functional markers for wheat quality, agronomic traits and disease resistance genes have been developed (Bagge et al. 2007; He et al. 2011; Liu et al. 2012). The processing quality of wheat-based products is highly associated with high- and low-molecular-weight glutenin subunits (HMW-GS and

LMW-GS), polyphenol oxidase (PPO) activity, lipoxynase (LOX) activity, yellow pigment content (YPC), kernel hardness, and starch properties. Most of the genes for these traits have been cloned and the functional markers have been developed. For the six genes cloned for disease resistance, functional markers were available for *Pm3* against powdery mildew and for *Lr34/Yr18/Pm38* locus against multiple diseases such as leaf rust, stripe rust and powdery mildew (Tommasini et al. 2006; Lagudah et al. 2009).

14.2.7 Decision Support Tools

A successful MAS program needs to be backed up with appropriate tools for data management, analysis, and decision making, particularly when large-scale genotyping and phenotyping is involved. Xu (2010) justified the requirements of breeding informatics and decision support tools in a series of molecular breeding procedures. Important tools include those for managing genotyping, phenotyping and environmental data and mining the data for marker-trait association, gene identification, G × E analysis, breeding simulation and MAS (Xu et al. 2012c). All these tools should be included in one-stop-shopping package and operated through webs and/or publicly available software. As an example of such effort, the Molecular Marker Toolkit was developed by the Generation Challenge Program (GCP) to include markers that are effectively used in breeding programs and compiled from a literature review and from contacts with crop experts (Van Damme et al. 2011). The information given for each marker is of immediate use and includes laboratory protocols, validation processes, and key references.

14.3 Marker-Assisted Selection Methodologies

Several MAS schemes have been designed and used in plant breeding (Fig. 14.2). Each is suitable for specific types of traits and breeding programs. They may be used individually or combined in one breeding program for improvement of a specific trait or multiple traits. Marker-assisted foreground selection and background selection have been proven very useful for breeding major gene controlled traits, which has been discussed in detail elsewhere (Hospital and Charcosset 1997; Stam 2003; Frisch 2004; Xu 2010). However, marker-assisted foreground selection is not effective for QTL with small effect, because marker-trait association mapping fails to detect rare or small effect QTL, only captures a portion of the genetic variance (Goddard and Hayes 2007), can lead to overestimated marker-effects (Lande and Thompson 1990; Beavis 1998), and may not be relevant across breeding populations, in different environments, or after several cycles of selection (Podlich et al. 2004). For complex traits that are generally controlled by QTL with small effect, two major MAS schemes, MARS and GS, have been proposed to be effective.

14.3.1 Marker-Assisted Recurrent Selection

Marker-assisted recurrent selection (MARS) was proposed in the 1990s (Edwards and Johnson 1994; Lee 1995; Stam 1995), which uses markers at each generation to target all traits of importance and for which genetic information can be obtained. When the QTL mapping is conducted based on a biparental population, parents often contribute different favorable alleles. As a result, the ideal genotype is a mosaic of chromosomal segments generated by recombination between the two parents. To produce or approach this ideal genotype, several successive generations of crossing individuals will be needed (Stam 1995; Peleman and van der Voort 2003). MARS refers to the improvement of an F_2 population by one cycle of marker-assisted selection (i.e., based on phenotypic data and marker scores) followed commonly by two or three cycles of marker-based selection (i.e., based on marker scores only). This idea can be extended to situations where favorable alleles come from more than two parents (Fig. 14.2). MARS can also start without any QTL information, and selection can be based on significant marker–trait association established during the MARS process. Simulation studies revealed that MARS was generally superior to phenotypic selection in accumulating favorable alleles and was between 3 % and almost 20 % more efficient than phenotypic selection (van Berloo and Stam 2001). The usefulness of including prior knowledge of QTL under genetic models has been examined that included QTL number, heritability, gene effects, linkage and epistasis. It is concluded that with known QTL, MARS is most beneficial for traits controlled by a moderately large number of QTL (Bernardo and Charcosset 2006).

14.3.2 Genomic Selection

GS, or genome-wide selection, as the phrase implies, has been defined in a very narrow sense to refer to marker-based selection without significant testing or identifying a subset of markers associated with the trait (Meuwissen et al. 2001). GS consists of three steps (Fig. 14.2; Rutkoski et al. 2011): (1) prediction model training and validation, (2) breeding value prediction, and (3) prediction-based selection (Rutkoski et al. 2011). In model training, marker effects are estimated for a training population that consists of germplasm having both phenotypic and genome-wide marker data. The combination of these marker effect estimates and the marker data of the single crosses are used to calculate the genomic estimated breeding values (GEBVs), the sum of all marker effects included in the model for an individual. Selection is then processed for the single crosses using GEBVs as the selection criterion. Thus, GS attempts to capture the total additive genetic variance with genome-wide marker coverage and effect estimates, contrasting with MARS strategies where only a small number of significant markers are used for prediction and selection (Rutkoski et al. 2011). Therefore, GS is more suitable for quantitative traits controlled by a large number of genes each with a small effect.

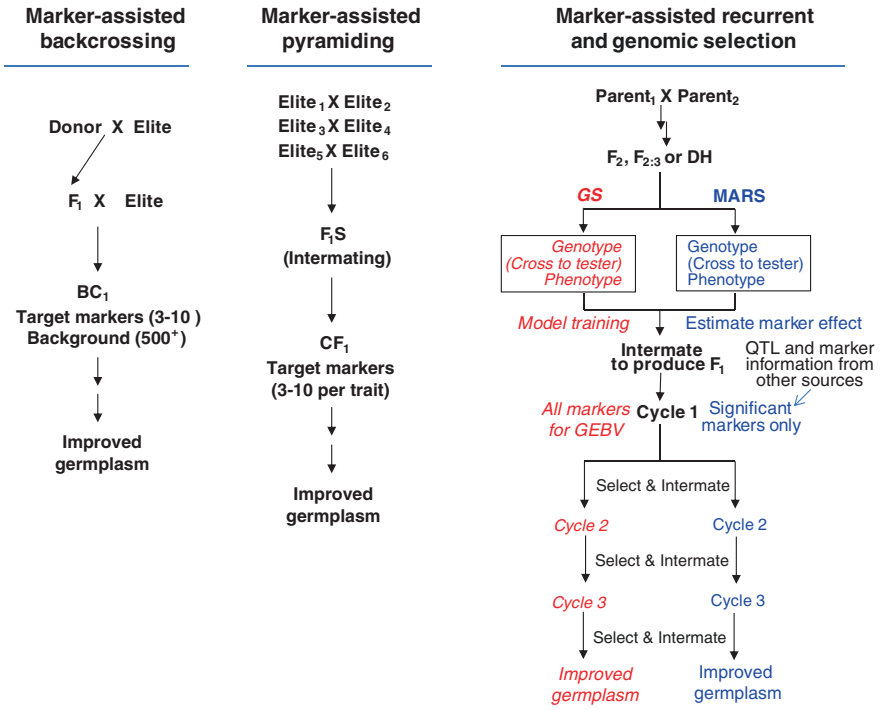

Fig. 14.2 Methodologies for marker-assisted breeding. Genomic selection (*GS*, *red*) and marker-assisted recurrent selected (*MARS*, *blue*) can started with the same type of population, F_2, $F_{2:3}$, BC, or DH. Crossing to tester can be included in the procedure for hybrid crops. Marker/QTL information from other sources can be incorporated for selection with the significant markers identified at the beginning stage of the current MARS. Results from GS of breeding populations can be used to improve the model training and prediction for next cycles of selection or other GS projects

GS is poised to revolutionize plant breeding. In GS, marker data are used to predict breeding line performance in one analysis; the breeding populations are analyzed directly; all markers are included in the model so that effect can be estimated unbiasedly with small effect QTL accounted for. The prospects for GS for improving quantitative traits in maize was analyzed, with a conclusion that this approach is superior to MARS for improving complex traits. GS is more expensive but effectively avoids issues associated with QTL number, allele distribution and epistatic effects (Bernardo and Yu 2007).

The frequency of phenotyping can be reduced with GS because selection is based on genotypic data. Selection in off-seasons using marker information can also reduce cycle time, thereby increasing annual genetic gains from selection. To date, the only publicly available results on large-scale GS performance are from dairy cattle breeding programs. Both simulation (Wong and Bernardo 2008; Zhong et al. 2009) and empirical studies (Lorenzana and Bernardo 2009) found that in plant populations GS would lead to greater gains per unit time than phenotypic selection.

A simulation study found that compared to MARS, GS in maize could increase response from selection, especially for traits of low heritability (Bernardo and Yu 2007). GS with GEBV accuracies of only 0.5 could lead to a twofold higher gain per year compared to MAS in a low-investment wheat breeding program and a threefold increase in a high-investment maize breeding program (Heffner et al. 2010).

Some theoretical issues associated with GS have been reviewed by taking durable stem rust resistance in wheat as an example (Rutkoski et al. 2011). First, marker density should be high enough to effectively cover the entire genome so that at least one marker should be in LD with each gene region. The minimum number of markers required to achieve genome-wide coverage depends on LD decay rates which vary widely across species, populations, and genomes due to forces of mutation, recombination, population size, population mating patterns, and admixture (Flint-Garcia et al. 2003). Second, the training population should be large in order to achieve the highest GS accuracies, and it should consist of the parents or very recent ancestors of the population under selection, with multiple generations of training. Third, the strength of marker effect determines the relative accuracy of the statistical methods for training the GS model including ridge regression best linear unbiased prediction (RR-BLUP), Bayes-A, and Bayes-B (Meuwissen et al. 2001). Fourth, traits with lower heritability require larger training populations to maintain high accuracies as indicated in cattle (Hayes et al. 2009). Fifth, for crops where insufficient seed production inhibits the use of selection and recombination in the F_1 generation, a modified recurrent selection scheme using GS among F_2 individuals is proposed (Bernardo 2010).

Although GS has been widely considered as a viable MAS approach for complex traits, three specific issues need to be considered (Rutkoski et al. 2011): (a) if landraces or exotic germplasm is used in GS, this may require very high marker densities and large training population sizes to retain accuracy (Meuwissen 2009); (b) for single crosses unrelated to the training population, marker effects will be inconsistent because of the presence of different alleles, allele frequencies, and genetic

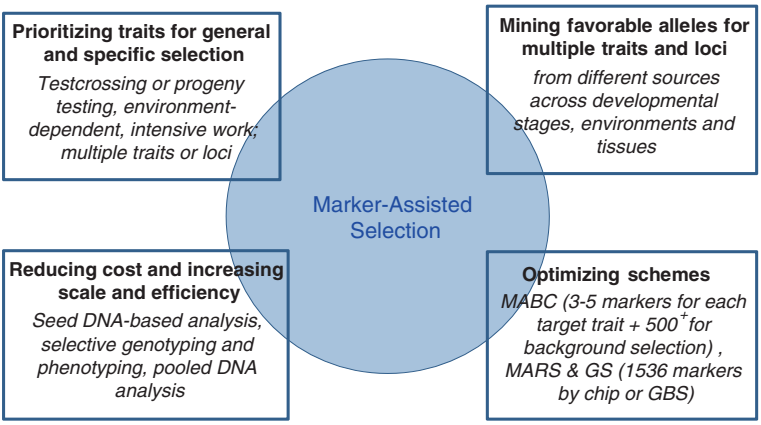

Fig. 14.3 Strategies for marker-assisted breeding

background effects (Bernardo 2008). These factors may result in inaccurate estimates of marker effect and thus, poor GEBV accuracies; (c) as the overall importance of the influence of QTL x genetic background interactions is not clear, QTL effects across unrelated individuals may be inconsistent and could undermine GEBV accuracies for the training population containing unrelated individuals. As it is difficult to have all these complications included a model training, simulation may hardly reflect the real situations GS is involved. In this case, the genetic gains and accuracy of GS should be evaluated with care, particularly when distant or multi-parental populations are used.

14.4 Strategies for Marker-Assisted Selection

Developing strategies for MAS needs to take many factors into consideration and in many cases, matters with which options are to be chosen (Xu 2003, 2010). Examples of some options available including leaf- versus seed- DNA-based genotyping, single versus bi-marker or multiple markers, targeting regional versus whole genome, heterogeneous versus homogeneous backgrounds, qualitative versus quantitative traits, single versus multiple traits, physical versus functional markers, and MABC versus MARS or GS approaches. In this section, several strategies for MAS will be discussed (Fig. 14.3).

14.4.1 Traits Most Suitable for Major Gene Selection

There are six situations that are most suitable for MAS of major genes (Xu 2002, 2010; Table 14.1): (1) selection without testcrossing or progeny test—many traits need testcrossing and a progeny test for unambiguous identification and typical examples include male-sterility restorability, wide compatibility, heterosis and combining ability; (2) selection independent of environments—many traits must be screened in specific or controlled environments (such as biotic and abiotic stress tolerances); (3) selection without laborious field or intensive laboratory work—many traits are phenotypically invisible or unscorable by visual observation and must be measured in the laboratory using sophisticated equipment or facilities; (4) selection at an early breeding stage—some traits are only measurable at or after the reproductive stage such as grain quality that can only be tested using mature seeds; (5) selection for multiple genes and multiple traits—in some cases, multiple pathogen races or insect biotypes must be used to identify plants for multiple resistances, but, in practice, this may be difficult or impossible, because different genes may produce similar phenotypes that cannot be distinguished from each other; (6) whole genome selection—MAS can also be practiced at the whole genome level to drastically reduce the donor genome in backcross breeding or to get rid of linkage drag when a wide cross is involved. Except for the first situation that only works for hybrid crops (e.g., maize and rice), the rest of the five situations work equally well for all cereals.

14.4.2 Pyramiding Multiple Loci and Favorable Alleles

To create a superior genotype, the breeder must assemble many genes which work well together and, for a specific trait, assemble the alleles with similar effects from different loci. This process is called "gene pyramiding", by which desirable alleles of different major QTL can be brought together and the true breeding lines associating alleles of similar (positive or negative) effect can be selected (Xu 1997). A general framework was developed to optimize breeding schemes to accumulate identified genes from multiple parents into a single genotype (gene pyramiding schemes) (Servin et al. 2004; Fig. 14.2). Whole genome selection schemes, including MARS and GS, can also be implemented to assist pyramiding favorable alleles from different sources, although not through marker-assisted backcrossing.

14.4.3 Selection for Multiple Traits

The methods for pyramiding favorable alleles affecting a specific trait can be used in the same way to accumulate QTL controlling different traits. A distinct difference in concept is that alleles at different trait loci to be accumulated may have different favorable directions, i.e. negative (decreasing) alleles are favorable for some traits but positive (increasing) alleles are favorable for others. Therefore, one may need to combine the positive QTL alleles of some traits with the negative alleles of others to meet breeding objectives. Marker-assisted gene pyramiding is also important when considering multiple traits, as in phenotypic selection each of these traits has to be tested in different environments, different developmental stages or different stages of a breeding program.

Strategy development for multiple trait improvement will include understanding the genetic relationships among different traits (including the interaction between component or secondary traits of a very complex trait such as drought tolerance); genetic dissection of the developmental correlation between multiple traits; understanding of genetic networks for correlated traits; and construction of selection indices with multiple traits. Much progress has already been made in this area especially for complex traits like drought tolerance in crops such as maize (Edmeades et al. 2000; Bänziger et al. 2006) and wheat (Babar et al. 2006, 2007).

Selection for multiple traits may be completed in one step as long as the population is large enough to allow desirable individuals to combine different traits. However, the number of trait loci that can be manipulated in one step is limited as the population size required to cover the recombinants increases exponentially with the increase of the number of traits/loci. To manipulate multiple genes/traits that are beyond the population sizes that are amenable, a two-stage selection strategy involved two generations proposed by Bonnet et al. (2005) and simulated by Wang et al. (2007) can be employed. In this approach, individuals are selected first by all target markers for both homozygous and heterozygous forms to obtain a subset of population that contain higher frequencies of the target alleles so that a much smaller population size is required in the following generation to obtain the homozygotes at the target loci.

Table 14.1 Examples of target traits for marker-assisted breeding in cereals

Priority	Category	Example
Testcrossing or a progeny test required	Crossing ability	Wide compatibility (hybrid rice)
	Sterility	Fertility restorability, CMS (hybrid crops)
	Combining ability	Hybrid performance across multiple crosses (hybrid crops)
Environment-dependent	Biotic stresses	Disease/insect resistance
	Abiotic stresses	Stresses caused by water, temperature, light, air, fertilization, soil problems, weeds
	Response to neutral environmental factors	Delayed flowering and male sterilities regulated by photoperiod and temperature
Intensive lab/field work involved	Quality	Physical and chemical properties and nutrient contents of final products
	Physiological features	Photosynthesis, carbon isotope discrimination (rice and wheat)
	Root characters	Numbers, sizes, distribution and profile
Selection made at late or advanced stages	Maturity	Days to flowering and maturity
	Final product	Grain yield, harvest index
Whole genome selection (MARS and GS)	Complex and multiple traits	Yield and hybrid performance controlled by large numbers of favorable alleles and their combinations across genome with small effects that cannot be separated from each other or from environmental effects

14.4.4 Reducing Costs and Increasing Scale and Efficiency

Wide adoption of MAS depends on the cost, scale and efficiency (Fig. 14.3). One of the strategies that have been developed to reduce the cost and increase scale and efficiency is seed DNA-based genotyping (Gao et al. 2008). Compared to MAS using DNA extracted from leaf tissues, seed DNA-based genotyping has many advantages, including: (1) discarding undesirable genotypes before planting; (2) increasing the speed of breeding cycles by selecting genotypes during the off season; (3) reducing the time-consuming and error-prone sample collecting step that currently involves harvesting leaf tissue from plants which then need to be retraced after genotyping; and (4) saving land because only selected genotypes (seeds) are planted. To develop a comprehensive and operational system for MAS using single-seed-based and non-destructive DNA extraction, the extracted DNA must have a high quality compared to leaf-tissue DNA. Similarly, the quantity of DNA should be large enough for whole genome genotyping and DNA extraction should

be high-throughput, while sampled seeds should maintain a high level of germination with no damage to seedling establishment. A seed DNA-based genotyping system that is feasible for crop species with relatively large seeds has been developed in CIMMYT for maize (Gao et al. 2008). Several multinational companies have established high-throughput, single seed-based genotyping platform for routine MAS in almost all crops including those with very small seeds, which usually consists of a seed chipping facility. To obtain high quality DNA, protocols will have to be optimized for crops that differ significantly in seed structure and composition. Caution should be taken for monocot species including all cereals as the DNA sample extracted from endosperm is triploid and one generation ahead of the mother plant. In addition, hetero-fertilization can result in the endosperm genotype different from the embryo (plant per se), causing a mismatching or wrong score. In maize, the frequency of hetero-fertilization is highly population-dependent, which could vary from 0.14 to 3.12 % (Gao et al. 2011). There is very limited information available about the hetero-fertilization frequencies in other crops.

There are some other approaches that can be used to reduce cost and increase scale and efficiency. The selective genotyping and pooled DNA analysis discussed for trait mapping can be extended and used for MAS related genotyping and for all types of extremes including those selected from germplasm collection and natural populations. Multiplexing can be utilized in different genotyping protocols for various types of markers to reduce the cost per data point and also increase the marker number per run. High-throughput phenotyping will not only provide high quality phenotyping data but also increase scale although the cost-effectiveness may vary case by case. The selective strategy developed for genotyping may be also exploited for selective phenotyping as genotyping becomes increasingly cheaper (Xu et al. 2012c). This has become a routine procedure in several companies where large numbers of DH lines created can be eliminated by genotyping for the presence of important target genes, alleles, haplotypes and genomic regions/blocks before they move to field trials, because genotyping a line is cheaper than a multi-location phenotyping. The last strategy proposed for reducing cost and increasing scale and efficiency (Xu and Crouch 2008; Xu 2010) is integrated genetic diversity analysis, genetic mapping and MAS, which have been separate, successive procedures and now can be combined through MARS and GS approaches.

14.4.5 Optimization of MAS Schemes

A suitable MAS scheme should be chosen based on breeding objectives (Fig. 14.3). For major gene introgression (selection for target genes only), 3–10 markers for each target trait will be needed to mark the gene (with functional markers available) and their flanking regions (with only closely linked markers available). Depending on how many traits to be introgressed at a time, the number of plants required for obtaining desirable target plants or recombinants may

vary from several hundred to several thousand or more. For MABC (selection for both target genes and background), five hundred or more evenly-distributed markers, plus 3–10 markers for each target trait, will be required to reach a reasonable genome coverage for background selection. For whole genome or genome-wide assay such as MARS and GS for complex traits, several thousands to millions of markers may be genotyped for hundreds or thousands of plants.

14.5 Application of Marker-Assisted Selection in Cereals

Molecular markers have already become important for predicting breeding value and are especially important for introgression of favorable alleles into adapted germplasm. Marker-assisted gene introgression has been facilitated by developing exotic genetic libraries (also known as CSSL, chromosome segment substitution line; IL, introgression line; or CL, contig line) to enhance utilization of wild relatives to expand crop gene pools. The use of MAS in plant breeding is now routine in commercial breeding programs to increase gains from selection per unit time (Eathington et al. 2007), although it has been limited in public breeding (Xu and Crouch 2008).

14.5.1 Gene Introgression from Wild Relatives or Alien Germplasm

Many genes and alleles that are favorable for agronomic traits can be only found in wild relatives or alien germplasm. McCouch et al. (2007) summarized results from a decade of collaborative research using advanced backcross populations derived from *Oryza rufipogon* to: (1) identify QTL-associated improved performance in cultivated rice *Oryza sativa*; and (2) to clone genes underlying key QTL of interest. They demonstrated that advanced backcrossing (AB)-QTL analysis is capable of: (1) successfully uncovering positive alleles in wild germplasm that were not obvious based on the phenotype of the parent; (2) offering an estimation of the breeding value of exotic germplasm; (3) generating NILs that can be used as the basis for gene isolation and also as parents for further crossing in a variety development program; and (4) providing gene-based markers for targeted introgression of alleles.

Using AB-QTL analysis, yield and grain quality enhancing alleles from wild relatives have been successfully introgressed in rice, wheat, barley, and sorghum. Dramatic yield advantages have been reported in rice, for example, through the introduction of two yield-enhancing QTL alleles (*yld1.1* and *yld2.1*) from *O. rufipogon* (AA genome) into 9311. This contributed in excess of 20 % yield increases in rice (Liang et al. 2004). In contrast, only 6–8 % increase in grain yield was reported when positive alleles from *Hordeum spontaneum* were introgressed into barley. Rice with genes introgressed from *Oryza rufipogon* has been shown to have better Al tolerance when grown on toxic soils (Nguyen et al. 2003), while

the introduction of genes from *O. logistaminata* has increased the drought tolerance of cultivated rice (Hajjar and Hodgkin 2007). Wild relatives also contributed positive alleles for improved grain characteristics in rice (long, slender and translucent grains and grain weight), wheat (grain weight and hardness) and barley (grain weight, protein content and some malt quality traits). Of particular interest is a locus for grain weight, *tgw2*, which contributed positive alleles from *O. grandiglumis* that are independent from undesirable effects of height and maturity (Yoon et al. 2006). More recently, two Na^+ exclusion genes, *Nax1* and *Nax2* (James et al. 2006; Lindsay et al. 2004), have been introgressed into the durum wheat variety Tamaroi, with those lines containing the *Nax2* gene having increased yield in saline soils, and importantly, no yield penalty under non-stressed conditions.

14.5.2 Gene Transfer or Introgression between Elite Germplasms

MABC has been widely used to transfer or introgress genes from one elite line to another. To improve the hybrid rice currently widely grown in China, a series of MAS were performed. (1) *Xa21*, *Xa7* and *Xa23*, three wide spectrum bacterial blight (BB) resistance genes, were introgressed to the restorers Minghui 63 and 9311 by MAS (Chen et al. 2000). (2) Two genes, *Pi1* and *Pi2*, showing broad-spectrum resistance to fungi blast, were introgressed into the maintainer Zhenshan 97. (3) Three genes, *Bph14*, *Bph15* and *Bph18*, highly resistant to brown planthopper (BPH), were introgressed to Zhenshan 97, Minghui 63 and 9311 to improve their BPH resistance. To improve the grain quality for hybrid rice, Zhou et al. (2003) successfully introduced the *wx-MH* fragment from Minghui 63 into the maintainer Zhenshan 97, which was subsequently transferred to Zhenshan 97A, using MABC. The introduction of this fragment has greatly improved the cooking and eating quality of inbred lines and their resultant hybrids. Introgression of *Wx-T* allele (conferring intermediate amylose content) into two maintainers (Longtefu and Zhenshan 97) and their relevant male-sterile lines resulted in improved *indica* hybrids. The *Lgc-1* locus conferring low glutelin has been successfully incorporated into *japonica* rice with 93–97 % selection efficiency using two SSR markers (Wang et al. 2005). To improve the disease resistance for Basmati rice, Pusa1460 was utilized as the donor for introgressing BB resistance genes *xa13* and *Xa21* into Pusa6B and PRR78, the two parental lines for aromatic hybrid rice Pusa RH10 (Basavaraj et al. 2010). Two BB-resistant rice cultivars, Improved Pusa Basmati-1 (Pusa 1460) and Improved Sambha Mahsuri, were developed (Joseph et al. 2004; Sundaram et al. 2009), and released for commercial cultivation. Traditional basmati varieties were also improved for BB resistance and plant height using MABC to transfer two BB resistance genes, *xa13* and *Xa21*, and semidwarfing gene, *sd-1*, into two traditional basmati varieties, Basmati 370 and Basmati 386 (Bhatia et al. 2011). MABC has also been used to transfer other genes important for hybrid rice breeding, including a photoperiod-sensitive male sterility gene from cv. Lunhui422S (indica) and the yellow leaf gene from

line Yellow249 (indica) into the elite japonica cv. Zhendao88. Fertility genes S5, S8, S7 and S9 have also been tagged using molecular markers (Chen et al. 2011).

Except for hybrid rice, MABC has been also used to improve elite rice inbred varieties. For example, the alleles of Azucena (upland rice) at four QTL for deeper roots on chromosomes 1, 2, 7, and 9 have been transferred from selected DH lines (Shen et al. 2001). Kalinga III, an upland *indica* variety, was improved for drought tolerance by MABC. The target segment on chromosome 9 (RM242-RM201) significantly increased root length under both irrigated and drought stress environments (Steele et al. 2006). Two QTL controlling resistance to rice yellow mottle virus were introgressed into the variety IR64 (Ahmadi et al. 2001).

In wheat, a study was undertaken to assess the effect on improving FHB resistance and on possible unwanted side effects (linkage drag) of two resistance QTL, *Fhb1* and *Qfhs.ifa-5A*, from the spring wheat line CM-82036 when transferred by MABC into several European winter wheat lines (Salameh et al. 2011). In USA, 27 different disease and pest resistance genes and 20 alleles with beneficial effects on bread-making and pasta quality were incorporated into about 180 lines adapted to the primary US wheat production regions (Sorrells 2007). In India, most of the marker-trait associations that were discovered during 1996–2009 were validated/utilized for introgression of QTL/genes for various traits including grain protein content, preharvest sprouting tolerance, grain weight, and resistance against leaf rusts (Kumar et al. 2010).

In maize, *opaque2*-specific SSR markers were utilized for conversion of the normal maize lines into quality protein maize (QPM) lines with enhanced nutritional quality (Babu et al. 2005; Gupta et al. 2009). The parental lines of 'Vivek Hybrid 9' (CM145 and CM212) were converted into QPM versions through transfer of *o2* gene using MAS and phenotypic screening for endosperm modifiers. QPM versions of six elite inbred lines, which are the parents of three single-cross hybrids, PEHM2, Parkash and PEEHM5, have also been developed (Prasanna et al. 2010). In African maize breeding, *o2* allele specific SSR markers were used to convert herbicide resistance maize lines into QPM which is the equivalent of modified *o2* phenotype. Using the three SSR markers, the result showed that 97 % of the lines were *o2* (Dwivedi et al. 2007). To improve drought tolerance, CIMMYT initiated a major MABC program to transfer five genomic regions involved in the expression of a short anthesis-silking interval from Ac7643 (a drought tolerant line) to CML247 (an elite tropical breeding line). Five genomic regions were transferred. The best five MABC-derived hybrids yielded, on average, at least 50 % more than the control hybrids under water stress conditions (Ribaut et al. 2002; Ribaut and Ragot 2007).

In barley, a variety of markers have been developed for selection of the *rym4* and *rym5* resistance genes for Barley Yellow Mosaic Virus complex (Rae et al. 2007). The regions of a typical Spanish barley line that carry VRNH1 and VRNH2 were introgressed into a winter variety, Plaisant. A set of 12 lines introgressed with all four possible combinations of VRNH1 and VRNH2 were evaluated for vernalization requirement and frost tolerance (Casao et al. 2011). To produce high yielding NILs that maintain traditional malting quality characteristics but transfer QTL associated with yield, via MABC, Schmierer et al. (2004) targeted Baronesse

chromosome 2HL and 3HL fragments presumed to contain QTL that affect yield. An identified NIL produced yield equal to Baronesse while maintaining a Harrington-like malt quality profile.

14.5.3 Selection for Multiple Targets and Gene Pyramiding

Gene Pyramiding for Major Genes

Many reports have been available for marker-assisted gene pyramiding. Dwivedi et al. (2007) listed some representative examples from barley, rice and wheat, most of which are for major genes. Gene pyramiding has been largely focused on the combinations of genes for the traits listed in Table 14.1 including resistance to multiple races of the same disease, resistance to different diseases, and resistance to both diseases and insects.

In rice, three blast resistance genes (*Pi-2*, *Pi-1* and *Pi-4*) have been pyramided into one variety. The pyramiding of *Bt*, Xa_{21}, and *wx* genes created an improved 'Shanyou 63' (He et al. 2002). QTL for increased grain number (*Gn1*) and QTL for reduced plant height [*Ph1*(*sd1*)] were pyramided in the Koshihikari background producing a 23 % increase in grain yield while reducing the plant height by 20 % compared with Koshihikari (Ashikari et al. 2005).

In wheat, powdery mildew (*Pm2*, *Pm4a*, *Pm6*, *Pm8*, and *Pm21*) pyramided lines and those with resistance to *Fusarium* head blight (six QTL), orange blossom midge (*Sm1*), and leaf rust (*Lr21*) were bred through MAS. DNA markers and DH technology were effectively used to produce gene pyramids of two or more of the stem rust resistance genes *Sr24* and new sources of *SrR*, *Sr31* and *Sr26* on reduced alien chromatin in the genetic backgrounds of Westonia and Pavon wheat (Mago et al. 2011).

In barley, pyramiding resistance gene has been carried out for several pathosystems (Friedt and Ordon 2007). An excellent example is for the genes against the barley yellow virus complex using the molecular markers identified for *rym4*, *rym5*, *rym9* and *rym11*. For pyramiding such recessive genes DH approach was used to increase homozygous recessive genotypes (Werner et al. 2005). Resistance to barley mild mosaic virus and barley yellow mosaic virus complex and stripe rust has been separately incorporated through MAS in barley. Many of these pyramided lines showed enhanced resistance to pests and diseases. In pearl millet, gene pyramiding has been used for the backcross transfer of QTL for downy mildew resistance (Witcombe and Hash 2000).

Gene Pyramiding for Complex Traits

Eathington (2005) and Crosbie et al. (2006) reported that the rates of genetic gain achieved for complex traits through MARS in maize were about twice those of phenotypic selection in some reference populations. Marker-only recurrent

selection schemes have been implemented for a variety of traits including grain yield and grain moisture (Eathington 2005), or abiotic stress tolerance (Ragot et al. 2000), and multiple traits have been targeted simultaneously. Selection indices were apparently based on 10 to probably more than 50 loci, these being either QTL identified in the experimental population where MARS was being initiated, QTL identified in other populations, or genes. Marker genotypes were generated for all markers flanking QTL included in the selection indices (Ragot et al. 2000). Plants were genotyped at each cycle and specific combinations of plants were selected for crossing, as proposed by van Berloo and Stam (1998). Several, probably three to four, cycles or MARS were conducted per year using continuous nurseries. As summarized by Ragot and Lee (2007), selection response of MARS can be attributed to: (1) rather large sizes of the populations submitted to selection at each cycle, (2) use of flanking versus single markers, (3) selection before flowering, (4) increased number of generations from one to four generations per year, and (5) lower cost of marker data points.

14.5.4 Marker-Assisted Breeding and Variety Development

MAS-derived varieties and advanced lines combining resistance to biotic and abiotic stresses or improved grain quality have been reported in rice, wheat, maize, barley, and pearl millet (Dwivedi et al. 2007; Jena and Mackill 2008; Gupta et al. 2010; Prasanna et al. 2010). Some of the improved germplasm lines were successfully utilized in breeding leading to release of varieties for commercial cultivation (Hospital 2009).

MAS, combined with conventional breeding, has resulted in a group of rice materials and released varieties with improved traits (Li et al. 2010). The major genes that have been used in MAS and variety development include disease resistance genes, such as *Xa4*, *Xa21*, *Xa23*, *R-sb2t*, *Pil*, *Pi-1*, *Pi-2*, *Pi-25*, *Pi-33*, *R-sbzt* etc., genes for grain quality such as *Wx*, sterility gene *Rf5* and genes for heading date (Wei et al. 2009; Wang et al. 2009). Grain quality traits such as 1,000-seed weight, kernel length/breadth ratio, basmati type aroma, and high amylose content have been combined with resistance to bacterial blight using MABC breeding (Ramalingam et al. 2002; Joseph et al. 2004). Development of submergence-tolerant varieties using MAS for *Sub 1* gene (Xu et al. 2006) has already been reported from Thailand (Siangliw et al. 2003) and tested for their adaptation to deepwater paddy cultivation in eastern India (Xu et al. 2006). Performance of tolerant varieties developed by MABC was evaluated to determine the effect of *Sub1* in different genetic backgrounds (Septiningsih et al. 2009). More recently, the SUB1 QTL was introduced into BR11, a rainfed lowland rice mega variety of Bangladesh through rapid and high-precision MABC (Iftekharuddaula et al. 2011). MAS-derived rice varieties are already being grown in the United States, China, Indonesia and India. These marker-aided rice varieties and hybrids have produced on average 11–34 % increased yield over popular inbred and hybrid varieties in

Asian countries. This has led to an estimated increase in grain harvest of 0.8 million Mt (worth US $20.5 million) of paddy rice per cropping season in India, Indonesia, the Philippines, and China as a result of the growing bacterial blight resistant inbred and hybrid varieties (Leung et al. 2004).

In maize, 'Vivek QPM Hybrid 9', developed through MAS, was released in India. The variety was developed through marker-assisted transfer of the *o2* gene and phenotypic selection for endosperm modifiers in the parental lines (CM145 and CM212) of Vivek Hybrid 9 (Babu et al. 2005; Gupta et al. 2009). The same approach was used to develop QPM versions of several elite, early maturing inbred lines adapted to the hill regions of India (Gupta et al. 2009) and QPM versions of six elite inbred lines (CM137, CM138, CM139, CM150 and CM151), which are the parents of three single-cross hybrids, PEHM2 (CM137 × CM138), Parkash (CM139 × CM140) and PEEHM5 (CM150 × CM151) (Khanduri et al. 2010). Major genes/QTL for resistance to turcicum leaf blight (*Exserohilum turcicum*) and Polysora rust (*Puccinia polysora*) have been pyramided in five elite lines, CM137, CM138, CM139, CM140 and CM212 (Prasanna et al. 2010). Similar efforts on MAS for generation of QPM lines and transfer of major QTL for SCMV resistance are being implemented in China, and the MAS products are in pipeline.

In wheat, MAS has been used in national and international breeding programs. In addition to validating molecular markers from other programs around the world, the Chinese Academy of Agricultural Sciences (CAAS)-CIMMYT joint wheat program plays a leading role in molecular marker development in collaboration with several provincial programs in China (Liu et al. 2012). The objectives were to clone genes such as the *Psy 1* genes on chromosome 7A and 7B associated with yellow pigment in flour, develop functional markers based on the allelic variants, and then validate them on Chinese wheat cultivars. More than 80 markers are available including these developed by other programs around the world, targeting such traits as quality, disease resistance, plant height, kernel weight, and adaptation. Currently, molecular markers are used to characterize parental lines, improve selection efficiency in backcross segregating populations, and confirm the presence of targeted genes in advanced lines. Three lines developed from MAS are being tested in regional trials, and they combined the high yield potential and outstanding quality of two parents (Zhonghu He, personal communications). In Western Australia Wheat Breeding Program, 42 traits/genes including a range of rust resistance genes (*Lr9, Lr19/Sr25, Lr24/Sr24, Lr34/Yr18, Lr46/Yr29, Lr47, Sr26, Sr32, Sr33* and *Sr36*) have been selected for variety development and germplasm enhancement through MAS (Cakir et al. 2008). In the US, 'Bringing Genomics to the Wheat Fields' project involved MABC to incorporate 27 different disease and pest resistance genes and 20 alleles with beneficial effects on bread-making and pasta quality into ~180 lines adapted to the primary US wheat production regions (Dubcovsky 2004). This involved more than 3,000 MAS backcrosses, and resulted in the development of ~240 backcross derived lines, and 45 released MAS-derived lines (Sorrells 2007). The germplasm was improved for a variety of traits using different genes including the following: (1) *pinB-D1b* for grain texture, (2) *Gpc-B1/Yr36* for grain protein content and stripe rust resistance, (3) *Lr47* for leaf rust resistance, (4) *Lr37/Yr17/Sr38* for resistance against all the three rusts, and (5) *H13*

for Hessian fly resistance. At CIMMYT, markers associated with 25 different genes controlling resistance against insect pests, protein quality, and other agronomic characters have been utilized in MAS-supplemented wheat breeding programs (William et al. 2007). At least 20 markers for *Rht, Ppd, Vrn* and the genes for resistance against a variety of pathogens have been used at CIMMYT for testing crossing blocks for designing crosses aimed at transfer/stacking of genes. In Europe, markers have been used to screen for resistance against soil borne viruses and track a range of rust resistance loci, vernalisation and photoperiod genes (*Vrn1, 2* and *3* and *Ppd*), dwarfing genes (*Rht 1, 2* and *8*), several quality related loci, and Fusarium head blight tolerance (*Fhb1* and *Fhb2*). MAS already led to the release of Lillian (DePauw et al. 2005), which possesses high grain protein content gene (*GPC-B1*), and Goodeve (DePauw et al. 2009), which possesses orange blossom wheat midge resistance gene (*Sm1*). The introgression of favorable rust and quality traits using MAS in combination with DH technology led to the development of a commercially useful line derived from 'Stylet' within 5 years as against 12 years needed in developing a variety through conventional breeding (Kuchel et al. 2008). MAS derived durum wheat variety 'Westmore' possessing *Yr36* gene responsible for resistance to stripe rust is also commercially available to US wheat growers. Gupta et al. (2010) provided a table to summarize the wheat varieties released and/or improved through MAS.

Among other crops, in sorghum three closely linked SSR markers (txp43, txp51 and txp211) were chosen to represent the three most significant QTL for seedling emergence and/or vigor. The result demonstrated the utility of linked markers for early-season cold tolerance in various genetic backgrounds and environments (Knoll and Ejeta 2008). In pearl millet, the first product of MAS was a downy mildew resistant pearl millet hybrid HHB-67-2 developed by ICRISAT and released for commercial cultivation (Hash 2005). Another example in pearl millet involved the use of a major QTL for increased grain yield and harvest index under terminal stress (Bidinger et al. 2005).

Although there are only few reports demonstrating successful use of MAS in plant breeding, the technology has demonstrated its potential as a tool to support conventional genetic enhancement of crops. It is possible that many more examples of successful applications of MAS are available with a number of commercial breeding companies around the world. There are also examples of desirable lines in the pipeline of public sector plant breeding, which will become available in future.

14.6 Summary and Outlook

The revolution in sequencing technology has significantly increased the availability of molecular markers and reduced drastically the genotyping cost, making it possible and also reasonable to work with a larger number of samples. High density molecular markers and large population size have significant impact on crop genetics and breeding. One such example is to break tight linkage between a target trait and undesirable characteristics when a wild relative is used in gene

introgression. For example, development of perennial cereals had a setback due to strong association of perennial nature with some undomesticated traits (Glove et al. 2010). Recent progress in developing perennial rice also indicated that the rare recombinants can be obtained when large segregating populations are used (Fengyi Hu, Yunnan Academy of Agricultural Sciences, China, personal communication). Similar progress can be achieved in maize and wheat, where progress in breeding for perennial nature has been slow during the last few decades.

As more genotypic data (G) are generated through high-throughput genotyping in various areas of genomics research, high-throughput and precision phenotyping will have to give more emphasis to keep pace with the genotypic data, and to generate precise phenotypic information (P). A third dimension of the data matrix will be the collection of data on environment (E), under which phenotyping data is recorded. This environmental assay has been called e-typing (Xu et al. 2012c). Although G × E interaction has been investigated through QTL mapping and statistical analysis using phenotypic data collected in different environments, it has been seldom exploited by integrative use of the three-dimensional data involving G-P-E. Since gene function and expression are eventually determined or regulated by environmental factors, incorporation of G-P-E model into MAS programs will be the next challenge, particularly for traits such as abiotic stresses which are influenced by the environment in a large measure.

Many traits of agronomic importance are complex in nature. Improvement of complex traits will depend on our ability to manipulate genes, which have minor effects, and exhibit interaction with each other. GS seems to be a powerful tool for improvement of complex traits, as shown by some simulation studies. However, the conditions and parameters included in simulations may not fully reflect the complex situations of diverse plant breeding programs; the germplasm involved, and the models used in such studies need to be appropriately optimized. The genetic gain per unit time and cost that has been achieved in simulations needs to be supported by the long-term selection response by comparing with other breeding approaches and historical genetic gain. Relative to animal breeding, genomic selection in plants has been taking the advantages of flexible breeding system, short growth period or life span, small plant size and fixed location for growth and development. Increased genetic gain would be expected if these advantages are better utilized along with a set of whole genome strategies in marker-assisted selection for all genes, alleles, haplotypes and their combinations (Xu et al. 2012c). The strategies should include whole genome sequences for the entire germplasm, high density molecular markers covering every gene and all important traits, high precision phenotyping performed for all major traits under multiple environments, and well dissected environmental factors that influence genes, genotypes and phenotypic performance,.

Our future plant breeding faces both challenges and opportunities. These include reducing cost while increasing scale and efficiency, availability of high-throughput genotyping and phenotyping platforms, accessibility and utilization of environmental factors, strong information management and data analysis tools, and decision support system. All these will finally make marker-assisted molecular breeding a routine practice for many breeding programs, especially in the developing world.

References

Ahmadi N, Albar L, Pressoir G, Pinel A, Fargette D, Ghesquiere A (2001) Genetic basis and mapping of the resistance to rice yellow mottle virus. III. Analysis of QTL efficiency in introgressed progenies confirmed the hypothesis of complementary epistasis between two resistance QTL. Theor Appl Genet 103:1084–1092

Akbari M, Wenzl P, Caig V, Carling J, Xia L, Yang S, Uszynski G, Mohler V, Lehmensiek A, Kuchel H, Hayden MJ, Howes N, Sharp P, Vaughan P, Rathnell B, Huttner E, Kilian A (2006) Diversity arrays technology (DArT) for high-throughput profiling of the hexaploid wheat genome. Theor Appl Genet 13:1409–1420

Ashikari M, Sakakibara H, Lin S, Yamamoto T, Takashi T, Nishimura A, Angeles RE, Qian Q, Kitano H, Matsuoka M (2005) Cytokinin oxidase regulates rice grain production. Science 309:741–745

Atwell S, Huang YS, Vilhjálmsson BJ, Willems G, Horton M, Li Y, Meng D, Platt A, Tarone AM, Hu TT, Jiang R, Muliyati NW, Zhang X, Amer MA, Baxter I, Brachi B, Chory J, Dean C, Debieu M, de Meaux J, Ecker JR, Faure N, Kniskern JM, Jones JD, Michael T, Nemri A, Roux F, Salt DE, Tang C, Todesco M, Traw MB, Weigel D, Marjoram P, Borevitz JO, Bergelson J, Nordborg M (2010) Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. Nature 465:627–631

Babar MA, Reynolds MP, van Ginkel M, Klatt AR, Raun WR, Stone ML (2006) Spectral reflectance to estimate genetic variation for in-season biomass, leaf chlorophyll and canopy temperature in wheat. Crop Sci 46:1046–1057

Babar MA, van Ginkel M, Klatt AR, Prasad B, Reynold MP (2007) The potential of using spectral reflectance indices to estimate yield in wheat grown under reduced irrigation. Euphytica 150:155–172

Babu R, Nair SK, Kumar A, Venkatesh S, Sekhar JC, Singh NN, Srinivasan G, Gupta HS (2005) Two-generation marker-aided backcrossing for rapid conversion of normal maize lines to Quality Protein Maize (QPM). Theor Appl Genet 111:888–897

Bagge M, Xia X, Lübberstedt TL (2007) Functional markers in wheat. Curr Opin Plant Biol 10:211–216

Bänziger M, Setimela PS, Hodson D, Vivek B (2006) Breeding for improved abiotic stress tolerance in maize adapted to southern Africa. Agr Water Manag 80:212–224

Barrière Y, Thomas J, Denoue D (2008) QTL mapping for lignin content, lignin monomeric composition, p-hydroxycinnamate content, and cell wall digestibility in the maize recombinant inbred line progeny F838 X F286. Plant Sci 175:585–595

Basavaraj SH, Singh VK, Singh A, Singh A, Singh A, Anand D, Yadav S, Ellur RK, Singh D, Krishnan SG, Nagarajan M, Mohapatra T, Prabhu KV, Singh AK (2010) Marker-assisted improvement of bacterial blight resistance in parental lines of Pusa RH10, a superfine grain aromatic rice hybrid. Mol Breeding 26:293–305

Beavis WD (1998) QTL analyses: power, precision and accuracy. In: Paterson AH (ed) Molecular dissection of complex traits. CRC Press, Boca Raton, pp 145–162

Bernardo R (2008) Molecular markers and selection for complex traits in plants: learning from the last 20 years. Crop Sci 48:1649–1664

Bernardo R (2010) Genomewide selection with minimal crossing in self-pollinated crops. Crop Sci 50:624–627

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 genomewide selection for quantitative traits in maize. Crop Sci 47:1082–1090

Bhatia D, Sharma R, Vikal Y, Mangat GS, Mahajan R, Sharma N, Lore JS, Singh N, Bharaj TS, Singh K (2011) Marker-assisted development of bacterial blight resistant, dwarf, and high yielding versions of two traditional basmati rice cultivars. Crop Sci 51:759–770

Bidinger FR, Serraj R, Rizvi SMH, Howarth C, Yadav RS, Hash CT (2005) Field evaluation of drought tolerance QTL effects on phenotype and adaptation in pearl millet (*Pennisetum glaucum* (L.) R. Br.) top cross hybrids. Field Crops Res 94:14–32

Bonnet DG, Rebetzke GJ, Spielmeyer W (2005) Strategies for efficient implementation of molecular markers in wheat breeding. Mol Breeding 15:75–85

Cakir M, Drake-Brockman F, Shankar M, Golzar H, McLean R, Bariana H, Wilson R, Barclay I, Moore C, Jones M, Loughman R (2008) Molecular mapping and marker-assisted improvement of rust resistance in the Australian wheat germplasm, p 1–3. In: Appels R, Eastwood R, Lagudah E, Langridge P, Mackay M, McIntyre L, Sharp P (eds) Proceedings of 11th International Wheat Genet Symposium, Brisbane Australia, 24–29 August 2008. Sydney University Press, Sydney. http://hdl.handle.net/2123/3317

Casao MC, Igartua E, Karsai I, Bhat PR, Cuadrado N, Gracia MP, Lasa JM, Casas AM (2011) Introgression of an intermediate VRNH1 allele in barley (*Hordeum vulgare* L.) leads to reduced vernalization requirement without affecting freezing tolerance. Mol Breeding 28:475–484

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

Chen S, Lin XH, Xu CG, Zhang Q (2000) Improvement of bacterial blight resistance of 'Minghui63', an elite restorer line of hybrid rice, by molecular marker-assisted selection. Crop Sci 40:239–244

Chen L, Zhao Z, Liu X, Liu L, Jiang L, Liu S, Zhang W, Wang Y, Liu Y, Wan J (2011) Marker-assisted breeding of a photoperiod-sensitive male sterile japonica rice with high cross-compatibility with indica rice. Mol Breeding 27:247–258

Chia JM, Song C, Bradbury PJ, Costich D, de Leon N, Doebley J, Elshire RJ, Gaut B, Geller L, Glaubitz JC, Gore M, Guill KE, Holland J, Hufford MB, Lai J, Li M, Liu X, Lu Y, McCombie R, Nelson R, Poland J, Prasanna BM, Pyhäjärvi T, Rong T, Sekhon RS, Sun Q,Tenaillon MI, Tian F, Wang J, Xu X, Zhang Z, Kaeppler SM, Ross-Ibarra J, McMullen MD, Buckler ES, Zhang G, Xu Y, Ware D (2012) Maize HapMap2 identifies extant variation from a genome in flux. Nat Genet 44:803–807

Close T, Bhat P, Lonardi S, Wu Y, Rostoks N, Ramsay L, Druka A, Stein N, Svensson J, Wanamaker S, Bozdag S, Roose M, Moscou M, Chao S, Varshney R, Szucs P, Sato K, Hayes P, Matthews D, Kleinhofs A, Muehlbauer G, DeYoung J, Marshall D, Madishetty K, Fenton R, Condamine P, Graner A, Waugh R (2009) Development and implementation of high-throughput SNP genotyping in barley. BMC Genomics 10:582

Collins NC, Tardieu F, Tuberosa R (2008) Quantitative trait loci and crop performance under abiotic stress: where do we stand? Plant Physiol 147:469–486

Crosbie TM, Eathington SR, Johnson GR, Edwards M, Reiter R, Stark S, Mohanty RG, Oyervides M, Buehler RE, Walker AK, Dobert R, Delannay X, Pershing JC, Hall MA, Lamkey KR (2006) Plant breeding: past, present, and future, p 3–50. In K.R. Lamkey and M. Lee (eds) Plant breeding: the Arnel R. Hallauer international symposium

Delannay X, McLaren G, Ribaut JM (2012) Fostering molecular breeding in developing countries. Mol Breeding 29:857–873

DePauw RM, Townley-Smith TF, Humphreys G, Knox RE, Clarke FR, Clarke JM (2005) Lillian hard red spring wheat. Can J Plant Sci 85:397–401

DePauw RM, Knox RE, Thomas JB, Smith M, Clarke JM, Clarke FR, McCaig TN, Fernandez MR (2009) Goodeve hard red spring wheat. Can J Plant Sci 89:937–944

Dubcovsky J (2004) Marker-assisted selection in public breeding programs: the wheat experience. Crop Sci 44:1895–1898

Dwivedi SL, Crouch JH, Mackill DJ, Xu Y, Blair MW, Ragot M, Upadhyaya HD, Ortiz R (2007) The molecularization of public sector crop breeding: progress, problems and prospects. Adv Agron 95:163–318

Eathington SR (2005) Practical applications of molecular technology in the development of commercial maize hybrids. In: Proceedings of the 60th Annual Corn and Sorghum Seed Research Conference, Chicago [CD-ROM]. 7–9 Dec 2005. American Seed Trade Association, Washington, DC

Eathington SR, Crosbie TM, Edwards MD, Reiter RS, Bull JK (2007) Molecular markers in a commercial breeding program. Crop Sci 47(S3): S154–S163

Edmeades GO, Bänziger M, Ribaut JM (2000) Maize improvement for drought-limited environments. In: Otegui ME, Slafer GA (eds) Physiological bases for maize improvement. Food Products Press, New York, pp 75–111

Edwards M, Johnson L (1994) RFLPs for rapid recurrent selection. In: Proceedings of Symposium on Analysis of Molecular Marker Data. American Society of Horticultural Science and Crop Science Society of America, Corvallis, Oregon, pp 33–40

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K et al (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE 6(5):e19379

Fleury D, Jefferies S, Kuchel H, Langridge P (2010) Genetic and genomic tools to improve drought tolerance in wheat. J Exp Bot 61:211–3222

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

Friedt W, Ordon F (2007) Molecular markers for gene pyramiding and disease resistance breeding in barley. In: Varshney RK, Tuberosa R (eds) Genomics-Assisted Crop Improvement. Vol.2 Genomics Application in Crops, Springer, pp 81–101

Frisch M (2004) Breeding strategies: optimum design of marker-assisted backcross programs. In: Lörz H, Wenzl G (eds) Biotechnology in agriculture and forestry, vol 55., Molecular marker systems in plant breeding and crop improvementSpringer, Berlin, pp 319–334

Gao S, Martinez C, Skinner DJ, Krivanek AF, Crouch JH, Xu Y (2008) Development of a seed DNA-based genotyping system for marker-assisted selection in maize. Mol Breeding 22:477–494

Gao S, Babu R, Lu Y, Martinez C, Hao Z, Krivanek AF, Wang J, Rong T, Crouch JH, Xu Y (2011) Revisiting the hetero-fertilization phenomenon in maize. PLoS ONE 6(1):e16101

Genc Y, Oldach K, Verbyla A, Lott G, Hassan M, Tester M, Wallwork H, McDonald G (2010) Sodium exclusion QTL associated with improved seedling growth in bread wheat under salinity stress. Theor Appl Genet 121:877–894

Glover JD, Reganold JP, Bell LW, Borevitz J, Brummer EC, Buckler ES, Cox CM, Cox TS, Crews TE, Culman SW, DeHaan LR, Eriksson D, Gill BS, Holland J, Hulke BS, Ibrahim AMH, Jackson W, Jones SS, Murray SC, Paterson AH, Ploschuk E, Sacks EJ, Snapp S, Tao D, Van Tassel DL, Wade LJ, Wyse DL, Xu Y (2010) Increased food and ecosystem security via perennial grains. Science 328:1638–1639

Goddard ME, Hayes BJ (2007) Genomic selection. J Anim Breeding Genet 124:323–330

Gore MA, Chia JM, Elshire RJ, Sun Q, Ersoz ES, Hurwitz BL, Peiffer JA, McMullen MD, Grills GS, Ross-Ibarra J, Ware DH, Buckler ES (2009) A first-generation haplotype map of maize. Science 326:1115–1117

Grewal TS, Rossnagel BG, Scoles GJ (2010) Validation of molecular markers associated with net blotch resistance and their utilization in barley breeding. Crop Sci 50:177–184

Gupta PK, Kumar J, Mir RR, Kumar A (2009) Marker-assisted selection as a component of conventional plant breeding. Plant Breeding Rev 33:145–217

Gupta PK, Langridge P, Mir RR (2010) Marker-assisted wheat breeding: present status and future possibilities. Mol Breeding 26:145–161

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

Harjes CE, Rocheford TR, Bai L, Brutnell TP, Kandianis CB, Sowinski SG, Stapleton AE, Vallabhaneni R, Williams M, Wurtzel ET, Yan J, Buckler ES (2008) Natural genetic variation in Lycopene Epsilon Cyclase tapped for maize biofortification. Science 319:330–333

Hash CT (2005) Opportunities for application of molecular markers for sustainable crop production in stress environments: sorghum and pearl millet. In: International Conference on Sustainable Crop Production in Stress Environments: Management and Genetic Options, pp 113 (abstract). Jawaharlal Nehru Krishi Vishwa Vidyalaya, Jabalpur, India

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

He Y, Chen C, Tu J, Zhou P, Jiang G, Tan Y, Xu C, Zhang Q (2002) Improvement of an elite rice hybrid, Shanyou 63, by transformation and maker-assisted selection. In: Abstracts of the Fourth International Symposium on Hybrid Rice, 14–17 May 2002, Hanoi, Vietnam, p 43

He ZH, Xia XC, Chen XM, Zhang QS (2011) Progress of wheat breeding in China and the future perspective. Acta Agronomica Sinica 37:202–215

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

Hospital F (2009) Challenges for effective marker-assisted selection in plants. Genetica 136:303–310

Hospital F, Charcosset A (1997) Marker-assisted introgression of quantitative trait loci. Genetics 147:1469–1485

Houle D, Govindaraju DR, Omholt S (2010) Phenomics: the next challenge. Nat Rev Genet 11:855–866

Huang X, Feng Q, Qian Q, Zhao Q, Wang L, Wang A, Guan J, Fan D, Wang Q, Huang T, Dong G, Sang T, Han B (2009) High-throughput genotyping by whole-genome resequencing. Genome Res 19:1068–1076

Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z, Li M, Fan D, Guo Y, Wang A, Wang L, Deng L, Li W, Lu Y, Weng Q, Liu K, Huang T, Zhou T, Jing Y, Li W, Lin Z (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. Nat Genet 42:961–967

Iftekharuddaula KM, Newaz MA, Salam MA, Ahmed HU, Mahbub MAA, Septiningsih EM, Collard BCY, Sanchez DL, Pamplona AM, Mackill DJ (2011) Rapid and high-precision marker assisted backcrossing to introgress the SUB1 QTL into BR11, the rainfed lowland rice mega variety of Bangladesh. Euphytica 178:83–97

International HapMap 3 Consortium (2010) Integrating common and rare genetic variation in diverse human populations. Nature 467:52–58

James RA, Davenport RJ, Munns R (2006) Physiological characterization of two genes for Na+ exclusion in durum wheat, *Nax1* and *Nax2*. Plant Physiol 142:1537–1547

Jannink JL, Walsh B (2002) Association mapping in plant populations. In: Kang MS (ed) Quantitative Genetics, Genomics and Plant Breeding. CAB International, Wallingford, UK, pp 59–68

Jena KK, Mackill DJ (2008) Molecular markers and their use in marker-assisted selection in rice. Crop Sci 48:1266–1276

Jaccoud D, Peng K, Feinstein D, Kilian A (2001) Diversity arrays: a solid state technology for sequence information independent genotyping. Nucl Acids Res 29:e25

Joseph M, Gopalakrishnan S, Sharma RK, Singh VP, Singh AK, Singh NK, Mohapatra T (2004) Combining bacterial blight resistance and basmati quality characteristics by phenotypic and molecular marker assisted selection in rice. Mol Breeding 13:377–387

Khanduri A, Tiwari A, Prasanna BM, Hossain F, Kumar R, Prakash O, Singh SB (2010) Conversion of elite maize lines in India into QPM versions using an integrated phenotypic and molecular marker-assisted selection strategy. In: Zaidi PH, Azrai M, Pixley KV (eds) Maize for Asia: emerging Trends and Technologies. Proceeding of The 10th Asian Regional Maize Workshop, Makassar, Indonesia, 20–23 Oct 2008. CIMMYT, Mexico D.F., p 233–236

Kim S, Plagnol V, Hu TT, Toomajian C, Clark RM, Ossowski S, Ecker JR, Weigel D, Nordborg M (2007) Recombination and linkage disequilibrium in *Arabidopsis thaliana*. Nat Genet 39:1151–1155

Knoll JE, Ejeta G (2008) Marker-assisted selection for early season cold tolerance in sorghum: QTL validation across populations and environments. Theor Appl Genet 116:541–553

Kuchel H, Fox R, Hollamby G, Reinheimer JL, Jefferies SP (2008) The challenges of integrating new technologies into a wheat breeding programme, p 1–5. In: Appels R, Eastwood R, Lagudah E, Langridge P, Mackay M, McIntyre L, Sharp P (eds) Proceedings of 11th International Wheat Genet Symposium. Brisbane, 24–29 Aug 2008. Sydney University Press. http://hdl.handle.net/2123/3400

Kumar J, Mir RR, Kumar N, Kumar A, Mohan A, Prabhu KV, Balyan HS, Gupta PK (2010) Marker assisted selection for pre-harvest sprouting tolerance and leaf rust resistance in bread wheat. Plant Breeding 12:617–621

Kump KL, Bradbury PJ, Wisser RJ, Buckler ES, Belcher AR, Oropeza-Rosas MA, Zwonitzer JC, Kresovich S, McMullen MD, Ware D, Balint-Kurti PJ, Holland JB (2011) Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. Nat Genet 43:163–168

Lagudah ES, Krattinger SG, Herrera-Foessel S, Singh RP, Huerta-Espino J, Spielmeyer W, Brown-Guedira G, Selter LL, Keller B (2009) Gene-specific markers for the wheat gene Lr34/Yr18/Pm38 which confers resistance to multiple fungal pathogens. Theor Appl Genet 119:889–898

Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 124:743–756

Lander ES, Botstein D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185–199

Langridge P, Paltridge N, Fincher G (2006) Functional genomics of abiotic stress tolerance in cereals. Briefings Funct Genomics Proteomics 4:343–354

Lebowitz RL, Soller M, Beckmann JS (1987) Trait-based analysis for the detection of linkage between marker loci and quantitative trait loci in cross between inbred lines. Theor Appl Genet 73:556–562

Lee M (1995) DNA markers and plant breeding programs. Adv Agron 55:265–344

Li Y, Wang JK, Qiu LJ, Ma YZ, Li XH, Wan JM (2010) Crop molecular breeding in China: current status and perspectives. Acta Agron Sinica 36:1425–1430

Li JZ, Zhang ZW, Li YL, Wang QL, Zhou YG (2011) QTL consistency and meta-analysis for grain yield components in three generations in maize. Theor Appl Genet 122:771–782

Liang F, Deng Q, Wang Y, Xiong Y, Jin D, Li J, Wang B (2004) Molecular marker-assisted selection for yield-enhancing genes in the progeny of '9311 × *O. rufipogon*' using SSR. Euphytica 139:159–165

Lindsay MP, Lagudah ES, Hare RA, Munns R (2004) A locus for sodium exclusion (*Nax1*), a trait for salt tolerance, mapped in durum wheat. Funct Plant Biol 31:1105–1114

Liu Y, He Z, Appels R, Xia X (2012) Functional markers in wheat: current status and future prospects. Theor Appl Genet 125:1–10

Löffler M, Schön CC, Miedaner T (2009) Revealing the genetic architecture of FHB resistance in hexaploid wheat (*Triticum aestivum* L.) by QTL meta-analysis. Mol Breeding 23:473–488

Lorenzana RE, Bernardo R (2009) Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. Theor Appl Genet 120:151–161

Lu Y, Zhang SH, Shah T, Xie C, Hao Z, Li X, Farkhari M, Ribaut JM, Cao M, Rong T, Xu Y (2010) Joint linkage–linkage disequilibrium mapping is a powerful approach to detecting quantitative trait loci underlying drought tolerance in maize. Proc Natl Acad Sci USA 107:19585–19590

Lu Y, Xu J, Yuan Z, Hao Z, Xie C, Li X, Shah T, Lan H, Zhang S, Rong T, Xu Y (2012) Comparative LD mapping using single SNPs and haplotypes identifies QTL for plant height and biomass as secondary traits of drought tolerance in maize. Mol Breeding 30:407–418

Mace ES, Xia L, Jordan DR, Halloran K, Parh DK, Huttner E, Wenzl P, Kilian A (2008) DArT markers: diversity analyses and mapping in *Sorghum bicolor*. BMC Genomics 9:26

Mace ES, Rami JF, Bouchet S, Klein PE, Klein RP, Kilian A, Wenzl P, Xia L, Halloran K, Jordan DR (2009) A consensus genetic map of sorghum that integrates multiple component maps and high-throughput Diversity Array Technology (DArT) markers. BMC Plant Biol 9:13

Mackay I, Powell W (2007) Methods for linkage disequilibrium mapping in crops. Trends Plant Sci 12:57–63

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 Breeding 27:329–335

Mammadov J, Chen W, Mingus J, Thompson S, Kumpatla S (2012) Development of versatile gene-based SNP assays in maize (*Zea mays* L.). Mol Breeding 29:779–790

Manenti G, Galvan A, Pettinicchio A, Trincucci G, Spada E, Zolin A, Milani S, Gonzalez-Neira A, Dragani TA (2009) Mouse genome-wide association mapping needs linkage analysis to avoid false-positive loci. PLoS Genet 5:e1000331

McCouch SR, Sweeney M, Li J, Jiang H, Thomson M, Septiningsih E, Edwards J, Moncada P, Xiao J, Garris A, Tai T, Martinez C, Tohme J, Sugiono M, McClung A, Yuan LP, Ahn SN (2007) Through the genetic bottleneck: *O. rufipogon* as a source of trait-enhancing alleles for *O. sativa*. Euphytica 154:317–339

McCouch SR, Zhao K, Wright M, Tung CW, Ebana K, Thomson M, Reynolds A, Wang D, DeClerck G, Ali ML, McClung A, Eizenga G, Bustamante C (2010) Development of genome-wide SNP assays for rice. Breeding Sci 60:524–535

McNally K, Childs K, Bohnert R, Davidson R, Zhao K, Ulat V, Zeller G, Clark R, Hoen D, Bureau T, Stokowski R, Ballinger D, Frazer K, Cox D, Padhukasahasram B, Bustamante C, Weigel D, Mackill D, Bruskiewich R, Rätsch G, Buell C, Leung H, Leach J (2009) Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. Proc Natl Acad Sci USA 106:12273–12278

Meuwissen TH (2009) Accuracy of breeding values of 'unrelated' individuals predicted by dense SNP genotyping. Genet Sel Evol 41:35

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genomewide dense marker maps. Genetics 157:1819–1829

Miura K, Ashikari M, Matsuoka M (2011) The role of QTLs in the breeding of high-yielding rice. Trends in Plant Science 16:319–326

Myles S, Peiffer J, Brown PJ, Ersoz ES, Zhang Z, Costich DE, Buckler ES (2009) Association mapping: critical considerations shift from genotyping to experimental design. Plant Cell 21:2194–2202

Nguyen B, Brar D, Bui B, Nguyen T, Pham L, Nguyen H (2003) Identification and mapping of the QTL for aluminum tolerance introgressed from the new source, *Oryza rufipogon* Griff., into indica rice (*Oryza sativa* L.). Theor Appl Genet 106:583–593

Peleman JD, van der Voort JR (2003) Breeding by design. Trends Plant Sci 8:330–334

Podlich DW, Winkler CR, Cooper M (2004) Mapping as you go: an effective approach for marker-assisted selection of complex traits. Crop Sci 44:560–1571

Prasanna BM, Pixley K, Warburton ML, Xie CX (2010) Molecular marker-assisted breeding options for maize improvement in Asia. Mol Breeding 26:339–356

Qiu LJ, Guo Y, Li Y, Wang XB, Zhou GA, Liu ZX, Zhou SR, Li XH, Ma YZ, Wang JK, Wan JM (2011) Novel gene discovery of crops in China: status, challenging, and perspective. Acta Agron Sinica 37:1–17

Rae SJ, Macaulay M, Ramsay L, Leigh F, Mathews D, O'Sullivan DM, Donini P, Morris PC, Powell W, Marshall DF, Waugh R, Thomas WTB (2007) Molecular barley breeding. Euphytica 158:295–303

Ragot M, Lee M (2007) Marker-assisted selection in maize: current status, potential, limitations and perspectives from the private and public sectors. In: Guimarães EP et al (eds) Marker-assisted selection, current status and future perspectives in crops, livestock, forestry, and fish. FAO, Rome, pp 117–150

Ragot M, Gay D, Muller JP, Durovray J (2000) Efficient selection for the adaptation to the environment through QTL mapping and manipulation in maize. In: Ribaut JM, Poland D (eds) Molecular approaches for the genetic improvement of cereals for stable production in water-limited environments. CIMMYT, Mexico, D.F, pp 128–130

Ramlingam J, Basharat HS, Zhang G (2002) STS and microsatellite marker-assisted selection for bacterial blight resistance and waxy gene in rice, *Oryza sativa* L. Euphytica 127:255–260

Ribaut JM, Ragot M (2007) Marker-assisted selection to improve drought adaptations in maize: the backcross approach, perspectives, limitations and alternatives. J Exp Bot 58:351–360

Ribaut JM, Bänziger M, Betran J, Jiang C, Edmeades GO, Dreher K, Hoisington D (2002) Use of molecular markers in plant breeding: drought tolerance improvement in tropical maize.

In: Kang MS (ed) Quantitative Genetics. Genomics and Plant Breeding CAB International, Wallingford, pp 85–99

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

Roy SJ, Tucker EJ, Tester M (2011) Genetic analysis of abiotic stress tolerance in crops. Curr Opin Plant Biol 14:1–8

Rutkoski JE, Heffner EL, Sorrells ME (2011) Genomic selection for durable stem rust resistance in wheat. Euphytica 179:161–173

Salameh A, Buerstmayr M, Steiner B, Neumayer A, Lemmens M, Buerstmayr H (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 Breeding 28:485–494

Schmierer DA, Kandemir N, Kudrna DA, Jones BL, Ullrich SE, Kleinhofs A (2004) Molecular marker-assisted selection for enhanced yield in malting barley. Mol Breeding 14:463–473

Septiningsih EM, Pamplona AM, Sanchez DL, Neeraja CN, Vergara GV, Heuer S, Ismail AM, Mackill DJ (2009) Development of submergence tolerant rice cultivars: the Sub1 locus and beyond. Ann Bot 103:151–160

Servin B, Martin OC, Mézard M, Hospital F (2004) Toward a theory of marker-assisted gene pyramiding. Genetics 168:513–523

Shen L, Courtois B, McNally KL, Robin S, Li Z (2001) Evaluation of near-isogenic lines of rice introgressed with QTLs for root depth through marker-aided selection. Theor Appl Genet 103:75–83

Siangliw M, Toojinda T, Tragoonrung S, Vanavichit A (2003) Thai Jasmine rice carrying QTLch9 (*Sub*QTL) is submergence tolerant. Ann Bot 91:255–261

Sorrells ME (2007) Application of new knowledge, technologies, and strategies to wheat improvement. Euphytica 157:299–306

Stam P (1995) Marker-assisted breeding. In: Van Ooijen JW and Jansen J (eds) Biometrics in plant breeding: applications of molecular markers. Proceedings of the 9th meeting of EUCARPIA section on biometrics in plant breeding (1994). Centre for Plant Breeding and Reproduction Research, Wageningen, Netherlands, pp 32–44

Stam P (2003) Marker-assisted introgression: speed at any cost? In: van Hintum Th.JL, Lebeda A, Pink D, Schut JW (eds) Proceedings of the Eucarpia meeting on leafy vegetables genetics and breeding, 19–21 March 2003, Noordwijkerhout, Netherlands. Centre for Genetic Resources (CGN), Wageningen, Netherlands, pp 117–124

Steele KA, Price AH, Shashidhar HE, Witcombe JR (2006) Marker-assisted selection to introgress rice QTL controlling root traits into an Indian upland rice variety. Theor Appl Genet 112:208–221

Sun Y, Wang J, Crouch JH, Xu Y (2010) Efficiency of selective genotyping for genetic analysis of complex traits and potential applications in crop improvement. Mol Breeding 26:493–511

Sundaram RM, Vishnupriya MR, Laha GS, Shobha Rani N, Srinivas Rao P, Balachandaran SM, Asho Reddy G, Sarma NP, Shonti RV (2009) Introduction of bacterial blight resistance into Triguna, a high yielding, mid-early duration rice variety. Biotechnol J 4:400–407

Tian F, Bradbury PJ, Brown PJ, Hung H, Sun Q, Flint-Garcia S, Rocheford TR, McMullen MD, Holland JB, Buckler ES (2011) Genome-wide association study of leaf architecture in the maize nested association mapping population. Nat Genet 43:159–162

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

Trebbi D, Maccaferri M, de Heer P, Sørensen A, Giuliani S, Salvi S, Sanguineti MC, Massi A, van der Vossen EAG, Tuberosa R (2011) High-throughput SNP discovery and genotyping in durum wheat (*Triticum durum* Desf.). Theor Appl Genet 123:555–569

Truntzler M, Barrière Y, Sawkins MC, Lespinasse D, Betran J, Charcosset A, Moreau L (2010) Meta-analysis of QTL involved in silage quality of maize and comparison with the position of candidate genes. Theor Appl Genet 121:1465–1482

van Berloo R, Stam P (1998) Marker-assisted selection in autogamous RIL populations: a simulation study. Theor Appl Genet 96:147–154
van Berloo R, Stam P (2001) Simultaneous marker-assisted selection for multiple traits in autogamous crops. Theor Appl Genet 102:1107–1112
Van Damme V, Gómez-Paniagua H, de Vicente MC (2011) The GCP molecular marker toolkit, an instrument for use in breeding food security crops. Mol Breeding 28:597–610
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
Wang YH, Liu SJ, Ji SL, Zhang WW, Wang CM, Jiang L, Wan JM (2005) Fine mapping and marker-assisted selection (MAS) of a low glutelin content gene in rice. Cell Res 15:622–630
Wang J, Chapman SC, Bonnett DG, Rebetzke GJ, Crouch J (2007) Application of population genetic theory and simulation models to efficiently pyramid multiple genes via marker-assisted selection. Crop Sci 47:582–588
Wang CL, Zhang YD, Zhu Z, Chen T, Zhao L, Lin J, Zhou LH (2009) Development of a new *japonica* rice variety Nanjing 46 with good eating quality by marker assisted selection. Mol Plant Breeding 7:1070–1076
Wang C, Chen S, Yu S (2011a) Functional markers developed from multiple loci in GS3 for fine marker-assisted selection of grain length in rice. Theor Appl Genet 122:905–913
Wang L, Wang A, Huang X, Zhao Q, Dong G, Qian Q, Sang T, Han B (2011b) Mapping 49 quantitative trait loci at high resolution through sequencing-based genotyping of rice recombinant inbred lines. Theor Appl Genet 122:327–340
Wei X, Jin Liu L L, Xu JF, Jiang L, Zhang WW, Wang JK, Zhai HQ, Wan JM (2009) Breeding strategies for optimum heading date using genotypic information in rice. Mol Breeding 25:287–298
Wenzl P, Li H, Carling J, Zhou M, Raman H, Paul E, Hearnden P, Maier C, Xia L, Caig V, Ovesná J, Cakir M, Poulsen D, Wang J, Raman R, Smith KP, Muehlbauer GJ, Chalmers KJ, Kleinhofs A, Huttner E, Kilian A (2006) A high-density consensus map of barley linking DArT markers to SSR and RFLP loci and agronomic traits. BMC Genomics 7:206–228
Werner K, Friedt W, Ordon F (2005) Strategies for pyramiding resistance genes against the barley yellow mosaic virus complex (BaMMV, BaYMV, BaYMV-2). Mol Breeding 16:45–55
William HM, Trethowan R, Crosby-Galvan EM (2007) Wheat breeding assisted by markers: CIMMYT's experience. Euphytica 157:307–319
Witcombe JR, Hash CT (2000) Resistance gene deployment strategies in cereal hybrids using marker-assisted selection: gene pyramiding, three-way hybrids and synthetic parent populations. Euphytica 112:175–186
Wong CK, 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
Xie W, Feng Q, Yu H, Huang X, Zhao Q, Xing Y, Yu S, Han B, Zhang Q (2010) Parent-independent genotyping for constructing an ultrahigh-density linkage map based on population sequencing. Proc Natl Acad Sci USA 107:10578–10583
Xu Y (1997) Quantitative trait loci: separating, pyramiding, and cloning. Plant Breeding Rev 15:85–139
Xu Y (2002) Global view of QTL: rice as a model. In: Kang MS (ed) Quantitative genetics, genomics and plant breeding. CABI Publishing, Wallingford, pp 109–134
Xu Y (2003) Developing marker-assisted selection strategies for breeding hybrid rice. Plant Breeding Rev 23:73–174
Xu Y (2010) Molecular plant breeding. CAB International, Wallingford, p 734
Xu Y, Crouch JH (2008) Marker-assisted selection in plant breeding: from publications to practice. Crop Sci 48:391–407
Xu K, Xia X, Fukao T, Canlas P, Maghirang-Rodriguez R, Heuer S, Ismail AI, Bailey-Serres J, Ronald PC, Mackill DJ (2006) Sub1A is an ethylene response factor-like gene that confers submergence tolerance to rice. Nature 442:705–708
Xu J, Liu Y, Liu J, Cao M, Wang J, Lan H, Xu Y, Lu Y, Guangtang Pan G, Rong T (2012a) The genetic architecture of flowering time and photoperiod sensitivity in maize as revealed by QTL review and meta analysis. J Integr Plant Biol 54:358–373

Xu X, Xin Liu X, Ge S, Jensen JD, Hu F, Li X, Dong Y, Gutenkunst RN, Fang L, Huang L, Li J, He W, Zhang G, Zheng X, Zhang F, Li Y, Yu C, Kristiansen K, Zhang X, Wang J, Wright M, McCouch S, Nielsen R, Wang J, Wang W (2012b) Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. Nat Biotech 30:105–111

Xu Y, Lu Y, Xie C, Gao S, Wan J, Prasanna BM (2012c) Whole genome strategies for marker-assisted plant breeding. Mol Breeding 29:833–854

Yamamoto T, Nagasaki H, Yonemaru J, Ebana K, Nakajima M, Shibaya T, Yano M (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 J, Shah T, Warburton ML, Buckler ES, McMullen MD, Crouch J (2009) Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. PLoS ONE 4:e8451

Yan J, Yang X, Shah T, Sánchez-Villeda H, Li J, Warburton M, Zhou Y, Crouch JH, Xu Y (2010a) High-throughput SNP genotyping with the GoldenGate assay in maize. Mol Breeding 25:441–451

Yan J, Kandianis CB, Harjes CE, Bai L, Kim E, Yang X, Skinner D, Fu Z, Mitchell S, Li Q, Fernandez MGS, Zaharieva M, Babu R, Fu Y, Palacios N, Li J, DellaPenna D, Brutnell T, Buckler ES, Warburton ML, Rocheford T (2010b) Rare genetic variation at *Zea mays crtRB1* increases β-carotene in maize grain. Nat Genet 42:322–327

Yoon DB, Kang KH, Kim HJ, Ju HG, Kwon SJ, Suh JP, Jeong OY, Ahu SN (2006) Mapping quantitative trait loci for yield components and morphological traits in an advanced back-cross population between *Oryza grandiglumis* and the *O. japonica* cultivar Hwaseongbyeo. Theor Appl Genet 112:1052–1062

Yu J, Hollan JB, McMullen MD, Buckler ES (2008) Genetic design and statistical power of nested association mapping in maize. Genetics 178:539–551

Zhong SQ, Dekkers JCM, 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

Zhou PH, Tan YF, He YA, Xu CG, Zhang A (2003) Simultaneous improvement of four quality traits of Zhenshan 97, an elite parent of hybrid rice, by molecular marker-assisted selection. Theor Appl Genet 106:326–331

Appendix I
Contributors

EDUARD AKHUNOV

Department of Plant Pathology, Kansas State University, Manhattan KS 66502 USA

ALINA AKHUNOVA

Integrated Genomics Facility, Kansas State University, Manhattan KS 66502 USA

VINDHYA AMARASINGHE

Department of Botany and Plant Pathology, 2082 Cordley Hall, Oregon State University, Corvallis, OR, 97331-2902, USA

JOSE L. ARAUS

Unitat de Fisiologia Vegetal, Facultat de Biologia, Universitat de Barcelona, Barcelona, Spain; International Maize and Wheat Improvement Center (CIMMYT), Km 45 Carretera Mexico-Veracruz, Texcoco, Mexico 56130, Mexico

JAN P. BUCHMANN

Institute of Plant Biology, University of Zurich, Zollikerstrasse 107, CH-8008 Zürich, Switzerland

JILL E. CAIRNS

International Maize and Wheat Improvement Center (CIMMYT), P.O. Box MP 163, Mount Pleasant, Harare, Zimbabwe

JOSE CROSSA

International Maize and Wheat Improvement Center (CIMMYT), Km 45 Carretera Mexico-Veracruz, Texcoco, Mexico 56130, Mexico

P. K. Gupta and R. K. Varshney (eds.), *Cereal Genomics II*,
DOI: 10.1007/978-94-007-6401-9,

BISWANATH DAS

International Maize and Wheat Improvement Center (CIMMYT), ICRAF House, United Nations Avenue, Gigiri, Nairobi 00621, Kenya

PALITHA DHARMAWARDHANA

Department of Botany and Plant Pathology, 2082 Cordley Hall, Oregon State University, Corvallis, OR, 97331-2902, USA

DAVID EDWARDS

University of Queensland, Brisbane, QLD, 4072, Australia

JUSTIN ELSER

Department of Botany and Plant Pathology, 2082 Cordley Hall, Oregon State University, Corvallis, OR, 97331-2902, USA

BIKRAM S. GILL

Wheat Genetic and Genomic Resources Center and Department of Plant Pathology, Throckmorton Plant Sciences Center, Kansas State University, Manhattan, KS 66506-5502, USA; King Abdulaziz University, Faculty of Science, Genomics and Biotechnology Section, Department of Biological Sciences, Jeddah 21589 Saudi Arabia

PUSHPENDRA K. GUPTA

Department of Genetics and Plant Breeding, CCS University, Meerut 250004, India

ZHONGHU HE

Institute of Crop Sciences/International Maize and Wheat Improvement Center (CIMMYT), Chinese Academy of Agricultural Sciences, 12 South Zhongguancun St., Beijing 100081, China

PANKAJ JAISWAL

Department of Botany and Plant Pathology, 2082 Cordley Hall, Oregon State University, Corvallis, OR, 97331-2902, USA

BEAT KELLER

Institute of Plant Biology, University of Zurich, Zollikerstrasse 107, CH-8008 Zürich, Switzerland

SEIFOLLAH KIANI

Department of Plant Pathology, Kansas State University, Manhattan KS 66502 USA

NATALIYA KOVALCHUK

Australian Centre for Plant Functional Genomics (ACPFG), University of Adelaide, Waite Campus, Urrbrae, South Australia

PAWAN L. KULWAL

State Level Biotechnology Centre, Mahatma Phule Agricultural University, Rahuri, Ahmednagar (MS), 413722, India

BARBARA LADDOMADA

Wheat Genetic and Genomic Resources Center and Department of Plant Pathology, Throckmorton Plant Sciences Center, Kansas State University, Manhattan, KS 66506-5502, USA; Istituto di Scienze delle Produzioni Alimentari, Consiglio Nazionale delle Ricerche (ISPA-CNR), Via Monteroni 73100 Lecce, Italy

PETER LANGRIDGE

Australian Centre for Plant Functional Genomics (ACPFG), University of Adelaide, Waite Campus, Urrbrae, South Australia

JIAYANG LI

State Key Laboratory of Plant Genomics and National Center for Plant Gene Research, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China

MING LI

Australian Centre for Plant Functional Genomics (ACPFG), University of Adelaide, Waite Campus, Urrbrae, South Australia

SERGIY LOPATO

Australian Centre for Plant Functional Genomics (ACPFG), University of Adelaide, Waite Campus, Urrbrae, South Australia

COSMOS MAGOROKOSHO

International Maize and Wheat Improvement Center (CIMMYT), P.O. Box MP 163, Mount Pleasant, Harare, Zimbabwe

REYAZUL R. MIR

Division of Plant Breeding & Genetics, Shere-Kashmir University of Agricultural Sciences & Technology of Jammu (SKUAST-J),Chatha-180 009, Jammu, India

NATALIA PALACIOS

International Maize and Wheat Improvement Center (CIMMYT), Km 45 Carretera Mexico-Veracruz, Texcoco, Mexico 56130, Mexico

ANDREW H. PATERSON

Plant Genome Mapping Laboratory, University of Georgia, Athens, GA 30602, USA

BODDUPALLI M. PRASANNA

International Maize and Wheat Improvement Center (CIMMYT), ICRAF House, United Nations Avenue, Gigiri, Nairobi 00621, Kenya

NIDHI RAWAT

Wheat Genetic and Genomic Resources Center and Department of Plant Pathology, Throckmorton Plant Sciences Center, Kansas State University, Manhattan, KS 66506-5502, USA

SACHIN RUSTGI

Washington State University, Pullman- 99164, W.A., USA

TAO SANG

State Key Laboratory of Systematic and Evolutionary Botany, Key Laboratory of Plant Resources and Beijing Botanical Garden, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China

NESE SREENIVASULU

Leibniz-Institute of Plant Genetics and Crop Plant Research (IPK), Interdiciplinary Center for Crop Plant Research (IZN) Research Group Stress Genomics, Corrensstraße 3, 06466 Gatersleben, Germany

PALAKOLANU S. REDDY

Leibniz-Institute of Plant Genetics and Crop Plant Research (IPK), Interdiciplinary Center for Crop Plant Research (IZN) Research Group Stress Genomics, Corrensstraße 3, 06466 Gatersleben, Germany

RAJEEV K. VARSHNEY

Center of Excellence in Genomics (CEG), International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru 502324, India; CGIAR Generation Challenge Programme (GCP), c/o CIMMYT, 06600 Mexico DF, Mexico

JIANMIN WAN

Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, 12 South Zhongguancun St., Beijing 100081, China

XI-YIN WANG

Plant Genome Mapping Laboratory, University of Georgia, Athens, GA 30602, USA; Center for Genomics and Computational Biology, School of Life Sciences, and School of Sciences, Hebei United University, Tangshan, Hebei 063009, China

THOMAS WICKER

Institute of Plant Biology, University of Zurich, Zollikerstrasse 107, CH-8008 Zürich, Switzerland

YUNBIXU

Institute of Crop Sciences/International Maize and Wheat Improvement Center (CIMMYT), Chinese Academy of Agricultural Sciences, 12 South Zhongguancun St., Beijing 100081, China

Appendix II
Reviewers

Andrew K. Borrell, University of Queensland, Brisbane, Australia
Dave Edwards, University of Queensland, Brisbane, Australia
Bikram Gill, Kansas State University, Manhattan, USA
Mukesh Jain, National Institute of Plant Genome Research, New Delhi, India
Takao Komatsuda, National Institute of Crop Science, Ibaraki, Japan
Vasudev Kumanduri, European Bioinformatics Institute, Cambridge, UK
Jianxin Ma, Purdue University, Indiana, USA
Reyazul R. Mir, Shere-Kashmir University of Agricultural Sciences & Technology of Jammu (SKUAST-J), Jammu, India
Eviatar Nevo, University of Haifa, Haifa, Israel
Manish Pandey, International Crops Research Institute for the Semi-Arid Tropics, Hyderabad, India
Saurabh Raghuvanshi, Delhi University – South Campus, Delhi, India
Heike U. Schneider, Forschungszentrum Jülich, Jülich, Germany
Trushar Shah, International Crops Research Institute for the Semi-Arid Tropics, Hyderabad, India
Nils Stein, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany
Mahendar Thudi, International Crops Research Institute for the Semi-Arid Tropics, Hyderabad, India
Roberto Tuberosa, University of Bologna, Bologna, Italy
Robbie Waugh, The James Hutton Institute, Dundee, UK

P. K. Gupta and R. K. Varshney (eds.), *Cereal Genomics II*,
DOI: 10.1007/978-94-007-6401-9,

Index

Note: Page numbers followed by *f* or *t* indicate figure or table, respectively

P. K. Gupta and R. K. Varshney (eds.), *Cereal Genomics II*,
DOI: 10.1007/978-94-007-6401-9,