International Journal of Plant Genomics

Genomics of Major Crops and Model Plant Species

Guest Editors: P. K. Gupta and Yunbi Xu

Genomics of Major Crops and Model Plant Species

International Journal of Plant Genomics

Genomics of Major Crops and Model Plant Species

Guest Editors: P. K. Gupta and Yunbi Xu

Contents

Genomics of Major Crops and Model Plant Species, P. K. Gupta and Yunbi Xu
Volume 2008, Article ID 171928, 2 pages

Wheat Genomics: Present Status and Future Prospects, P. K. Gupta, R. R. Mir, A. Mohan, and J. Kumar
Volume 2008, Article ID 896451, 36 pages

Barley Genomics: An Overview, Nese Sreenivasulu, Andreas Graner, and Ulrich Wobus
Volume 2008, Article ID 486258, 13 pages

Development in Rice Genome Research Based on Accurate Genome Sequence, Takashi Matsumoto, Jianzhong Wu, Baltazar A. Antonio, and Takuji Sasaki
Volume 2008, Article ID 348621, 9 pages

Rice Molecular Breeding Laboratories in the Genomics Era: Current Status and Future Considerations, Bert C. Y. Collard, Casiana M. Vera Cruz, Kenneth L. McNally, Parminder S. Virk, and David J. Mackill
Volume 2008, Article ID 524847, 25 pages

Genomics of Sorghum, Andrew H. Paterson
Volume 2008, Article ID 362451, 6 pages

Brachypodium Genomics, Bahar Sogutmaz Ozdemir, Pilar Hernandez, Ertugrul Filiz, and Hikmet Budak
Volume 2008, Article ID 536104, 7 pages

Sugarcane Functional Genomics: Gene Discovery for Agronomic Trait Development, M. Menossi, M. C. Silva-Filho, M. Vincentz, M.-A. Van-Sluys, and G. M. Souza
Volume 2008, Article ID 458732, 11 pages

Soybean Genomics: Developments through the Use of Cultivar "Forrest", David A. Lightfoot
Volume 2008, Article ID 793158, 22 pages

Recent Advances in *Medicago truncatula* Genomics, Jean-Michel Ané, Hongyan Zhu, and Julia Frugoli
Volume 2008, Article ID 256597, 11 pages

Progress in Understanding and Sequencing the Genome of *Brassica rapa*, Chang Pyo Hong, Soo-Jin Kwon, Jung Sun Kim, Tae-Jin Yang, Beom-Seok Park, and Yong Pyo Lim
Volume 2008, Article ID 582837, 9 pages

Recent Advances in Cotton Genomics, Hong-Bin Zhang, Yaning Li, Baohua Wang, and Peng W. Chee
Volume 2008, Article ID 742304, 20 pages

Structural and Functional Genomics of Tomato, Amalia Barone, Maria Luisa Chiusano, Maria Raffaella Ercolano, Giovanni Giuliano, Silvana Grandillo, and Luigi Frusciante
Volume 2008, Article ID 820274, 12 pages

Genomic Resources and Tools for Gene Function Analysis in Potato, Glenn J. Bryan and Ingo Hein
Volume 2008, Article ID 216513, 9 pages

Citrus Genomics, Manuel Talon and Fred G. Gmitter Jr.
Volume 2008, Article ID 528361, 17 pages

Hindawi Publishing Corporation
International Journal of Plant Genomics
Volume 2008, Article ID 171928, 2 pages
doi:10.1155/2008/171928

Editorial

Genomics of Major Crops and Model Plant Species

P. K. Gupta[1] and Yunbi Xu[2]

[1] *Molecular Biology Laboratory, Department of Genetics and Plant Breeding, Chaudhary Charan Singh University, Meerut, 250 004, India*
[2] *International Maize and Wheat Improvement Center (CIMMYT), Apartado Postal 6-64, 06600 Mexico, DF, Mexico*

Correspondence should be addressed to P. K. Gupta, pkgupta36@gmail.com

Received 24 December 2008; Accepted 24 December 2008

Plant genomics research had its beginning in December 2000, with the publication of the whole genome sequence of the model plant species *Arabidopsis thaliana*. Rapid progress has since been made in this area. The significant developments include the publication of a high-quality rice genome sequence in August 2005, draft genome of poplar in September 2006, whole genome sequence of two grapevine genotypes in 2007, and that of transgenic papaya in 2008. Draft sequences of corn gene-space and those of the genomes of *Lotus japonicus* and *Glycine max* have also become available in 2008. Genomes of several other plant species (e.g., *Sorghum bicolor, Manihot esculenta* (cassava), barley, wheat, potato, cotton, tomato, maize, *Brachypodium distachyon* (a small model grass genome), *Medicago truncatula*, shepherd's purse, peach) are also currently being sequenced. Multinational genome projects on *Brassica* and Solanaceous genomes are also in progress. In still other cases (e.g., wheat, corn, barley), where the large genome size prohibits whole genome sequencing, the gene rich regions (GRRs) of the genomes are being identified to bring down the sequencing work to a manageable level. The 10-year-old US National Plant Genome Initiative (NPGI) also made a call for more plant genomes to be sequenced. While making a choice for additional plant genomes to be sequenced, it has also been emphasized that much of plant diversity is available in tropical plants so that during the next decade, more genomes from tropics (e.g., *Carica, Saccharum, Psychoria, Opuntia*) need to be sequenced.

The sequencing information obtained as above will be utilized for both basic and applied research so that while this will help in elucidating evolutionary relationships and developing better phylogenetic classification, this will also help in the discovery of new genes, allele-mining, and large-scale SNP genotyping. In order to achieve these objectives, there has also been a call for sequencing genomes of diverse cultivars of each crop like rice. As a result, the concept of plant pan genome (initially developed for microbial genomes), each composed of "core genome" and "dispensable genome," has also been introduced. The sequence information from diverse cultivars in a crop will be utilized for molecular breeding. For instance, new technologies have been used for the improvement of *indica* rice, but similar efforts are now being made for improvement of *japonica* rice also. An overview of the present status of plant genomics research and its impact is also available in a recent special issue of Science (April 25, 2008).

The future plant genomics research will certainly derive benefit from the recent development of new-generation sequencing technologies. These new technologies include improvements in sequencing systems based on Sanger's sequencing approach, as well as a number of non-Sanger sequencing technologies that became available during 2005–2008. The non-Sanger technologies include both sequencing based on amplified DNA molecules, and those based on single DNA molecules including Helicos true single molecule sequencing (tSMS) technology commercially launched in 2008. These new-generation sequencing technologies will certainly help in plant genomics research in a big way and may include a variety of research projects. While more plant genomes will be sequenced, epigenomes, transcriptomes, and metabolomes will also be worked out with much higher speed and at a cost reduced by several orders in magnitude. The science of plant genomics will also be influenced by the new emerging areas of "chemogenomics" and "synthetic genomics."

This special issue of the International Journal of Plant Genomics is devoted to "Genomics of Major Crops and Model Plant Species" with the aim to present an updated account of the genomics of major crop species and the model plant species. Articles published in this special issue involve almost all fields of genomics, including structural genomics, functional genomics, proteomics, metabolomics, and comparative genomics. Discussions also extend to cover phenomics, bioinformatics, epigenetics, and organellar genomics. Translational genomics from model plant species to cultivated crops and applications of genomics in crop improvement are topics for several articles. Structural genomics, as a major field for most crop plants, received a greater attention in this special issue, compared to other fields, including various types of molecular markers from RFLP to SNP and their use in construction of genetic, cytogenetic, and physical maps, QTL/gene mapping, genome sequencing, and generation of genomics resources. Functional genomics is the second field that received more attention, and some issues addressed significantly include gene isolation through map-based cloning and candidate gene approach, as well as functional analysis through insertional mutagenesis, RNAi, TILLING, and transcription profiling.

There are 14 review articles in this special issue, seven belonging to grass family, two devoted to legumes (soybean and *Medicago*), one devoted to oil-seed crop (*Brassica rapa*), and one each to cotton, tomato, potato, and *Citrus*. The special issue starts with several articles on genomics of food crops including wheat, barley and rice. There is a comprehensive article on wheat genomics written by P. K. Gupta et al. (Meerut, India) followed by an article giving an overview on barley genomics by N. Sreenivasulu et al. from IPK (Gatersleben, Germany). On rice genomics, there are two articles: one with emphasis on genome sequencing (written by T. Matsumoto et al. (Japan)) gives an account of international collaboration in sequencing rice genome and its annotation (including structure and composition of rice centromeres and telomeres), and the other on rice molecular breeding (written jointly by B. Collard (Queensland, Australia) and the rice genomics group (including D. J. Mackill) from International Rice Research Institute (IRRI) (Manila, Philippines)) gives a detailed account of how rice genomics resources can be utilized for molecular breeding. A. H. Paterson has written a review on *Sorghum* genomics (giving information on both markers and whole genome sequencing) and H. Budak et al. (from Turkey and Spain) give an updated account of the development of genomics resources for the grass genus *Brachypodium*, which is being preferred over the rice genus *Oryza* as a model for temperate grasses (including cereals and forage grasses). G. M. Souza et al. (Brazil) have written an article on sugarcane functional genomics, outlining the development and the use of ESTs and cDNA microarrays for gene discovery.

Among legume species, soybean (*Glycine max*) is an important crop world-wide, while *Medicago truncatula* and *Lotus japonicus* emerged as model systems for legume biology during the last decade. Therefore, one article on soybean genomics and another on *Medicago truncatula* have been included in this special issue. D. A. Lightfoot from The Illinois Soybean Center (Illinois, USA) discusses the use of Forresr cultivar for the development of genomic resources in this crop. Similarly, Julia Frugoli (SC, USA) with his two other colleagues elsewhere wrote an article on *Medicago truncatula* giving an updated account on the developments of genomic resources in this model legume. C. P. Hong et al. report the current understanding of the genome structure of *Brassica rapa* and efforts for the whole-genome sequencing of the species.

Hong-Bin Zhang from College Station (Texas, USA) and his coworkers (from University of Georgia and China) discuss advances on genomics research in cotton, highlighting the development of DNA marker linkage/physical maps, QTL mapping, ESTs, and whole genome sequencing. The last three articles deal with genomics of two related Solanaceous crops, namely, tomato and potato, as well as a fruit-tree genus (*Citrus*). L. Frusciante et al. from (Portici and Roma, Italy) give an updated account of tomato genomics, G. J. Bryan and I. Hein from Scottish Crop Research Institute (SCRI, Dundee, UK) give an account of potato genomics, and M. Talon (Valencia, Spain) with F. G. Gmitter (Citrus Research and Education Center, University of Florida, USA) give an account for *Citrus* genomics.

ACKNOWLEDGMENTS

We hope that the readers and the research workers in the field of plant genomics will find this special issue useful as a resource both for teaching and research. We would like to thank the reviewers who helped us in reviewing the articles submitted for this special issue and timely submitted their recommendations. We also would like to thank the Chief-editor Professor Hong-Bin Zhang from College Station (Texas, USA), and other staff of the Editorial Section of The International Journal of Plant Genomics (IJPG), who showed confidence in us and cooperated at all stages of the production of this special issue.

P. K. Gupta
Yunbi Xu

Hindawi Publishing Corporation
International Journal of Plant Genomics
Volume 2008, Article ID 896451, 36 pages
doi:10.1155/2008/896451

Review Article

Wheat Genomics: Present Status and Future Prospects

P. K. Gupta, R. R. Mir, A. Mohan, and J. Kumar

Molecular Biology Laboratory, Department of Genetics and Plant Breeding, Ch. Charan Singh University, Meerut 250 004, India

Correspondence should be addressed to P. K. Gupta, pkgupta36@gmail.com

Received 6 November 2007; Accepted 15 March 2008

Recommended by Yunbi Xu

Wheat (*Triticum aestivum* L.), with a large genome (16000 Mb) and high proportion (~80%) of repetitive sequences, has been a difficult crop for genomics research. However, the availability of extensive cytogenetics stocks has been an asset, which facilitated significant progress in wheat genomic research in recent years. For instance, fairly dense molecular maps (both genetic and physical maps) and a large set of ESTs allowed genome-wide identification of gene-rich and gene-poor regions as well as QTL including eQTL. The availability of markers associated with major economic traits also allowed development of major programs on marker-assisted selection (MAS) in some countries, and facilitated map-based cloning of a number of genes/QTL. Resources for functional genomics including TILLING and RNA interference (RNAi) along with some new approaches like epigenetics and association mapping are also being successfully used for wheat genomics research. BAC/BIBAC libraries for the subgenome D and some individual chromosomes have also been prepared to facilitate sequencing of gene space. In this brief review, we discuss all these advances in some detail, and also describe briefly the available resources, which can be used for future genomics research in this important crop.

1. INTRODUCTION

Wheat is one of the most important staple food crops of the world, occupying 17% (one sixth) of crop acreage worldwide, feeding about 40% (nearly half) of the world population and providing 20% (one fifth) of total food calories and protein in human nutrition. Although wheat production during the last four decades had a steady significant increase, a fatigue has been witnessed during the last few years, leading to the lowest current global wheat stocks ever since 1948/49. Consequently, wheat prices have also been soaring, reaching the highest level of US $ 10 a bushel as against US $ 4.50 a year ago (http://www.planetark.com/dailynewsstory.cfm/newsid/44968/story.htm). As against this, it is projected that, in order to meet growing human needs, wheat grain production must increase at an annual rate of 2%, without any additional land to become available for this crop [1]. In order to meet this challenge, new level of understanding of the structure and function of the wheat genome is required.

Wheat is adapted to temperate regions of the world and was one of the first crops to be domesticated some 10000 years ago. At the cytogenetics level, common wheat is known to have three subgenomes (each subgenome has 7 chromosomes, making $n = 21$) that are organized in seven homoeologous groups, each homoeologous group has three closely related chromosomes, one from each of the three related subgenomes. The diploid progenitors of the A, B, and D subgenomes have been identified, although there has always been a debate regarding the progenitor of B genome (reviewed in [1]). It has also been found that common wheat behaves much like a diploid organism during meiosis, but its genome can tolerate aneuploidy because of the presence of triplicate genes. These features along with the availability of a large number of aneuploids [particularly including a complete set of monosomics, a set of 42 compensating nullisomic-tetrasomics and a complete set of 42 ditelocentrics developed by Sears [2]] and more than 400 segmental deletion lines [developed later by Endo and Gill [3]] facilitated greatly the wheat genomics research.

Molecular tools have recently been used in a big way for cytogenetic studies in wheat, so that all recent cytogenetic studies in wheat now have a molecular component, thus paving the path for wheat genomics research. However, these studies in the area of molecular cytogenetics have been relatively difficult in bread wheat due to its three closely related subgenomes and a large genome (1C = >16 billion base pairs) with high proportion (>80%) of repetitive DNA. Despite this, significant progress in the area of molecular cytogenetics and cytogenomics of wheat has been made during the last two decades, thus making it amenable to genomics research. For instance, molecular maps in bread wheat, emmer wheat, and einkorn wheat utilizing a variety of molecular markers are now available, where gene rich regions (GRRs) and recombination hotspots have also been identified (for a review, see [4, 5]).

In recent years, a number of initiatives have been taken to develop new tools for wheat genomics research. These include construction of large insert libraries and development of massive EST collections, genetic and physical molecular maps, and gene targeting systems. For instance, the number of wheat ESTs has increased from a mere ~5 in 1999 [6] to a massive >1 240 000 in January 2008 (http://www.ncbi.nlm.nih.gov/), thus forming the largest EST collection in any crop as a resource for genome analysis. These ESTs are being used for a variety of activities including development of functional molecular markers, preparation of transcript maps, and construction of cDNA arrays. A variety of molecular markers that were developed either from ESTs or from genomic DNA also helped to discover relationships between genomes [7] and to compare marker-trait associations in different crops. Comparative genomics, involving major crop grasses including wheat, has also been used not only to study evolutionary relationships, but also to design crop improvement programs [8]. Functional genomics research in wheat, which though lagged far behind relative to that in other major food crops like maize and rice, has also recently witnessed significant progress. For instance, RNA interference, TILLING, and "expression genetics" leading to mapping of eQTLs have been used to identify functions of individual genes [9]. This allowed development of sets of candidate genes for individual traits, which can be used for understanding the biology of these traits and for development of perfect diagnostic marker(s) to be used not only for map-based cloning of genes, but also for MAS [9, 10]. In order to sequence the GRRs of wheat genome, a multinational collaborative program named International Genome Research on Wheat (IGROW) was earlier launched, which later took the shape of International Wheat Genome Sequencing Consortium (IWGSC) [11]. This will accelerate the progress on genome sequencing and will allow analysis of structure and function of the wheat genome. Keeping the above background in mind, Somers [12] identified the following five thrust areas of research for wheat improvement: (i) genetic mapping, (ii) QTL analysis, (iii) molecular breeding, (iv) association mapping, and (v) software development. In this communication, we briefly review the recent advances in all these areas of wheat genomics and discuss their impact on wheat improvement programs.

2. MOLECULAR MAPS OF WHEAT GENOME

2.1. *Molecular genetic maps*

Although some efforts toward mapping of molecular markers on wheat genome were initially made during late 1980s [13], a systematic construction of molecular maps in wheat started only in 1990, with the organization of International Triticeae Mapping Initiative (ITMI), which coordinated the construction of molecular maps of wheat genome. Individual groups (headed by R Appels, PJ Sharp, ME Sorrells, J Dvorak, BS Gill, GE Hart, and MD Gale) prepared the maps for chromosomes belonging to each of the seven different homoeologous groups. A detailed account on mapping of chromosomes of individual homoeologous groups and that of the whole wheat genome is available elsewhere [14]; an updated version is available at GrainGenes (http://wheat.pw.usda.gov/), and summarized in Table 1. Integrated or composite maps involving more than one type of molecular markers have also been prepared in wheat (particularly the SSR, AFLP, SNP, and DArT markers (see Table 1)). Consensus maps, where map information from multiple genomes or multiple maps was merged into a single comprehensive map, were also prepared in wheat [15, 16]. On these maps, classical and newly identified genes of economic importance are being placed to facilitate marker-assisted selection (MAS). Many genes controlling a variety of traits (both qualitative and quantitative) have already been tagged/mapped using a variety of molecular markers (for references, see [14, 17]). The density of wheat genetic maps was improved with the development of microsatellite (SSR) markers leading to construction of SSR maps of wheat [18–20]. Later, Somers et al. [16] added more SSR markers to these earlier maps and prepared a high-density SSR consensus map. At present, >2500 mapped genomic SSR (gSSR) markers are available in wheat, which will greatly facilitate the preparation of high-density genetic maps, so that we will be able to identify key recombination events in breeding populations and fine-map genes. In addition to gSSRs, more then 300 EST-SSR could also be placed on the genetic map of wheat genome [21–23]. However, more markers are still needed, particularly for preparation of high-density physical maps for gene cloning [24]. Availability of a number of molecular markers associated each with individual traits will also facilitate marker-assisted selection (MAS) during plant breeding.

In addition to random DNA markers (RDM), gene targeted markers (GTMs) and functional markers (FMs) are also being used in wheat to facilitate identification of genes responsible for individual traits and to improve possibilities of using MAS in wheat breeding. As a corollary, functional markers (FMs) are also being developed from the available gene sequences [10]. These markers were also used to construct transcript and molecular functional maps. Recently, microarray-based high-throughput diversity array technology (DArT) markers were also developed and used for preparing genetic maps in wheat [53, 54]. Large-scale genotyping for dozens to thousands of SNPs is also being undertaken

Table 1: A list of some important molecular maps developed in wheat.

Map type/class of wheat	Population used for mapping	No. of loci mapped	Genetic map length (cM)	Reference
RFLP maps				
Diploid wheat (D-genome)	F_2 [*T. tauschii* (TA1691 var. meyeri × TA1704 var. typica)]	152	1554	[25]
Diploid wheat (D-genome)	F_2 [*Aegilops tauschii* var. meyeri(TA1691) × *Ae. tauschii* var. typica(TA 1704)]	546	—	[26]
SSR maps				
Bread wheat	ITMI RILs (W7984 × Opata85)	279	—	[18]
Bread wheat	RILs (Synthetic × Opata)	1235	2569	[16]
Bread wheat	RILs (W7984 × Opata85)	1406	2654	[27]
Bread wheat	DHs (Kitamoe × Munstertaler)	464	3441	[28]
Bread wheat*	RILs (Chuan-Mai18 × Vigour18)	244	3150	[29]
AFLP maps				
Bread wheat*	RILs (Wangshuibai × Alondra's)	250	2430	[30]
Composite maps				
Einkorn wheat	F_{2s}/F_{3s} (*T. monococcum*ssp . *monococcum* DV92 × *T. monococcum* ssp. *aegilopoides* C3116) (marker loci-mainly RFLPs)	3335	714	[31]
Einkorn wheat	RILs (*Triticum boeoticum* × *T. monococcum*) marker loci-RFLPs, SSR	177	1262	[5]
Durum wheat	RILs (*T. durum* var. Messapia × *T. turgidium* var. MG4343) (marker loci-RFLP, Glu3B, others)	213	1352	[32]
Durum wheat	RILs (*T. durum* var. Messapia × *T. turgidium* var. MG4343) (marker loci-AFLPs, RFLPs)	88	2063	[33]
Durum wheat	RILs (Jennah Khetifa × Cham10 (marker loci-RFLPs, SSRs, AFLPs)	206	3598	[34]
Durum wheat*	RILs (Omrabi 5 × *T dicoccoides* 600545) (marker loci-SSRs, AFLPs)	312	2289	[35]
Bread wheat	RILs (*T. aestivum* L. var. Forno × *T. spelta* L. var. Oberkulmer) (marker loci-RFLPs, SSRs)	230	2469	[36]
Bread wheat*	DHs (CM-82036 × Remus) (marker loci-RFLPs, AFLPs, SSRs, etc.)	384	1860	[37]
Bread wheat*	DHs (Savannah × Senat) (marker loci-SSRs, AFLPs)	345 (17)	2300	[38]
Bread wheat*	RILs (Renan × Récital) (marker loci-SSRs, RFPLs, AFLPs)	265 (17)	2722	[39, 40]
Bread wheat	F_{5s}(Arina × Forno) (marker loci-RFLPs, SSRs)	396	3086	[41]
Bread wheat	DHs (Courtot × Chinese Spring) (marker loci-RFLPs, SSRs, AFLPs)	659	3685	[42]
Bread wheat*	DHs (Frontana × Remus) (marker loci-SSRs, STSs, AFLPs, etc.)	535	2840	[43]
Bread wheat	RILs (Grandin × BR34) (marker loci-TRAPs, SSRs)	352	3045	[44]
Bread wheat*	DHs (Spring × SQ1) (marker loci-AFLPs, SSRs)	567	3521	[45]
Bread wheat*	RILs (Dream × Lynx) (marker loci-SSRs, STSs, AFLPs)	283 (17)	1734	[46]
Bread wheat*	DHs (AC Karma × 87E03-S2B1) (marker loci-STSs, SSRs, etc.)	167 (15)	2403	[47]

Table 1: Continued.

Map type/class of wheat	Population used for mapping	No. of loci mapped	Genetic map length (cM)	Reference
Bread wheat*	DHs (Trident × Molineux) (marker loci-SSRs, STSs, RFLPs, etc.)	251	3061	[48]
Bread wheat*	DH (Arina × Riband) (marker loci-AFLPs, SSRs)	279	1199	[49]
Bread wheat*	DHs (RL4452 × AC Domain) (marker loic-SSRs, genes, etc.)	369	2793	[50]
Bread wheat*	RILs (Chuan 35050 × Shannong 483) (marker loci-SSRs, EST-SSRs, ISSRs, SRAPs,TRAPs, Glu loci)	381	3636	[51]
Bread wheat*	DHs (Shamrock × Shango) (marker loci-SSRs, DArTs)	263	1337	[52]
Bread wheat	DHs Cranbrook × Halberd (Marker loci-SSRs, RFLPs, AFLPs, DArTs, STSs)	749	2937	[53]

* These are framework linkage map prepared for QTL analyses.

using several high-density platforms including Illumina's GoldenGate and ABI's SNaPshot platforms (http://wheat.pw.usda.gov/SNP/new/index.shtml). The genotyping activity may be extended further through the use of Solexa's high throughput and low-cost resequencing technology.

2.2. Molecular marker-based physical maps

Molecular markers in bread wheat have also been used for the preparation of physical maps, which were then compared with the available genetic maps involving same markers. These maps allowed comparisons between genetic and physical distances to give information about variations in recombination frequencies and cryptic structural changes (if any) in different regions of individual chromosomes. Several methods have been employed for the construction of physical maps.

2.2.1. Deletion mapping

In wheat, physical mapping of genes to individual chromosomes began with the development of aneuploids [55], which led to mapping of genes to individual chromosomes. Later, deletion lines of wheat chromosomes developed by Endo and Gill [3] were extensively used as a tool for physical mapping of molecular markers. Using these deletion stocks, genes for morphological characters were also mapped to physical segments of wheat chromosomes directly in case of unique and genome specific markers or indirectly in case of duplicate or triplicate loci through the use of intergenomic polymorphism between the A, B, and D subgenomes (see Table 2 for details of available physical maps). In addition to physical mapping of genomic SSRs, ESTs and EST-SSRs were also subjected to physical mapping (see Table 2). As a part of this effort, a major project (funded by National Science Foundation, USA) on mapping of ESTs in wheat was successfully completed by a consortium of 13 laboratories in USA leading to physical mapping of ~16000 EST loci (http://wheat.pw.usda.gov/NSF/progress_mapping.html; [56] (see Table 2)).

2.2.2. In silico physical mapping

As many as 16000 wheat EST loci assigned to deletion bins, as mentioned above, constitute a useful source for in silico mapping, so that markers with known sequences can be mapped to wheat chromosomes through sequence similarity with mapped EST loci available at GrainGene database (http://wheat.pw.usda.gov/GG2/blast.shtml). Using the above approach, Parida et al. [80] were able to map 157 SSR containing wheat unique sequences (out of 429 class I unigene-derived microsatellites (UGMS) markers developed in wheat) to chromosome bins. These bin-mapped UGMS markers provide valuable information for a targeted mapping of genes for useful traits, for comparative genomics, and for sequencing of gene-rich regions of the wheat genome. Another set of 672 loci belonging to 275 EST-SSRs of wheat and rye was assigned to individual bins through in silico and wet-lab approaches by Mohan et al. [79]. A few cDNA clones associated with QTL for FHB resistance in wheat were also successfully mapped using in silico approach [81].

2.2.3. Radiation-hybrid mapping

Radiation hybrid (RH) mapping was first described by Goss and Harris [82] and was initially used by Cox et al. [83] for physical mapping in animals/humans. In wheat, the approach has been used at North Dakota State University (NDSU) utilizaing addition and substituition of individual D-genome chromosomes into tetraploid durum wheat. For RH mapping of 1D, durum wheat alien substitution line for chromosome 1D (DWRH-1D), harboring nuclear-cytoplasmic compatibility gene scs^{ae} was used. These RH lines initially allowed detection of 88 radiation-induced breaks involving 39 1D specific markers. Later, this 1D RH map was further expanded to a resolution of one break every 199 kb of DNA, utilizing 378 markers [84]. Using the

Table 2: Deletion-based physical maps of common wheat.

Homoeologous group/ chromosome/arm	Marker loci mapped	No. of deletion stocks used	Reference
1	19 RFLPs	18	[57]
1	50 RFLPs	56	[58]
2	30 RFLPs	21	[59]
2	43 SSRs	25	[60]
3	29 RFLPs	25	[61]
4	40 RFLPs	39	[62]
5	155 RFLPs	65	[63]
5	245 RFLPs, 3 SSRs	36	[64]
5S	100 RFLPs	17	[65]
5A	22 RFLPs	19	[66]
6	24 RFLPs	26	[67]
6	210 RFLPs	45	[68]
6S	82 RFLPs	14	[69]
7	16 RFLPs	41	[70]
7	91 RFLPs, 6 RAPDs	54	[71]
6B, 2D, and 7D	16 SSRs	13	[72]
1BS	24 AFLPs	8	[73]
4DL	61 AFLPs, 2 SSRs, 2 RFLPs	8	[74]
1BS	22 ESTs	2	[75]
Whole genome	725 SSRs	118	[76]
Whole genome	260 BARC	117	[27]
Whole genome	313 SSRs	162	[77]
Whole genome	16000 ESTs	101	http://wheat.pw.usda.gov/NSF/progressmapping.html
Whole genome	266 eSSRs	105	[78]
Whole genome	672 EST-SSRs	101	[79]

same approach, construction of radiation hybrid map for chromosome 3B is currently in progress (S. Kianian personal communication).

2.3. *BAC-based physical maps*

BAC-based physical map of wheat D genome is being constructed using the diploid species, *Aegilops tauschii*, with the aim to identify and map genes and later sequence the gene-rich regions (GRRs). For this purpose, a large number of BACs were first fingerprinted and assembled into contigs. Fingerprint contigs (FPCs) and the data related to physical mapping of the D genome are available in the database (http://wheat.pw.usda.gov/PhysicalMapping/index.html). BACs belonging to chromosome 3B are also being fingerprinted (with few BACs already anchored to wheat bins), and a whole genome BAC-based physical map of hexaploid wheat is proposed to be constructed under the aegis of IWGSC in its pilot studies (see later).

3. IN SITU HYBRIDIZATION STUDIES IN WHEAT

In bread wheat, in situ hybridization (ISH) involving radioactively labeled probes was initially used to localize repetitive DNA sequences, rRNA and alien DNA segments [104–106]. Later, fluorescence in situ hybridization (FISH), multicolor FISH (McFISH, simultaneous detection of more than one probe), and genome in situ hybridization (GISH, total genomic DNA as probe) were used in several studies. FISH with some repeated sequences as probes was used for identification of individual chromosomes [107–110]. FISH was also utilized to physically map rRNA multigene family [111, 112], RFLP markers [110, 113], and unique sequences [114–116] and also for detecting and locating alien chromatin introgressed into wheat [117–119].

A novel high-resolution FISH strategy using super-stretched flow-sorted chromosomes was also used (extended DNA fibre-FISH; [120–122]) to fine map DNA sequences [123, 124] and to confirm integration of transgenes into the wheat genome [125].

Recently, BACs were also utilized as probes for the so called BAC-FISH which helped not only in the discrimination between the three subgenomes, but also in the identification of intergenomic translocations, molecular cytogenetic markers, and individual chromosomes [126]. BAC-FISH also helped in localization of genes (BACs carrying genes) and in studying genome evolution and organization among wheat and its relatives [110, 127, 128].

Table 3: Genes already cloned or likely to be cloned through map-based cloning in wheat.

Gene/QTL	Trait	Reference
Lr1	Leaf rust resistance	[85, 86]
Lr10	Leaf rust resistance	[87]
Lr21	Leaf rust resistance	[88]
VRN1	Vernalization response	[89]
VRN2	Vernalization response	[90]
VRN3	Vernalization response	[91]
Q	Free threshing character	[92, 93]
Pm3b	Powdery mildew resistance	[94, 95]
GPC-B1	High grain protein content	[96, 97]
Qfhs.Ndsu-3bs	Fusarium head blight resistance	[98]
Yr5	Resistance to stripe rust	[99]
B	Boron tolerance	[100]
Fr2	Frost resistance	http://www.agronomy.ucdavis.edu/Dubcovsky
EPS-1	Flowering time	http://www.agronomy.ucdavis.edu/Dubcovsky
Tsn1	Host-selective toxin *Ptr ToxA*	[101]
Ph1	Chromosome pairing locus	[102]
Sr2	Stem rust resistance	[103]

4. MAP-BASED CLONING IN WHEAT

In wheat, a number of genes for some important traits including disease resistance, vernalization response, grain protein content, free threshing habit, and tolerance to abiotic stresses have been recently cloned/likely to be cloned via map-based cloning (see Table 3). The first genes to be isolated from wheat by map-based cloning included three resistance genes, against fungal diseases, including leaf rust (*Lr21*; [88, 129, 130] and *Lr10*; [87]) and powdery mildew (*Pm3b* ; [94]). A candidate gene for the *Q* locus conferring free threshing character to domesticated wheat was also cloned [92]. This gene influences many other domestication-related traits like glume shape and tenacity, rachis fragility, plant height, spike length, and ear-emergence time. Another important QTL, *Gpc-B1*, associated with increased grain protein, zinc, and iron content has been cloned, which will contribute in breeding enhanced nutritional value wheat in future [96]. Cloning of three genes for vernalization response (*VRN1, VRN2, VRN3*) helped in postulating a hypothetical model summarizing interactions among these three genes [89–91, 131].

5. EST DATABASES AND THEIR USES

During the last 8–10 years, more than 1240455 wheat ESTs have become available in the public domain as in January 2008 (http://www.ncbi.nlm.nih.gov/). A number of cDNA libraries have been used for this purpose. These ESTs proved to be an enormous resource for a variety of studies including development of functional molecular markers (particularly SSRs and SNPs), construction of a DNA chip, gene expression, genome organization, and comparative genomics research.

5.1. *EST-derived SSRs*

Wheat ESTs have been extensively used for SSR mining (1SSR/10.6 kb; [80]), so that in our own laboratory and elsewhere detected by author, a large number of SSRs have already been developed from EST sequences [22, 78, 80, 132–134]. These EST-SSRs served as a valuable source for a variety of studies including gene mapping, marker-aided selection (MAS), and eventually positional cloning of genes. The ESTs and EST-derived SSRs were also subjected to genetic and physical mapping (see above).

Since EST-SSRs are derived from the expressed portion of the genome, which is relatively more conserved, these markers show high level of transferability among species and genera [133, 135]. However, the transferability of wheat EST-SSRs to closely related triticeae species (*Triticum* and *Aegilops* species) is higher as compared to more distant relatives such as barley, maize, rice, sorghum, oats, and rye. The EST-SSRs thus also prove useful in comparative mapping, transfer of markers to orphanage wild species, and for genetic diversity estimates [79, 132, 134, 136–139].

5.2. *EST-derived SNPs and the International SNP Consortium*

In recent years, single nucleotide polymorphisms (SNPs) have become the markers of choice. Therefore, with the aim to discover and map SNPs in tetraploid and hexaploid wheats, an International Wheat SNP Consortium was constituted, and comprehensive wheat SNP database was developed (http://wheat.pw.usda.gov/SNP/new/index.shtml). Approximately 6000 EST unigenes from the database of mapped ESTs and other EST databases were distributed to consortium members for locating SNPs, for designing conserved primers for these SNPs and for validation of these SNP. Considerable progress has been made in this direction in different laboratories; the project data are accessible through http://wheat.pw.usda.gov/SNP/snpdb.html. In May 2006, the database contained 17174 primers (forward and reverse), 1102 wheat polymorphic loci, and 2224 polymorphic sequence tagged sites in diploid ancestors of polyploid wheat. Zhang et al. [140] also reported 246 gene loci with SNPs and/or small insertions/deletions from wheat homoeologous group 5. Another set of 101 SNPs (1SNP/212 bp) was discovered from genomic sequence analysis in 26-bread wheat lines and one synthetic line (http://urgi.versailles.inra.fr/GnpSNP/, [141]).

6. BAC/BIBAC RESOURCES

BAC/BIBAC libraries have been produced in diploid, tetraploid, and hexaploid wheats (see Table 4). Chromosome-specific BAC libraries were also prepared in hexaploid wheat [142–144]. These BAC resources proved useful for a variety of studies including map-based cloning (see Table 3), organization of wheat genome into gene-rich and gene-poor regions that are loaded with retroelements [8, 145–147], and for physical mapping and sequencing of wheat genome (http://wheatdb.ucdavis.edu:8080/wheatdb/, [11]).

7. GENE DISTRIBUTION IN WHEAT: GENE-RICH AND GENE-POOR REGIONS

Genetic and physical maps of the wheat genome, discussed above, have been utilized for a study of gene distribution within the genome [58, 63, 148]. In order to identify and demarcate the gene-containing regions, 3025 loci including 252 phenotypically characterized genes and 17 quantitative trait loci (QTL) were physically mapped with the help of deletion stocks [149, 150]. It was shown that within the genome, genes are not distributed randomly and that there are gene-rich regions (GRRs) and gene-poor regions (GPRs), not only within the wheat genome, but perhaps in all eukaryotes (for reviews, see [4, 151]).

In wheat genome, 48 GRRs containing 94% of gene markers were identified with an average of ∼7 such GRRs (range 5–8) per homoeologous group. It was also shown that different wheat chromosomes differed for number and location of GRRs, with 21 GRRs on the short arms containing 35% of the wheat genes, and the remaining 27 GRRs on the long arms containing about 59% of the genes. The GRRs also vary in their size and in gene-density with a general trend of increased gene-density toward the distal parts of individual chromosome arms. This is evident from the fact that more than 80% of the total marker loci were mapped in the distal half of the chromosomes and ∼58% mapped in the distal 20%.

Among 48 GRRs, there were 18 GRRs (major GRRs), which contained nearly 60% of the wheat genes, covering only 11% of the genome, suggesting a very high density of genes in these GRRs, although the number and density of genes in these 18 GRRs was also variable [149, 150]. It has also been shown that the size of GRRs decreases and the number of GRRs increases, as the genome size increases from rice to wheat [4]. For instance, the average size of gene clusters in rice is ∼300 kb as compared to less than 50 kb in wheat and barley. However, no correlation was observed between the chromosome size and the proportion of genes or the size of the GRRs. For instance, group 3 has the longest chromosomes among the wheat homoeologous groups but contained only 13% of the genes compared to group 5 chromosomes that contained 20% of genes [150].

For the chromosomes of homoeologous group 1, the distribution of genes and recombination rates have been studied in a relatively greater detail. Each chromosome of this group (1A, 1B, 1D) has eight GRRs (ranging in size from 3 Mb to 35 Mb), occupying ∼119 Mb of the 800-Mb-long chromosome. Using this homoeologous group, it was confirmed that the GRRs differ in the number of genes and gene-density even within a chromosome or its arms. For instance, the "1S0.8 region" is the smallest of all GRRs, but has the highest gene-density, which is ∼12 times that in the "1L1.0 region."

The distribution of GRRs has also been compared with the distribution of chromosome breaks involved in the generation of deletion stocks that are currently available and have been used for physical mapping of wheat genome. It was found that the breakpoints are nonrandom, and occur more frequently around the GRRs (one break every 7 Mb; [58, 67]); they seem to occur around GRRs twice as frequently as one would expect on random basis (one break every 16 Mb; [149]). Consequently, GRRs interspersed by <7-Mb-long GPRs will not be resolved and better resolution would be needed to partition the currently known GRRs into mini-GRRs and GPRs.

It has also been inferred that perhaps in eukaryotic genomes, the "gene-poor" regions preferentially enlarged during evolution, as is obvious in wheat, where large, essentially, "gene-empty" blocks of up to ∼192 Mb are common. Taking polyploidy into account, 30% gene-rich part of the genome is still ∼4 times larger than the entire rice genome [149]. Therefore, gene distribution within the currently defined GRRs of wheat would probably be similar to that in the rice genome, except that the gene-clusters would be smaller and the interspersing "gene-empty" regions would be larger, similar to barley as described above. It has also been shown that the "gene-empty" regions of the higher eukaryotic genomes are mainly comprised of retrotransposons and pseudogenes [152, 153]. The proportion of retrotransposons is significantly higher than pseudogenes, especially in the larger genomes, like those of maize and bread wheat.

8. VARIABLE RECOMBINATION RATES

The recombination rate has also been recently shown to vary in different regions of the wheat genome. This was demonstrated through a comparison of consensus physical and genetic maps involving 428 common markers [149, 150]. Recombination in the distal regions was generally found to be much higher than that in the proximal half of individual chromosomes, and a strong suppression of recombination was observed in the centromeric regions. Recombination rate

Table 4: BAC libraries available in wheat.

Species (accession)	Coverage	Restriction site	No. of clones (clone size in kb)	Curator
T. monococcum (DV92)	5.6 X	*Hin*d III	276000 (115)	J. Dubcovsky
T. dicoccoides (Langdon)	5.0 X	*Hin*d III	516000 (130)	J. Dubcovsky
T. urartu (G1812)	4.9 X	*Bam*H I	163200 (110)	J. Dvorak
Ae. tauschii (AL8/78)	2.2 X	*Eco*R I	54000 (167)	H.B. Zhang
Ae. tauschii (AL8/78)	2.2 X	*Hin*d III	59000 (189)	H.B. Zhang
Ae. tauschii (AL8/78)	3.2 X	*Hin*d III	52000 (190)	H.B. Zhang
Ae. tauschii (AL8/78)	2.8 X	*Bam*H I	59000 (149)	H.B. Zhang
Ae. tauschii (AL8/78)	2.4 X	*Bam*H I	76000 (174)	H.B. Zhang
Ae. tauschii (Aus 18913)	4.2 X	*Hin*d III	144000 (120)	E. Lagudah
Ae. tauschii (AS75)	4.1 X	*Bam*H I	181248 (115)	J. Dvorak
Ae. speltoides (2-12-4-8-1-1-1)	5.4 X	*Bam*H I	237312 (115)	J. Dvorak
T. aestivum (Glenlea)	3.1 X	*Bam*H I & *Hin*d III	656640 (80)	S. Cloutier
T. aestivum (Renan)	3.2 X	*Hin*d III	478840 (150)	B. Chalhoub
T. aestivum (Renan)	2.2 X	*Eco*R I	285312 (132)	B. Chalhoub
T. aestivum (Renan)	1.5 X	*Bam*H I	236160 (122)	B. Chalhoub
T. aestivum (Chinese Spring)		*Hin*d III	950000 (54)	Y. Ogihara
T. aestivum (Chinese Spring)	< 4%	*Mlu* I	>12000 (45)	K. Willars
		Not I	>1000	
T. aestivum (Chinese Spring) 3B	6.2 X	*Hin*d III	67968 (103)	J. Dolezel & B. Chalhoub
T. aestivum, (Chinese Spring) 1D, 4D & 6D	3.4 X	*Hin*d III	87168 (85)	J. Dolezel & B. Chalhoub
T. aestivum (Pavon) 1BS	14.5 X	*Hin*d III	65280 (82)	J. Dolezel & B. Chalhoub
T. aestivum (*AVS-Yr5*)	3.6 X	*Hin*d III	422400 (140)	X.M. Chen
T. aestivum (Norstar)	5.5 X	*Hin*d III	1200000 (75)	R. Chibbar

among GRRs present in the distal half of the chromosome was highly variable with higher recombination in some proximal GRRs than in the distal GRRs [149, 150]. The gene poor-regions accounted for only ~5% of recombination.

It has also been reported that the distribution of recombination rates along individual chromosomes is uneven in all eukaryotes studied so far (for more references, see [154, 155]). Among cereals, the average frequency of recombination in rice (with the smallest genome) is translated into a genetic distance of about 0.003 cM per kb with a range of 0 to 0.06 cM per kb (http://rgp.dna.affrc.go.jp/Publicdata.html) and that of wheat (the largest genome) is 0.0003 cM per kb with a range from 0 to 0.007 cM per kb. Non-recombinogenic regions were observed in yeast as well as in rice, but the highest recombination rate for a region appears to be ~35-fold lower in rice and 140-fold lower in bread wheat (relative to yeast). It may be due to differences in the resolution of recombination rates, which is ~400 kb in rice (in wheat the resolution is much lower than in rice), whereas the resolution in recombination hotspots in yeast may be as high as only <1 kb in length. Due to averaging over larger regions, recombination in hotspots in rice and wheat may appear to be low relative to that in yeast [4, 150, 151, 156].

9. FLOW CYTOGENETICS AND MICRODISSECTION OF CHROMOSOMES IN WHEAT

Flow cytogenetics and microdissection facilitated physical dissection of the large wheat genome into smaller and defined segments for the purpose of gene discovery and genome sequencing. Flow karyotypes of wheat chromosomes were also prepared [157–159]. DNA obtained from the flow-sorted chromosomes has been used for the construction of chromosome-specific large-insert DNA libraries, as has been done for chromosome 4A [157, 159]. Later, all individual 42 chromosome arms involving 21 wheat chromosomes were also sorted out using flow cytometry [160]. In another study, it was also possible to microdissect 5BL isochromosomes from meiotic cells and to use their DNA with degenerate oligonucleotide primer PCR (DOP-PCR) to amplify chromosome arm-specific DNA sequences. These amplified PCR sequences were then used as probes for exclusive painting of 5BL [161].

Flow sorting in wheat has also been used for efficient construction of bacterial artificial chromosome (BAC) libraries for individual chromosomes [143, 162]. The use of these chromosome- and chromosome arm-specific BAC libraries is expected to have major impact on wheat genomics research [1]. For instance, the availability of 3B-specific BAC library facilitated map-based cloning of agronomically important genes such as major QTL for Fusarium head blight resistance [98]. Flow cytometry can also be used to detect numerical and structural changes in chromosomes and for the detection of alien chromosomes or segments thereof (reviewed in detail by [163]). For instance, a 1BL.1RS translocation could be detected by a characteristic change in the flow karyotype [164]. In addition, DNA from flow-sorted chromosomes can be used for hybridization on DNA arrays and chips, with the aim of assigning DNA sequences to specific chromosome arms. This technique will be extensively used now with the availability of Affymetrix wheat GeneChip [165].

10. WHEAT GENE SPACE SEQUENCING

International Triticeae Mapping Initiative (ITMI), at its meeting held at Winnipeg, Canada during June 1–4, 2003, took the first initiative toward whole genome sequencing (WGS) in wheat and decided to launch a project that was described as International Genome Research of Wheat (IGROW) by B. S. Gill. A workshop on wheat genome sequencing was later organized in Washington, DC during November 11–13, 2003, which was followed by another meeting of IGROW during the National Wheat Workers Workshop organized at Kansas, USA, during Feb 22–25, 2004 [166]. Consequently, IGROW developed into an International Wheat Genome Sequencing Consortium (IWGSC). Chinese Spring (common wheat) was selected for WGS, since it already had ample genetic and molecular resources [1].

Three phases were proposed for sequencing the wheat genome: pilot, assessment, and scale up. The first phase was recommended for 5 years and is mainly focused on the short-term goal of IWGSC, involving physical and genetic mapping along with sample sequencing of the wheat genome aimed at better understanding of the wheat genome structure. The assessment phase will involve determining which method(s) can be used in a cost-effective manner to generate the sequence of the wheat genome. After a full assessment, the scale-up phase will involve the deployment of optimal methods on the whole genome, obtaining the genome sequence and annotation, which is the long-term goal of IWGSC. With the availability of new sequencing technologies provided by 454/Roche and those provided by Illumina/Solexa and ABI SOLiD [167]; sequencing of gene space of the wheat genome, which was once thought to be almost impossible, should become possible within the foreseeable future.

First pilot project for sequencing of gene space of wheat genome, led by INRA in France, was initiated in 2004 using the largest wheat chromosome, 3B (1GB = 2x the rice genome) of hexaploid wheat as a model. As many as 68000 BAC clones from a 3B chromosome specific BAC library [143] were fingerprinted and assembled into contigs, which were then anchored to wheat bins, covering ~80% of chromosome 3B. Currently, one or more of these contigs are being sequenced [11], which will demonstrate the feasibility of large-scale sequencing of complete gene space of wheat genome.

11. FUNCTIONAL GENOMICS

The determination of the functions of all the genes in a plant genome is the most challenging task in the postgenomic era of plant biology. However, several techniques or platforms, like serial analysis of gene expression (SAGE), massively parallel signature sequencing (MPSS), and micro- and macroarrays, are now available in several crops for the estimation of mRNA abundance for large number of genes simultaneously. The microarrays have also been successfully used in wheat for understanding alterations in the transcriptome of hexaploid wheat during grain development, germination and plant development under abiotic stresses [168, 169]. Recently, a comparison was made between Affymetrix GeneChip Wheat Genome Array (an in-house custom-spotted complementary DNA array) and quantitative reverse transcription-polymerase chain reaction (RT-PCR) for the study of gene expression in hexaploid wheat [170]. Also, functional genomics approach in combination with "expression genetics" or "genetical genomics" provides a set of candidate genes that can be used for understanding the biology of a trait and for the development of perfect or diagnostic marker(s) to be used in map-based cloning of genes and MAS [9]. A similar example was provided by Jordan et al. [9], when they identified regions of wheat genome controlling seed development by mapping 542 eQTLs, using a DH mapping popultion that was earlier used for mapping of SSRs and QTL analysis of agronomic and seed quality traits [171]. Expression analysis using mRNA from developing seeds from the same mapping population was also conducted using Affymetrix GeneChip Wheat Genome Array [172].

11.1. RNA interference for wheat functional genomics

RNA interference (RNAi), which was the subject of the 2006 Nobel Prize in Physiology or Medicine, is also being extensively utilized for improvement of crop plants [173]. This technique does not involve introduction of foreign genes and thus provides an alternative to the most controversial elements of genetic modification. Plans in Australia are underway, where the knowledge gained from RNAi approach will be used for developing similar wheats by conventional method of plant breeding, as suggested by CSIRO scientists for developing high-fibre wheat [174]. In bread wheat, in particular, the technology provides an additional advantage of silencing all genes of a multigene family including homoeoloci for individual genes, which are often simultaneously expressed, leading to a high degree of functional gene redundancy [175]. It has been shown that delivery of specific dsRNA into single epidermal cells in wheat transiently interfered with gene function [176, 177]. Yan et al. [90] and Loukoianov et al. [178] used RNAi for stable transformation and to demonstrate that RNAi-mediated reduction of *VRN2* and *VRN1* transcript levels, respectively, accelerated and delayed flowering initiation in winter wheat. Similarly, Regina et al. [179] used RNAi to generate high-amylose wheat. However, none of the above studies reported long-term phenotypic stability of RNAi-mediated gene silencing over several generations, neither did they report any molecular details on silencing of homoeologous genes. However, Travella et al. [180] showed RNAi results in stably inherited phenotypes suggesting that RNAi can be used as an efficient tool for functional genomic studies in polyploid wheat. They introduced dsRNA-expressing constructs containing fragments of genes encoding *Phytoene Desaturase* (*PDS*) or the signal transducer of ethylene, *Ethylene Insensitive 2* (*EIN2*) and showed stably inherited phenotypes of transformed wheat plants that were similar to mutant phenotypes of the two genes in diploid model plants. Synthetic microRNA constructs can also be used as an alternative to large RNA fragments for gene silencing, as has been demonstrated for the first time in wheat by Yao et al. [181] by discovering and predicting targets for 58 miRNAs, belonging to 43 miRNA families (20 of these are conserved and 23 are novel to wheat); more importantly four of these miRNAs are monocot specific. This study will serve as a foundation for the future functional genomic studies. The subject of the use of RNAi for functional genomics in wheat has recently been reviewed [173].

11.2. TILLING in wheat

Recently, Targeting Induced Local Lesions IN Genomes (TILLING) was developed as a reverse genetic approach to take advantage of DNA sequence information and to investigate functions of specific genes [182]. TILLING was initially developed for model plant *Arabidopsis thaliana* [183] having fully sequenced diploid genome and now has also been successfully used in complex allohexaploid genome of wheat, which was once considered most challenging candidate for reverse genetics [184].

To demonstrate the utility of TILLING for complex genome of bread wheat, Slade et al. [185] created TILLING library in both bread and durum wheat and targeted *waxy* locus, a well characterized gene in wheat encoding granule bound *starch synthase I* (*GBSSI*). Loss of all copies of this gene results in the production of *waxy* starch (lacking amylose). Production of waxy wheat by traditional breeding was difficult due to lack of genetic variation at one of the *waxy* loci. However, targeting *waxy* loci by TILLING [185], using locus specific PCR primers led to identification of 246 alleles (196 alleles in hexaploid and 50 alleles in tetraploid) using 1920 cultivars of wheat (1152 hexaploid and 768 tetraploid). This made available novel genetic diversity at *waxy* loci and provided a way for allele mining in important germplasm of wheat. The approach also allowed evaluation of a triple homozygous mutant line containing mutations in two waxy loci (in addition to a naturally occurring deletion of the third locus) and exhibiting a near waxy phenotype.

Another example of on-going research using TILLING in wheat is the development of EMS mutagenised populations of *T. aestivum* (cv. Cadenza, 4200 lines, cv. Paragon, 6000 lines), *T. durum* (cv. Cham1, 4,200 lines), and *T. monococcum* (Accession DV92, 3000 lines) under the Wheat Genetic Improvement Network (WGIN; funded by Defra and BBSRC in the UK and by the EU Optiwheat programme). The aim of this program is to search noval variant alleles for *Rht-b1c,RAR-1, SGT-1*, and *NPR-1* genes (personal communication: andy.phillips@bbsrc.ac.uk and Simon.Orford@bbsrc.ac.uk).

The above examples provide proof-of-concept for TILLING other genes, whose mutations may be desired in wheat or other crops. However, homoeolog-specific primers are required in order to identify new alleles via TILLING in wheat. In case of *waxy*, the sequences of the three homoeologous sequences were already known, which facilitated primer designing, but TILLING of other genes may require cloning and sequencing of these specific genes in order to develop homoeolog-specific target primers.

12. COMPARATIVE GENOMICS

In cereals, a consensus map of 12 grass genomes including wheat is now available, representing chromosome segments of each genome relative to those in rice on the basis of mapping of anchor DNA markers [186]. Some of the immediate applications of comparative genomics in wheat include a study of evolution [187] and isolation/characterization of genes using the model genome of rice. The genes, which have been examined using comperative genomics approach include the pairing gene, *Ph1* [102, 188], gene(s) controlling preharvest sprouting (PHS; [189]), receptor-like kinase loci [190], gene for grain hardness [191], genes for glume coloration and pubescence (*Bg, Rg*; [192]), and the *Pm3* gene, responsible for resistance against powdery mildew [187].

Conservation of colinearity and synteny

Among cereals, using molecular markers, colinearity was first reported among A, B, and D subgenomes of wheat [13, 193], and later in the high-gene density regions of wheat and barley. At the *Lrk10* locus in wheat and its orthologous region in barley, a gene density of one gene per 4-5 kb was observed, which was similar to that found in *A. thaliana* [6]. Conservation of colinearity between homoeologous A genomes of diploid einkorn wheat and the hexaploid was also exploited for chromosome walking leading to cloning of candidate gene for the leaf rust resistance locus *Lr10* in bread wheat [194]. *Lr10* locus along with LMW/HMW loci of diploid wheat, when compared with their orthologs from tetraploid and hexaploid wheats, was found to be largely conserved except some changes that took place in intergenic regions [195–197]. On the basis of divergence of intergenic DNA (mostly transposable elements), tetraploid and hexaploid wheats were shown to have diverged about 800 000 years ago [197]. Similarly, the divergence of diploid from the tetraploid/hexaploid lineage was estimated to have occurred about 2.6–3 million years ago [195, 196].

Notwithstanding the above initial demonstration of colinearity using molecular markers, later studies based on genome sequences suggested disruption of microcolinearity in many regions thus complicating the use of rice as a model for cross-species transfer of information in these genomic regions. For instance, Guyot et al. [198] conducted an in silico study and reported a mosaic conservation of genes within a novel colinear region in wheat chromosome 1AS and rice chromosome 5S. Similarly, Sorrells et al. [199] while comparing 4485 physically mapped wheat ESTs to rice genome sequence data belonging to 2251 BAC/PAC clones, resolved numerous chromosomal rearrangements. The above findings also received support from sequence analysis of the long arm of rice chromosome 11 for rice-wheat synteny [200].

More recently, the grass genus *Brachypodium* is emerging as a better model system for wheat belonging to the genus *Triticum*, because of a more recent divergence of these two genera (35–40 million years) relative to wheat-rice divergence [201–203]. Also, sequence of *Brachypodium*, which is likely to become available in the near future, may help further detailed analyses of colinearity and synteny among grass genomes. This has already been demonstrated through a comparison of 371 kb sequence of *B. sylvaticum* with orthologous regions from rice and wheat [204]. In this region, *Brachypodium* and wheat showed perfect macrocolinearity, but rice was shown to contain ~220 kb inversion relative to *Brachypodium* sequence. Also, in *Ph1* region, more orthologous genes were identified between the related species *B. sylvaticum* and wheat than between wheat and rice, thus once again demonstrating relative utility of *Brachypodium* genome as a better model than rice genome for wheat comparative genomics [102, 188].

13. EPIGENETICS IN WHEAT

Epigenetics refers to a heritable change that is not a result of a change in DNA sequence, but, instead, results due to a chemical modification of nucleotides in the DNA or its associated histone proteins in the chromatin. Several studies have recently been intiated to study the epigenetic modifications in the wheat genome. For instance, methylation-sensitive amplified polymorphism (MSAP) has been used to analyze the levels of DNA methylation at four different stages (2d, 4d, 8d, and 30d after pollination) of seed development in bread wheat [205]. It was found that 36–38% of CCGG sites were either fully methylated at the internal C's and/or hemimethylated at the external C's at the four corresponding stages. Similarly, Shitsukawa et al. [206] also studied genetic and epigenetic alterations among three homoeologs in the two class E-type wheat genes for flower development, namely, *wheat SEPALLATA* (*WSEP*) and *wheat LEAFY HULL STERILE1* (*WLHS1*). Analyses of gene structure, expression patterns, and protein functions showed that no alterations were present in the *WSEP* homoeologs. By contrast, the three *WLHS1* homoeologs showed genetic and epigenetic alterations. It was shown that *WLHS1-B* was predominantly silenced by cytosine methylation, suggesting that the expression of three homoeologous genes is differentially regulated by genetic or epigenetic mechanisms. Similar results were reported for several other genes like *TaHd1* involved in photoperiodic flowering pathway, *Ha* for grain hardness, and *TaBx* for benzoxazinone biosynthesis [207–209].

A prebreeding program in wheat (along with barley and canola) based on epigenetically modified genes has also been initiated in Australia at CSIRO, under the leadership of Dr. Liz Dennis and Dr. Jim Peacock, with the support from Dr. Ben Trevaskis (http://www.grdc.com.au/director/events/groundcover?item_id=A5B55D1DED8B9C20860C0CDE8C6EE077&article_id=A97C28B1F1614E34835D6BDB8CBDC75C). This pioneering work will involve vernalization, the mechanism that allows winter crops to avoid flowering until spring, when long days and mild conditions favor seed setting and grain filling. They plan to breed varieties with a wider range of heading dates and improved frost tolerance during flowering. In wheat (as also in other cereals), the epigenetic component is also built around *VRN1* gene, which plays a role analogous to that of *Flowering Locus C* (*FLC*) in *Arabidopsis* and canola. *VRN1* is one of the most important determinants of heading dates in winter cereals including wheat and also accounts for difference between winter and spring wheat varieties. It has been shown that during vegetative growth, *VRN1* is repressed epigenetically; this repression is lifted in spring, allowing the protein encoded by *VRN1* to activate other genes involved in reproduction. As many as ~3000 wheat varieties are being looked at for variation in their *VRN1* gene so as to breed better combinations of heading date and frost tolerance (http://www.grdc.com.au/director/events/groundcover?item_id=A5B55D1DED8B9C20860C0CDE8C6EE077&article_id=A97C28B1F1614E34835D6BDB8CBDC75C).

Wheat allopolyploidy and epigenetics

Polyploidization induces genetic and epigenetic modifications in the genomes of higher plants including wheat (reviewed in [210, 211]). Elimination of noncoding and low-copy DNA sequences has been reported in synthetic allopolyploids of *Triticum* and *Aegilops* species [212–214]. In two other studies, patterns of cytosine methylation were also examined throughout the genome in two synthetic allotetraploids, using methylation-sensitive amplification polymorphism (MSAP; [215, 216]). This analysis indicated that the parental patterns of methylation were altered in the allotetraploid in 13% of the genomic DNA analyzed. Gene silencing and activation were also observed when 3072 transcribed loci were analyzed, using cDNA-AFLP [217, 218]. This study demonstrated new, nonadditive patterns of gene expression in allotetraploid, as indicated by the fact that 48 transcripts disappeared and 12 transcripts that were absent in the diploid parents, appeared in the allotetraploid. These results were found reproducible in two independent synthetic allotetraploids. The disappearance of transcripts could be related to gene silencing rather than gene loss and was partly associated with cytosine methylation. In another similar study involving artificially synthesized hexaploid wheats and their parents, down-regulation of some genes and activation of some other genes, selected in a nonrandom manner, was observed [219]. The genome-wide genetic and epigenetic alterations triggered by allopolyploidy thus suggested plasticity of wheat genome. The reproducibility of genetic and epigenetic events indicated a programmed rather than a chaotic response and suggests that allopolyploidy is sensed in a specific way that triggers specific response rather than a random mutator response [218].

14. QUANTITATIVE TRAIT LOCI (QTL) AND PROTEIN QUANTITATIVE LOCI (PQLs) IN WHEAT

A large number of QTL studies for various traits have been conducted in bread wheat, leading to mapping of QTL for these traits on different chromosomes. In most of these studies, either single marker regression approach or QTL interval mapping has been utilized. Although most of these studies involved mapping of QTL with main effects only, there are also reports of QTL, which have no main effects but have significant digenic epistatic interactions and QTL × environment interactions [220–222]. A detailed account of studies involving gene tagging and QTL analyses for various traits conducted in wheat is available elsewhere [14, 223]. More up-to-date accounts on QTL studies (summarized in Table 5) are also available for disease resistance [224], for resistance against abiotic stresses [225], grain size, and grain number [226], and for several other traits including yield and yield contributing characters, plant type, and flowering time [222, 227]. Advanced backcross QTL (AB-QTL) analysis, proposed by Tanksley and Nelson [228], has also been utilized in wheat to identify QTL for a number of traits including yield and yield components, plant height, and ear emergence [129, 229]. More recently AB-QTL analysis was practiced for the identification of QTL for baking quality traits in two BC_2F_3 populations of winter wheat [230].

Quantitative variation in protein spots was also used for detection of protein quantitative loci (PQL) in wheat. For instance, in a study, 170-amphiphilic protein spots that were specific to either of the two parents of ITMIpop were used for genotyping 101 inbred lines; 72 out of these 170 proteins spots were assigned to 15 different chromosomes, with highest number of spots mapped to Group-1 chromosomes. QTL mapping approaches were also used to map PQL; 96 spots out of the 170 specific ones showed at least one PQL. These PQL were distributed throughout the genome. With the help of MALDI-TOF spectrometry and database search, functions were also assigned to 93 specific and 41 common protein spots. It was shown in the above study that majority of these proteins are associated with membranes and/or play a role in plant defense against external invasions [231].

15. RECENT INSIGHTS INTO THE ORIGIN/EVOLUTION OF WHEAT GENOMES

In the genomics era, the subject of origin and evolution of bread wheat has also been revisited. This gave new insights into the identity of progenitors of the three subgenomes (A, B, D) of bread wheat, and into the genome alterations, which presumably accompanied the course of its evolution and domestication (see Figure 1). These aspects of evolution of bread wheat will be discussed briefly in this section.

15.1. Origin of A, B, and D subgenomes

As mentioned earlier, bread wheat is a segmental allo-hexaploid having three closely related subgenomes A, B, and D. Initial analysis of the three subgenomes of bread wheat was mainly based on studies involving chromosome pairing in interspecific hybrids, and karyotype analysis in bread wheat as well as in the probable donors of the subgenomes (for reviews, see [232–236]). However, more recently, molecular markers and DNA sequence data have been used for the analysis of these subgenomes (see [237–239]). As a result, we have known with some degree of certainty that *T. urartu* ($2n = 14$) is the donor of subgenome A and *Ae. tauschii* (synonyms, *T. tauschii, Ae. squarrosa*) is the donor of subgenome D; this has recently been confirmed through analysis of DNA sequences of two genes, namely, *Acc-1* (plastid acetyl-CoA carboxylase) and *Pgk-1* (plastid 3-phosphoglycerate kinase) [240]. In contrast to this, although *Ae. speltoides* was once considered as the probable donor of the B subgenome ([241], for a review, see [237]), studies carried out later showed that *Ae. speltoides* more closely resembles the subgenome G of *T. timopheevii* rather than to the subgenome B of bread wheat. DNA sequences of the above genes, *Acc-1* and *Pgk-1* also proved to be of no help in identification of the progenitor of the subgenome B. There is, thus still no unanimity on the progenitor of the subgenome B of bread wheat (for more details, see [242]), and there are speculations that the donor of the subgenome B might have lost its identity during evolution and may never be discovered.

DNA sequences of genes other than the above two genes have also been used for the study of origin and evolution of the component subgenomes of bread wheat. For instance, in one such study, sequences from 14 loci (2 sequences from each of the 7 chromosomes) belonging to the subgenome B of bread wheat, when compared with those from five diploid species (from section Sitopsis) closely related to the B subgenome of bread wheat, indicated that the B subgenome of bread wheat and the genomes of the above five diploid species diverged greatly after the origin of tetraploid wheat [243]. The above study also received support from the recent evidence of independent origins of wheat B and G subgenomes [244]. In this study, 70 AFLP loci were used to sample diversity among 480 wheat lines collected from their natural habitats, which encompassed the entire range of habitats for all S genome *Aegilops* species. Also, a comparison of 59 *Aegilops* representatives of S genome diversity with 2x, 4x chromosome number, and 11 nulli-tetrasomic wheat lines at 375 AFLP loci suggested that B genome chromosomes of 6x wheat were derived from chromosomes of *Ae. speltoides*, and no other species. Further, an analysis of the haplotypes at nuclear and chloroplast loci *ACC1*, *G6PDH*, *GPT*, *PGK1*, *Q*, *VRN1*, and *ndhF* for ∼70 *Aegilops* and *Triticum* lines (0.73 Mb sequenced) revealed that both B and G genomes of polyploid wheats are unique samples of *A. speltoides* haplotype diversity. However, it is likely that due to the outbreeding nature of *A. speltoides*, no modern *A. speltoides* lines have preserved the B donor genotype in its ancestral state. The above findings can be incorporated into a broader scheme of wheat genome evolution (see Figure 1) with resolved positions of the B genome relative to S progenitors and G sisters. Similar analysis of the D subgenome and its progenitor showed that the D subgenome had more than one allele for a single locus derived from a progenitor, suggesting that hexaploid wheat perhaps originated from tetraploid wheat more than once utilizing different sources of *Ae. tauschii* [245]. Also, it was realized that major part of the large genome (16000 Mb) of bread wheat is composed of transposable elements (TEs). Therefore, the role of TEs in the evolution of bread-wheat and allied genomes has also been examined [246, 247]. In these studies, some specific

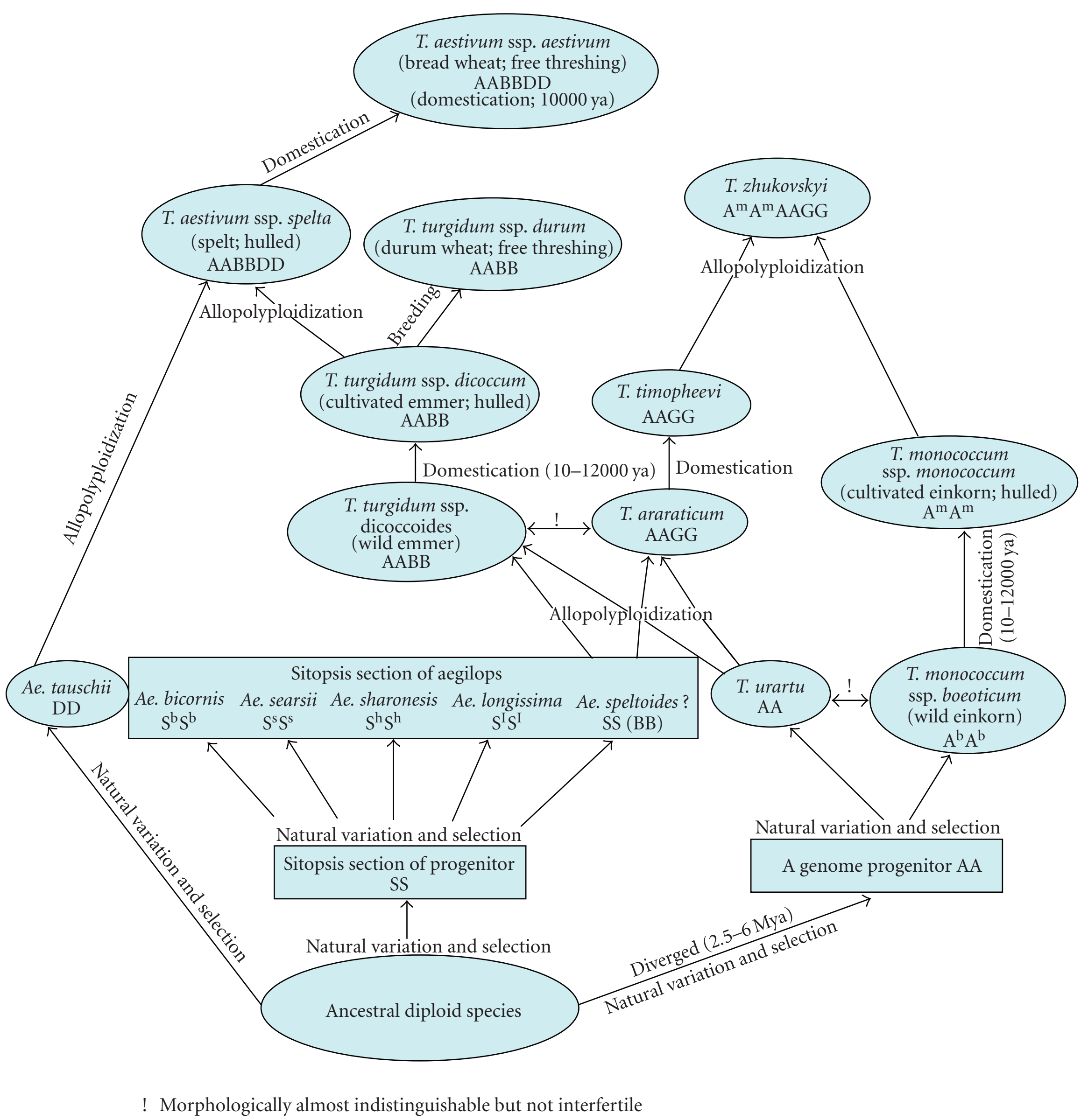


Figure 1: Schematic representation of the evolutionary history of wheat species (*Triticum* and *Aegilops*).

sequences from A and B genomes of diploid species were located, respectively, in B- and A-subgenomes of bread wheat, suggesting the role of TEs in transfer of sequences between A and B subgenomes. A bioinformatics approach was also used on a large genomic region (microgenomic approach) sequenced from *T. monococcum* (AA) and *Ae. tauschii* (DD). This approach allowed a comparison of variation within coding regions with that in the noncoding regions of the subgenomes.

15.2. Alterations that accompanied domestication

Domestication of most crop plants including wheat involved transition from short day, small-seeded plants with natural seed dispersal to photoperiod insensitive, large-seeded nonshattering plants. A study of genetic loci underlying domestication-related traits in *T. dicoccoides* was also conduced [430], where seven domestication syndrome factors (DSFs) were proposed, each affecting 5–11 traits. Following conclusions were made with respect to the domestication-related QTL. (i) Some of these QTL had strong effect and were clustered. (ii) Strong QTL were mainly associated with GRRs, where recombination rates are high. (iii) These QTL predominantly occurred in the A genome, suggesting that A genome has played a more important role than the B genome in evolution during domestication; this is understandable, because einkorn diploid wheat (*T. monococcum*) carrying the A genome was the first wheat to be domesticated, so that most of the domestication related traits in different wheats must have been selected within the A genome. Similar studies

Table 5: A list of gene/QTL tagged/mapped in wheat. RSL = recombinant substitution line; CSL = chromosome substitution line; RIL = recombinant inbred lines; DH = double haploid; RICL = recombinant inbred chromosome lines; SCRI = single-chromosome recombinant lines; AL = addition lines; BIL = backcross inbred lines; NIL = near isogenic lines; TC = test cross.

Trait	Gene/QTL (chromosome)	Mapping population	Reference
Disease			
(i) Leaf rust resistance	*Lr9* (6BL)	NILs	[248]
	Lr1 (5DL)	F_2	[249]
	Lr24 (3DL)	F_2	[250]
	Lr10 (1AS)	F_2	[251]
	Lr28 (4AL)	$F_{2:3}$	[252]
	Lr3 (6BL)	F_2	[253]
	Lr35 (2B)	F_2	[254]
	Lr47 (7A)	BC_1F_2	[255]
	LrTr (4BS)	F_2	[256]
	Lr19 (7DL)	Deletion lines	[257]
	Lr39 (=*Lr41*)(2DS)	F_2	[258]
	Lr37 (2AS)	NILs	[259]
	Lr20 (7AL)	F_2	[260]
	Lr19 (7D)	F_2	[261]
	Lr21/*Lr40* (1DS)	F_2	[88]
	Lr1 (5DL)	$F_{2:3}$ families	[262]
	Lr28 (-)	$F_{2:3}$	[263]
	Lr34 (7D)	RILs	[264]
	Lr52 (*LrW*) (5B)	F_2	[265]
	Lr16 (2BS)	DH	[50]
	Lr19 (7DL)	F_2	[266]
	Lr24 (3DL)	F_2	[266]
	Lr34 (7DS)	RILs	[267]
	Lr22a (2DS)	F_2	[268]
	Lr1 (5DL)	RILs	[269]
	Unknown (5B)	$F_{2:3}$ lines	[270]
	QTL (7D, 1BS)	RILs	[271]
	QTL (2D, 2B)	F_2	[272]
	QTL (7DS, linked to *Lr34*)	RILs	[273]
(ii) Stripe rust resistance	*Yr15* (1B)	F_2	[274]
	YrH52 (1B)	F_2	[275]
	Yrns-B1 (3BS)	F_3 lines	[276]
	Yr15 (1B)	F_2 lines	[277]
	Yr28 (4DS)	RILs	[273]
	Yr9 (1B/1R)	$BC_7F_{2:3}$	[278]
	Yr17 (2A)	NILs	[259]
	Yr26 (1BS)	F_2 lines	[279]
	Yr10 (1B)	F_2 lines	[280]
	Yr5 (2B)	BC_7F_3	[131]
	Yr18 (7D)	RILs	[264]
	Yr36 (6B)	RILs	[281]
	YrCH42 (1B)	F_2	[282]
	YrZH84 (7BL)	F_2, F_3	[283]
	Yr34 (5AL)	DH	[284]
	Yr26 (1B)	$F_{2:3}$ lines	[285]

Table 5: Continued.

Trait	Gene/QTL (chromosome)	Mapping population	Reference
	QTL (2D, 5B, 2B, 2A)	RILs	[286]
	QTL (2AL, 2AS, 2BL, 6BL)	DH	[287]
(iii) Stem rust resistance	*Sr22* (7A)	F_2	[288]
	Sr38 (2AS)	NILs	[259]
	Sr2 (3BS)	F_3 lines	[289]
(iv) Fusarium head blight resistance	*Fhb2* (6BS)	RILs	[290]
	QTL (5A, 3B, 1B)	DH	[37]
	QTL (3BS, 3A, 5B)	RILs	[291]
	QTL (3B)	Advanced lines	[292]
	QTL (3B, 6B, 2B)	RILs	[293, 294]
	QTL (6D, 4A, 5B)	RILs	[295]
	QTL (3A, 5A)	DH	[43]
	QTL (1B, 3B)	RILs	[30]
	QTL (2B)	RILs	[296]
	QTL (3B)	DH	[297]
	QTL (6AL, 1B, 2BL, 7BS)	RILs	[46]
	QTL (3A)	RICLs	[298]
	QTL (4D)	DH	[49]
	QTL (3BS, 5AS, 2DL)	RILs	[299]
	QTL (1BS, 1DS, 3B, 3DL, 5BL, 7BS, 7AL)	RILs	[300]
	QTL (7E)	RILs	[301]
(v) Scab resistance	QTL (2AS, 2BL, 3BS)	RILs	[302]
	QTL (3BS)	$F_{3:4}$ lines	[303]
(vi) Powdery mildew resistance	*Pm2* (5DS)	F_2	[304]
	Pm18 (5DS)	F_2	[304]
	Pm12 (6B)	F_2	[305]
	Pm21 (6AL)	BC lines	[306]
	Pm3g (1A)	DH	[307]
	Pm24 (1DS)	$F_{2:3}$ lines	[308]
	Pm26 (2BS)	RSI	[309]
	Pm6 (2BL)	NILs	[310]
	Pm27 (6B)	F_2	[311]
	Pm8/Pm17 (1BL)	F_3 families	[312]
	Pm3 (1AS)	RILs	[313]
	Pm1 (7AL)	F_2	[260]
	Pm29 (7D)	F_2 & F_4 lines	[314]
	Pm30 (5BS)	BC_2F_2 lines	[315]
	Pm13 (3S)	AL	[316]
	Pm5e (7BL)	F_2	[130]
	Pm4a (2A)	F_2	[317]
	PmU (7AL)	F_2	[318]
	Pm34 (5D)	$F_{2:3}$ lines	[319]
	PmY39 (2B)	$BC_3F_{4:5}$	[320]
	Pm35 (5DL)	$F_{2:3}$ lines	[321]

Table 5: Continued.

Trait	Gene/QTL (chromosome)	Mapping population	Reference
	Pm5d (7BL)	F_3 lines	[322]
	Pm12 (6B)	BC_3F_2	[323]
	MlRE, QTL (6A, 5D)	F_3 lines	[324]
	MlG (6AL)	BC_2F_3	[325]
	mlRD30 (7AL)	F_2	[326]
	Mlm2033, Mlm80 (7A)	F_2	[181]
	QTL (5A, 7B, 3D)	RILs	[327]
	QTL (1B, 2A, 2B)	$F_{2:3}$ lines	[328]
	QTL (2B, 5D, 6A)	DH	[329]
	QTL (2B)	F_2	[330]
	QTL (1BL, 2AL, 2BL)	RILs	[331]
(vii) Common bunt resistance	*Bt-10* (-)	F_2	[332]
	QTL (1B, 7A)	DH	[333]
(viii) Tan spot and *Stagonospora nodorum* blotch resistance	QTL (1A, 4A, 1B, 3B)	RILs	[334]
	QTL (5B, 3B)	Inbred, CS lines	[335]
	tsn3a, *tsn3b*, *tsn3c* (3D)	$F_{2:3}$ lines	[336]
(ix) Septoria tritici blotch resistance	*Stb5* (7D)	SCRI	[337]
	QTL (3A)	DH	[38]
	QTL (1D, 2D, 6B)	RILs	[338]
(x) Barley yellow dwarf tolerance	QTL (12 chromosomes)	RILs	[339]
(xi) Leaf and glume blotch resistance	QTL (4B, 7B, 5A)	RILs	[340]
(xii) Wheat streak mosaic virus resistance	*Wms1* (4D)	F_2	[341]
	WSSMV (2DL)	RILs	[342]
(xiii) Yellow mosaic virus resistance	*YmYF* (2D)	F_2	[343]
(xiv) Eyespot (straw breaker foot rot) resistance	*Pch2* (7AL)	F_2	[344]
	Pch1 (7A)	F_3 lines	[345]
	Pch1, Ep-D1 (7D)	TC	[346]
Insect-pest			
(i) Green bug resistance	*Gb3* (7D)	$F_{2:3}$ lines	[347]
	Gby (7A)	$F_{2:3}$ lines	[348]
	Gb7 (7DL)	RILs	[349]
	Gb (7DL)	$F_{4:5}$ lines, F_2	[350]
(ii) Hessian fly resistance gene	*H23* (6D)	F_2	[351]
	H24 (3D)	F_2	[351]
	H3, H6, H9, H10, H12, H16, H17 (5A)	NILs, F_2	[352]
	H5, H11, H13, H14 (1A)	NILs, F_2	[353]
	H21 (2B)	NILs, F_2	[354]
	H6 (-)	F_2	[355]
	H13 (6DS)	$F_{2:3}$	[356]
	H26, H13 (3D, 6D)	$F_{2:3}$ lines	[357]
	H22 (1D)	$F_{2:3}$ lines	[358]
	H16 and H17 (1AS)	BC_1F_2, $F_{2:3}$ lines	[359]

Table 5: Continued.

Trait	Gene/QTL (chromosome)	Mapping population	Reference
(iii) Russian wheat aphid resistance	*Dn8, Dn9* (7DS, 1DL)	F_2	[360]
	Dn1, Dn2, Dn5, Dn8	F_2	[360]
	Dnx (7DS)		
	Dn2 (7DS)	F_2	[361]
	Dn4 (1D)	F_2	[362]
	Dn6 (7D)	F_2	[362]
Nematodes			
(i) Cereal cyst nematode resistance	*Cre1* (2B)	NILs, F_2	[363]
	Cre5 (2AS)	NILs	[364]
	Cre6 (5A)	F_2	[365]
	QTL (1B)	DH	[48]
(ii) Root-knot nematode resistance	*Rkn-mn1* (3BL)	BC_3F_2, F_3 lines	[366]
(iii) Root-lesion nematode resistance	*Rlnn1* (7AL)	DH	[367]
Quality and quality related traits			
(i) Seed dormancy or preharvest sprouting	QTL (4A)	DH	[368]
	QTL (4A)	RILs, DH	[369]
	QTL (3A)	BC_1F_2	[370]
	QTL (3A)	RILs	[371]
	QTL (3A)	RILs	[372]
	QTL (4A)	DH	[373]
(ii) Grain protein content	QTL (6B)	RILs	[374]
	QTL (2A, 3A, 4D, 7D, 2B, 5B, 7A)	RILs	[39]
	QTL (2A, 2B, 2D, 3D, 4A, 6B, 7A, 7D)	RILs	[375]
	QTL (2AS, 6AS, 7BL)	BILs	[376]
(iii) Others			
Flour colour	QTL (3A, 7A)	RILs	[377]
Milling yield	QTL (3A, 7D)	RILs	[378]
Bread-making quality	QTL (5DS, 1B, 6A, 3B, 1A)	DH	[379]
Milling traits	QTL (7A, 6B)	RILs	[35]
Grain dry matter and N accumulation, protein composition	QTL (1A, 2B, 3A, 6A, 5A, 7A, 7D)	RILs	[380]
Mixograph-extensibility	QTL (5A)	DH	[381]
Kernel hardness and dough strength	QTL (1A, 5D, 1B, 1D, 5B)	Inbred lines	[382]
Purple grain colour	*Pp1, Pp3b, Pp3a* (2A, 7BL)	F_2	[383]
Quality traits	QTL (5DS, 6DS, 2DS, 1AS, 1BS, 6DS)	RILs	[384]
Low-molecular-weight glutenin	*LMW-GS* (-)	F_5 lines	[385]
Bread-making quality	QTL (3A, 7A)	RILs	[40]
Milling and baking quality	QTL (4B, 6D)	BC_2F_3	[230]
Endosperm colour	QTL, *Psy1-1* (2A, 4B, 6B, 7B)	DH	[386]

Table 5: Continued.

Trait	Gene/QTL (chromosome)	Mapping population	Reference
Agronomic traits			
(i) Plant height	*Rht-B1*, *Rht-D1* (4BS, 4DS)	DH	[387]
	Rht8 (2D)	RILs	[388]
	Rht8 (2DS)	DH, Inbred lines	[389]
(ii) Tiller inhibition gene	*tin3* (3A)	F_2	[390]
(iii) Spherical grain and compact spikes	*s16219, C17648 B1* (3B, 5A)	F_2	[391]
(iv) Ear-emergence time and plant height	QTL (5A)	RILs	[392]
(v) Heading date	QTL (2BS)	DH	[393]
	QTL, *Ppd-B1, Ppd-D1* (2B, 2D, 5A, 2B)	RILs	[394]
	QTL (2DS)	RILs	[395]
	QTL (2A, 2B, 2D, 5A, 5B, 5D, 4A, 4B)	RILs	[396]
(vi) Grain yield and related traits	QTL (5A)	RILs	[397]
	QTL (2D, 3B, 3D, 5D, 7D)	$BC_2F_{2:4}$ lines	[398]
	QTL (1D, 2A, 6B, 7D)	RILs	[51]
	QTL (7AL, 7BL, 1D, 5A)	DH	[45]
	QTL (4AL)	RILs	[399]
	QTL (1B, 4D, 7D)	DH	[400]
(vii) Spike-related traits	QTL (7D)	F_2	[401]
(viii) Grain weight	QTL (1A, 2B, 7A)	RILs	[402]
(ix) Others	QTL (4A, 4B, 4D, 7D, 3B, 3D)	DH	[171]
	QTL (1D, 4D)	DH	[47]
Growth related traits			
(i) Spike morphology, awn development, vernalization	*B, Q, VRN1* (5A)	RILs	[403]
(ii) Supernumerary spikelet	*bh* (2D, 4A, 4B, 5A)	$F_{2:3}$ lines	[404]
(iii) Sphaerococcum-like growth habit	*S1, S2, S3* (3D, 3B, 3A)	F_2	[405]
(iv) Thermosensitive earliness	*Eps-Am1* (1AL)	F_2	[406]
(v) Coleoptiles pigmentation	*Rc-A1*, *Rc-B1*, *Rc-D1* (7A, 7B, 7D, 4BL)	RILs	[407]
(vi) Thermosensitive genic male-sterile	*wtms1* (2B)	F_2	[408]
(vii) Hybrid necrosis	*Ne1, Ne2* (5BL, 2BS)	F_2	[409]
(viii) Leaf pubescence and hairy leaf	*Hl1, Hl2, Aesp*, QTL (4BL, 7BS)	F_2	[410]
(ix) Stem solidness	*sst* (3BL)	DH	[411]
(x) Lodging resistance	QTL (1BS, 2AS, 2D, 3AS, 4AS, 5AL, 5BL, 6BL, 7BL)	RILs	[412]
(xi) Stem strength and related traits	QTL (3A, 3B, 1A, 2D)	DH	[413]

Table 5: Continued.

Trait	Gene/QTL (chromosome)	Mapping population	Reference
(xii) Brittle rachis	QTL (3A, 3B)	RICLs	[414]
(xiii) Coleoptiles growth	QTL (2B, 2D, 4A, 5D, 6B)	DH	[415]
(xiv) Kernel shattering	QTL (2B, 3B, 7A)	RILs	[416]
(xv) Seed development	QTL (1D, 4B)	DH	[9]
(xvi) Longer coleoptiles	QTL (6A)	RILs	[29]
(xvii) Viridescent phenotype	QTL (2B)	DH	[52]
Biochemical			
(i) Casein kinase	*CK2α* (5A)	F_2	[417]
(ii) Nonglaucousness	*Iw3672* (2DS)	F_2	[418]
(iii) Low lipoxygenase	*Lpx-B1.1, Lpx-A3* (4B, 4A)	RILs	[419]
(iv) Polyphenol oxidase (PPO) genes	*PPO* (2A, 2D)	DH	[420]
(v) ABA signaling genes	QTL (3A, 5A)	RILs	[421]
(vi) Polyphenol oxidase	QTL (2A)	DH	[422]
(vii) Water-soluble carbohydrates	QTL (21 chromosome)	DH	[423]
Abiotic stress			
(i) Photoperiod insensitive	*Ppd-B1* (2BS)	RILs, DH	[424]
(ii) Aluminum tolerance	*ALMT1* (4D)	DH	[425]
	QTL (4D, 3BL)	RIL	[426]
(iii) Boron toxicity tolerance	QTL (*Bo1*) (7BL)	DH	[427]
(iv) Frost resistance	QTL (5B)	RSI	[428]
(v) Salt tolerance	QTL (3A, 3B, 4DL, 6DL)	RILs	[429]

involving study of evolution during domestication were also conducted in hexaploid wheats for seed size, free threshing habit, rachis stiffness, photoperiod insensitivity, and so forth (for a review, see [431]). In wheat, a primary component of domestication syndrome was the loss of spike shattering, controlled by *Br* (brittle rachis) loci on chromosome 3A and 3B [414]. Other traits of wheat domestication syndrome shared by all domesticated wheats are the soft glumes, increased seed size, reduced number of tillers, more erect growth, and reduced dormancy [432]. A gene *GPC-B1*, which is an early regulator of senescence with pleiotropic effects on grain nutrient content, has also been found to affect seed size [96]. However, in some genotypes and environments, the accelerated grain maturity conferred by functional *GPC-B1* allele has been found associated with smaller seeds [433], suggesting that indirect selection for large seeds may explain the fixation of the nonfunctional *GPC-B1* allele in both durum and bread wheats [96]. Among many genes relevant to wheat domestication syndrome, only *Q* and *GPC-B1* have been successfully isolated so far, suggesting a need for systematic effort to clone other genes, since it is possible that genetic variation at these loci might have played an important role in the success of wheat as a modern crop.

16. APPLICATION OF GENOMICS TO MOLECULAR BREEDING OF WHEAT

16.1. *Association mapping in wheat*

Association mapping is a high-resolution method for mapping QTL based on linkage disequilibrium (LD) and holds great promise for genetic dissection of complex traits. It offers several advantages, which have been widely discussed [434, 435]. In wheat, some parts of the genome relative to other parts are more amenable to LD/association mapping for QTL detection and fine mapping, since the level of LD is variable across the length of a chromosome. As we know, LD decay over longer distances will facilitate initial association of trait data with the haplotypes in a chromosome region and

LD decay over short distances will facilitate fine mapping of QTL [12].

Several studies involving association mapping in wheat have been conducted in the recent past. For instance, association mapping has been conducted for kernel morphology and milling quality [436] and for the quantity of a high-molecular-weight glutenin [141, 437]. In another study, 242 diversity array technology (DArT) markers were utilized for association mapping of genes/QTL controlling resistance against stem rust (SR), leaf rust (LR), yellow rust (YR), powdery mildew (PM), and those controlling grain yield (GY). Phenotypic data from five historical CIMMYT elite spring wheat yield trials (ESWYT) conducted in a large number of international environments were utilized for this purpose and two linear mixed models were applied to assess marker-trait associations after a study of population structure and additive genetic covariance between relatives [438]. A total of 122, 213, 87, 63, and 61 DArT markers were found to be significantly associated with YR, GY, LR, SR, and PM, respectively. Association analysis was also conducted between markers in the region of a major QTL responsible for resistance to *Stagonospora nodorum* (causing glume blotch); it was concluded that association mapping had a marker resolution, which was 390-fold more powerful than QTL analysis conducted using an RIL mapping population [439]. Such high-resolution mapping of traits and/or QTL to the level of individual genes, using improved statistical methods, will provide new possibilities for studying molecular and biochemical basis of quantitative trait variation and will help to identify specific targets for crop improvement.

16.2. Marker-assisted selection in wheat

A large number of marker-trait associations determined during the last decades facilitated the use of molecular markers for marker-assisted selection (MAS) in bread wheat, which is gaining momentum in several countries. In particular, major programs involving MAS in wheat are currently underway in USA, Australia, and at CIMMYT in Mexico. In USA, a wheat MAS consortium comprosing more than 20 wheat-breeding programs was constituted at the end of 2001. The objective of this consortium was to apply and to integrate MAS in public wheat breeding programs [440]. Under these programs, MAS has been utilized for transfer of as many as 27 different insect and pest resistance genes and 20 alleles with beneficial effects on bread making and pasta quality into ~180 lines adapted to the primary US production regions. These programs led to release of germplasm consisting of 45 MAS-derived lines [441]. Similarly, the program in Australia involved improvement of 20 different traits (including resistance to some abiotic stresses) and has already led to release of some improved cultivars ([442], Peter Langridge personal communication). Among these traits, MAS has become a method of choice for those agronomically important traits, where conventional bioassays were expensive and unconvincing, as was the case in selection for cereal cyst nematodes resistance carried out by Agriculture Victoria [443]. In addition to this, MAS has been incorporated in backcross breeding in order to introgress QTL for improvement of transpiration efficiency and for negative selection for undesirable traits such as yellow flour color [444]. Australian scientists also conducted a computer simulation in order to design a genetically effective and economically efficient marker-assisted wheat-breeding strategy for a specific outcome. This investigation involved an integration of both restricted backcrossing and doubled haploid (DH) technology. Use of MAS at the BC_1F_1 followed by MAS in haploids derived from pollen of BC_1F_1 (prior to chromosome doubling) led to reduction of cost of marker-assisted breeding up to 40% [445]. Later, this MAS strategy was validated practically in a marker-assisted wheat-breeding program in order to improve quality and resistance against rust disease (for review, see [446]). At CIMMYT, markers associated with 25 different genes governing insect pest resistance, protein quality, homoeologous pairing, and other agronomic characters are currently being utilized in wheat breeding programs in order to develop improved wheat cultivars [447]. Some of the markers used in these programs are perfect markers that have been developed from available nucleotide sequences of these genes. In future, large-scale sequencing of GRRs (gene-rich regions), to be undertaken by IWGSC, will also facilitate isolation of important genes for production of improved transgenic crops, and for development of "perfect markers" for agronomically important traits to be used in MAS [448, 449].

17. ORGANELLAR GENOMES AND THEIR ORGANIZATION

The genomes of wheat chloroplast and mitochondrion have also been subjected to a detailed study during the last decade. The results of these studies will be briefly discussed in this section.

17.1. Chloroplast genome

In bread wheat, 130–155 chloroplasts, each containing 125–170 circular DNA molecules (135 kb), are present in each mesophyll cell, thus making 16000–26000 copies of cpDNA within a cell. This makes 5–7% of the cellular DNA in the leaf and 10–14% of the DNA in a mesophyll cell. In the related diploid species, there are 4900–6600 copies and in tetraploid species, there are 9600–12400 copies of cpDNA per mesophyll cell.

The wheat chloroplast genome, like all other plant chloroplast genomes, has two inverted repeat regions, each copy (21-kb-long) separated from the other by two single copy regions (12.8 kb, 80.2 kb). The gene content of wheat chloroplast is the same as those of rice and maize plastomes, however some structural divergence was reported in the gene coding regions, due to illegitimate recombination between two short direct repeats and/or replication slippage; this included the presence of some hotspot regions for length mutations. The study of deletion patterns of open reading frames (ORFs) in the inverted-repeat regions and in the borders between the inverted repeats and the small single-copy regions supports the view that wheat and rice are related more closely to each other than to maize (see [450, 451]).

Deletions, insertions, and inversions have also been detected during RFLP analysis of cpDNA, which gave eleven different cpDNA types, in the genus *Triticum*, the bread wheat sharing entirely the cpDNA type with durum wheats, but not with that of any of the diploid species. The cpDNA of *Ae. speltoides* showed maximum similarity to those of *T. aestivum*, *T. timopheevii*, and *T. zhukovskyi*, suggesting that *Ae. speltoides* should be the donor of the B subgenome of common wheat [452].

17.2. Mitochondrial genome

Wheat mtDNA is larger (430 kb) than cpDNA (135 kb) with a minimum of 10 repeats but encodes only 30–50% polypeptides relative to cpDNA. Thus, large amount of mtDNA is noncoding, there being about 50 genes involved in RNA synthesis [453]. Mitochondrial genome of Chinese Spring has been sequenced using 25 cosmid clones of mitochondrial DNA, selected on the basis of their gene content. This led to the identification of 55 (71) genes including the following: 18 genes (20) for electron transport system, 4 genes for mitochondrial biogenesis, 11 genes for ribosomal proteins, 2 genes for splicing and other function, 3 genes (10) for rRNAs, and 17 genes (24) for tRNAs (the numerals in parentheses represent number of genes, taking multiple copies of a gene as separate genes). When mitochondrial gene maps were compared among wheat, rice, and maize, no major synteny was found between them other than a block of two to five genes. Therefore, mitochondrial genes seem to have thoroughly reshuffled during speciation of cereals. In contrast, chloroplast genes show perfect synteny among wheat, rice, and maize [451].

18. CONCLUSIONS

Significant progress during the last two decades has been made in different areas of wheat genomics research. These include development of thousands of molecular markers (including RFLPs, SSRs, AFLPs, SNPs, and DArT markers), construction of molecular genetic and physical maps (including radiation hybrid maps for some chromosomes) with reasonably high density of markers, development of more than 1 million ESTs and their use for developing functional markers, and the development of BAC/BIBAC resources for individual chromosomes and entire subgenomes to facilitate genome sequencing. Functional genomics approaches like TILLING, RNAi, and epigenetics have also been utilized successfully, and a number of genes/QTL have been cloned to be used in future wheat improvement programs. Organellar genomes including chloroplast and mitochondrial genomes have been fully sequenced, and we are at the threshold of initiating a major program of sequencing the gene space of the whole nuclear genome in this major cereal. The available molecular tools also facilitated a revisit of the wheat community to the problem of origin and evolution of the wheat genome and helped QTL analysis (including studies involving LD and association mapping) for identification of markers associated with all major economic traits leading to the development of major marker-aided selection (MAS) programs for wheat improvement in several countries.

ACKNOWLEDGMENTS

The authors would like to thank the Indian National Science Academy (INSA) for award of an INSA Honorary Scientist Position to PKG, Department of Biotechnology (DBT) for providing financial support to R. R. Mir, A. Mohan, J. Kumar, and also to the Head of Department of Genetics and Plant Breeding, Chaudhary Charan Singh University, Meerut, for providing the facilities.

REFERENCES

[1] B. S. Gill, R. Appels, A.-M. Botha-Oberholster, et al., “A workshop report on wheat genome sequencing: international genome research on wheat consortium,” *Genetics*, vol. 168, no. 2, pp. 1087–1096, 2004.

[2] E. R. Sears, “Nullisomic-tetrasomic combinations in hexaploid wheat,” in *Chromosome Manipulation and Plant Genetics*, R. Riley and K. R. Lewis, Eds., pp. 29–45, Oliver and Boyd, Edinburgh, UK, 1966.

[3] T. R. Endo and B. S. Gill, “The deletion stocks of common wheat,” *Journal of Heredity*, vol. 87, no. 4, pp. 295–307, 1996.

[4] K. S. Gill, “Gene distribution in cereal genomes,” in *Cereal Genomics*, P. K. Gupta and R. K. Varshney, Eds., pp. 361–385, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2004.

[5] K. Singh, M. Ghai, M. Garg, et al., “An integrated molecular linkage map of diploid wheat based on a *Triticum boeoticum* × *T. monococcum* RIL population,” *Theoretical and Applied Genetics*, vol. 115, no. 3, pp. 301–312, 2007.

[6] C. Feuillet and B. Keller, “Comparative genomics in the grass family: molecular characterization of grass genome structure and evolution,” *Annals of Botany*, vol. 89, no. 1, pp. 3–10, 2002.

[7] M. D. Gale and K. M. Devos, “Plant comparative genetics after 10 years,” *Science*, vol. 282, no. 5389, pp. 656–659, 1998.

[8] K. M. Devos, “Updating the ‘crop circle’,” *Current Opinion in Plant Biology*, vol. 8, no. 2, pp. 155–162, 2005.

[9] M. C. Jordan, D. J. Somers, and T. W. Banks, “Identifying regions of the wheat genome controlling seed development by mapping expression quantitative trait loci,” *Plant Biotechnology Journal*, vol. 5, no. 3, pp. 442–453, 2007.

[10] M. Bagge, X. Xia, and T. Lübberstedt, “Functional markers in wheat,” *Current Opinion in Plant Biology*, vol. 10, no. 2, pp. 211–216, 2007.

[11] P. Moolhuijzen, D. S. Dunn, M. Bellgard, et al., “Wheat genome structure and function: genome sequence data and the international wheat genome sequencing consortium,” *Australian Journal of Agricultural Research*, vol. 58, no. 6, pp. 470–475, 2007.

[12] D. J. Somers, “Molecular breeding and assembly of complex genotypes in wheat,” in *Frontiers of Wheat Bioscience. The 100 Memorial Issue of Wheat Information Service*, K. Tsunewaki, Ed., pp. 235–246, Kihara Memorial Yokohama Foundation for the Advancement of Life Sciences, Yokohama, Japan, 2005.

[13] S. Chao, P. J. Sharp, A. J. Worland, E. J. Warham, R. M. D. Koebner, and M. D. Gale, “RFLP-based genetic maps of wheat homoeologous group 7 chromosomes,” *Theoretical and Applied Genetics*, vol. 78, no. 4, pp. 495–504, 1989.

[14] P. K. Gupta, R. K. Varshney, P. C. Sharma, and B. Ramesh, "Molecular markers and their applications in wheat breeding," *Plant Breeding*, vol. 118, no. 5, pp. 369–390, 1999.

[15] R. Appels, "A consensus molecular genetic map of wheat-a cooperative international effort," in *Proceedings of the 10th International Wheat Genetics Symposium*, N. E. Pogna, Ed., pp. 211–214, Paestum, Italy, September 2003.

[16] D. J. Somers, P. Isaac, and K. Edwards, "A high-density microsatellite consensus map for bread wheat (*Triticum aestivum* L.)," *Theoretical and Applied Genetics*, vol. 109, no. 6, pp. 1105–1114, 2004.

[17] R. A. McIntosh, K. M. Devos, J. Dubcovsky, C. F. Morris, and W. J. Rogers, "Catalogue of gene symbols for wheat," 2003, http://wheat.pw.usda.gov/ggpages/wgc/2003upd.html.

[18] M. S. Röder, V. Korzun, K. Wendehake, et al., "A microsatellite map of wheat," *Genetics*, vol. 149, no. 4, pp. 2007–2023, 1998.

[19] E. Pestsova, M. W. Ganal, and M. S. Röder, "Isolation and mapping of microsatellite markers specific for the D genome of bread wheat," *Genome*, vol. 43, no. 4, pp. 689–697, 2000.

[20] P. K. Gupta, H. S. Balyan, K. J. Edwards, et al., "Genetic mapping of 66 new microsatellite (SSR) loci in bread wheat," *Theoretical and Applied Genetics*, vol. 105, no. 2-3, pp. 413–422, 2002.

[21] L. F. Gao, R. L. Jing, N. X. Huo, et al., "One hundred and one new microsatellite loci derived from ESTs (EST-SSRs) in bread wheat," *Theoretical and Applied Genetics*, vol. 108, no. 7, pp. 1392–1400, 2004.

[22] J.-K. Yu, T. M. Dake, S. Singh, et al., "Development and mapping of EST-derived simple sequence repeat markers for hexaploid wheat," *Genome*, vol. 47, no. 5, pp. 805–818, 2004.

[23] N. Nicot, V. Chiquet, B. Gandon, et al., "Study of simple sequence repeat (SSR) markers from wheat expressed sequence tags (ESTs)," *Theoretical and Applied Genetics*, vol. 109, no. 4, pp. 800–805, 2004.

[24] J. W. Snape and G. Moore, "Reflections and opportunities: gene discovery in the complex wheat genome," in *Wheat Production in Stressed Environments*, H. T. Buck, Ed., pp. 677–684, Springer, Dordrecht, The Netherlands, 2007.

[25] K. S. Gill, E. L. Lubbers, B. S. Gill, W. J. Raupp, and T. S. Cox, "A genetic linkage map of *Triticum tauschii* (DD) and its relationship to the D genome of bread wheat (AABBDD)," *Genome*, vol. 34, no. 3, pp. 362–374, 1991.

[26] E. V. Boyko, B. S. Gill, L. Mickelson-Young, et al., "A high-density genetic linkage map of *Aegilops tauschii*, the D-genome progenitor of bread wheat," *Theoretical and Applied Genetics*, vol. 99, no. 1-2, pp. 16–26, 1999.

[27] Q. J. Song, J. R. Shi, S. Singh, et al., "Development and mapping of microsatellite (SSR) markers in wheat," *Theoretical and Applied Genetics*, vol. 110, no. 3, pp. 550–560, 2005.

[28] A. Torada, M. Koike, K. Mochida, and Y. Ogihara, "SSR-based linkage map with new markers using an intraspecific population of common wheat," *Theoretical and Applied Genetics*, vol. 112, no. 6, pp. 1042–1051, 2006.

[29] W. Spielmeyer, J. Hyles, P. Joaquim, et al., "A QTL on chromosome 6A in bread wheat (*Triticum aestivum*) is associated with longer coleoptiles, greater seedling vigour and final plant height," *Theoretical and Applied Genetics*, vol. 115, no. 1, pp. 59–66, 2007.

[30] X. Zhang, M. Zhou, L. Ren, et al., "Molecular characterization of *Fusarium* head blight resistance from wheat variety Wangshuibai," *Euphytica*, vol. 139, no. 1, pp. 59–64, 2004.

[31] J. Dubcovsky, M.-C. Luo, G.-Y. Zhong, et al., "Genetic map of diploid wheat, *Triticum monococcum* L., and its comparison with maps of *Hordeum vulgare* L.," *Genetics*, vol. 143, no. 2, pp. 983–999, 1996.

[32] A. Blanco, M. P. Bellomo, A. Cenci, et al., "A genetic linkage map of durum wheat," *Theoretical and Applied Genetics*, vol. 97, no. 5-6, pp. 721–728, 1998.

[33] C. Lotti, S. Salvi, A. Pasqualone, R. Tuberosa, and A. Blanco, "Integration of AFLP markers into an RFLP-based map of durum wheat," *Plant Breeding*, vol. 119, no. 5, pp. 393–401, 2000.

[34] M. M. Nachit, I. Elouafi, M. A. Pagnotta, et al., "Molecular linkage map for an intraspecific recombinant inbred population of durum wheat (*Triticum turgidum* L. var. *durum*)," *Theoretical and Applied Genetics*, vol. 102, no. 2-3, pp. 177–186, 2001.

[35] I. Elouafi and M. M. Nachit, "A genetic linkage map of the Durum× *Triticum dicoccoides* backcross population based on SSRs and AFLP markers, and QTL analysis for milling traits," *Theoretical and Applied Genetics*, vol. 108, no. 3, pp. 401–413, 2004.

[36] M. M. Messmer, M. Keller, S. Zanetti, and B. Keller, "Genetic linkage map of a wheat×spelt cross," *Theoretical and Applied Genetics*, vol. 98, no. 6-7, pp. 1163–1170, 1999.

[37] H. Buerstmayr, M. Lemmens, L. Hartl, et al., "Molecular mapping of QTLs for Fusarium head blight resistance in spring wheat. I. Resistance to fungal spread (type II resistance)," *Theoretical and Applied Genetics*, vol. 104, no. 1, pp. 84–91, 2002.

[38] L. Eriksen, F. Borum, and A. Jahoor, "Inheritance and localisation of resistance to *Mycosphaerella graminicola* causing septoria tritici blotch and plant height in the wheat (*Triticum aestivum* L.) genome with DNA markers," *Theoretical and Applied Genetics*, vol. 107, no. 3, pp. 515–527, 2003.

[39] C. Groos, N. Robert, E. Bervas, and G. Charmet, "Genetic analysis of grain protein-content, grain yield and thousand-kernel weight in bread wheat," *Theoretical and Applied Genetics*, vol. 106, no. 6, pp. 1032–1040, 2003.

[40] C. Groos, E. Bervas, E. Chanliaud, and G. Charmet, "Genetic analysis of bread-making quality scores in bread wheat using a recombinant inbred line population," *Theoretical and Applied Genetics*, vol. 115, no. 3, pp. 313–323, 2007.

[41] S. Paillard, T. Schnurbusch, M. Winzeler, et al., "An integrative genetic linkage map of winter wheat (*Triticum aestivum* L.)," *Theoretical and Applied Genetics*, vol. 107, no. 7, pp. 1235–1242, 2003.

[42] P. Sourdille, T. Cadalen, H. Guyomarc'h, et al., "An update of the Courtot × Chinese Spring intervarietal molecular marker linkage map for the QTL detection of agronomic traits in wheat," *Theoretical and Applied Genetics*, vol. 106, no. 3, pp. 530–538, 2003.

[43] B. Steiner, M. Lemmens, M. Griesser, U. Scholz, J. Schondelmaier, and H. Buerstmayr, "Molecular mapping of resistance to *Fusarium* head blight in the spring wheat cultivar Frontana," *Theoretical and Applied Genetics*, vol. 109, no. 1, pp. 215–224, 2004.

[44] Z. H. Liu, J. A. Anderson, J. Hu, T. L. Friesen, J. B. Rasmussen, and J. D. Faris, "A wheat intervarietal genetic linkage map based on microsatellite and target region amplified polymorphism markers and its utility for detecting quantitative trait loci," *Theoretical and Applied Genetics*, vol. 111, no. 4, pp. 782–794, 2005.

[45] S. A. Quarrie, A. Steed, C. Calestani, et al., “A high-density genetic map of hexaploid wheat (*Triticum aestivum* L.) from the cross Chinese Spring × SQ1 and its use to compare QTLs for grain yield across a range of environments,” *Theoretical and Applied Genetics*, vol. 110, no. 5, pp. 865–880, 2005.

[46] M. Schmolke, G. Zimmermann, H. Buerstmayr, et al., “Molecular mapping of Fusarium head blight resistance in the winter wheat population Dream/Lynx,” *Theoretical and Applied Genetics*, vol. 111, no. 4, pp. 747–756, 2005.

[47] X. Q. Huang, S. Cloutier, L. Lycar, et al., “Molecular detection of QTLs for agronomic and quality traits in a doubled haploid population derived from two Canadian wheats (*Triticum aestivum* L.),” *Theoretical and Applied Genetics*, vol. 113, no. 4, pp. 753–766, 2006.

[48] K. J. Williams, K. L. Willsmore, S. Olson, M. Matic, and H. Kuchel, “Mapping of a novel QTL for resistance to cereal cyst nematode in wheat,” *Theoretical and Applied Genetics*, vol. 112, no. 8, pp. 1480–1486, 2006.

[49] R. Draeger, N. Gosman, A. Steed, et al., “Identification of QTLs for resistance to Fusarium head blight, DON accumulation and associated traits in the winter wheat variety Arina,” *Theoretical and Applied Genetics*, vol. 115, no. 5, pp. 617–625, 2007.

[50] C. A. McCartney, D. J. Somers, B. D. McCallum, et al., “Microsatellite tagging of the leaf rust resistance gene *Lr16* on wheat chromosome 2BSc,” *Molecular Breeding*, vol. 15, no. 4, pp. 329–337, 2005.

[51] S. Li, J. Jia, X. Wei, et al., “A intervarietal genetic map and QTL analysis for yield traits in wheat,” *Molecular Breeding*, vol. 20, no. 2, pp. 167–178, 2007.

[52] J. R. Simmonds, L. J. Fish, M. A. Leverington-Waite, Y. Wang, P. Howell, and J. W. Snape, “Mapping of a gene (*Vir*) for a non-glaucous, viridescent phenotype in bread wheat derived from *Triticum dicoccoides*, and its association with yield variation,” *Euphytica*, vol. 159, no. 3, pp. 333–341, 2008.

[53] M. Akbari, P. Wenzl, V. Caig, et al., “Diversity arrays technology (DArT) for high-throughput profiling of the hexaploid wheat genome,” *Theoretical and Applied Genetics*, vol. 113, no. 8, pp. 1409–1420, 2006.

[54] K. Semagn, Å. Bjørnstad, H. Skinnes, A. G. Marøy, Y. Tarkegne, and M. William, “Distribution of DArT, AFLP, and SSR markers in a genetic linkage map of a doubled-haploid hexaploid wheat population,” *Genome*, vol. 49, no. 5, pp. 545–555, 2006.

[55] E. R. Sears, “The aneuploids of common wheat,” *University of Missouri Agriculture Experiment Station, Bulleten*, vol. 572, pp. 1–58, 1954.

[56] L. L. Qi, B. Echalier, S. Chao, et al., “A chromosome bin map of 16,000 expressed sequence tag loci and distribution of genes among the three genomes of polyploid wheat,” *Genetics*, vol. 168, no. 2, pp. 701–712, 2004.

[57] R. S. Kota, K. S. Gill, B. S. Gill, and T. R. Endo, “A cytogenetically based physical map of chromosome 1B in common wheat,” *Genome*, vol. 36, no. 3, pp. 548–554, 1993.

[58] K. S. Gill, B. S. Gill, T. R. Endo, and T. Taylor, “Identification and high-density mapping of gene-rich regions in chromosome group 1 of wheat,” *Genetics*, vol. 144, no. 4, pp. 1883–1891, 1996.

[59] D. E. Delaney, S. Nasuda, T. R. Endo, B. S. Gill, and S. H. Hulbert, “Cytologically based physical maps of the group-2 chromosomes of wheat,” *Theoretical and Applied Genetics*, vol. 91, no. 4, pp. 568–573, 1995.

[60] M. S. Röder, V. Korzun, B. S. Gill, and M. W. Ganal, “The physical mapping of microsatellite markers in wheat,” *Genome*, vol. 41, no. 2, pp. 278–283, 1998.

[61] D. E. Delaney, S. Nasuda, T. R. Endo, B. S. Gill, and S. H. Hulbert, “Cytologically based physical maps of the group 3 chromosomes of wheat,” *Theoretical and Applied Genetics*, vol. 91, no. 5, pp. 780–782, 1995.

[62] L. Mickelson-Young, T. R. Endo, and B. S. Gill, “A cytogenetic ladder-map of the wheat homoeologous group-4 chromosomes,” *Theoretical and Applied Genetics*, vol. 90, no. 7-8, pp. 1007–1011, 1995.

[63] K. S. Gill, B. S. Gill, T. R. Endo, and E. V. Boyko, “Identification and high-density mapping of gene-rich regions in chromosome group 5 of wheat,” *Genetics*, vol. 143, no. 2, pp. 1001–1012, 1996.

[64] J. D. Faris, K. M. Haen, and B. S. Gill, “Saturation mapping of a gene-rich recombination hot spot region in wheat,” *Genetics*, vol. 154, no. 2, pp. 823–835, 2000.

[65] L. L. Qi and B. S. Gill, “High-density physical maps reveal that the dominant male-sterile gene Ms3 is located in a genomic region of low recombination in wheat and is not amenable to map-based cloning,” *Theoretical and Applied Genetics*, vol. 103, no. 6-7, pp. 998–1006, 2001.

[66] Y. Ogihara, K. Hasegawa, and H. Tsujimoto, “High-resolution cytological mapping of the long arm of chromosome 5A in common wheat using a series of deletion lines induced by gametocidal (Gc) genes of *Aegilops speltoides*,” *Molecular and General Genetics*, vol. 244, no. 3, pp. 253–259, 1994.

[67] K. S. Gill, B. S. Gill, and T. R. Endo, “A chromosome region-specific mapping strategy reveals gene-rich telomeric ends in wheat,” *Chromosoma*, vol. 102, no. 6, pp. 374–381, 1993.

[68] Y. Weng, N. A. Tuleen, and G. E. Hart, “Extended physical maps and a consensus physical map of the homoeologous group-6 chromosomes of wheat (*Triticum aestivum* L. em Thell.),” *Theoretical and Applied Genetics*, vol. 100, no. 3-4, pp. 519–527, 2000.

[69] Y. Weng and M. D. Lazar, “Comparison of homoeologous group-6 short arm physical maps of wheat and barley reveals a similar distribution of recombinogenic and gene-rich regions,” *Theoretical and Applied Genetics*, vol. 104, no. 6-7, pp. 1078–1085, 2002.

[70] J. E. Werner, T. R. Endo, and B. S. Gill, “Towards a cytogenetically based physical map of the wheat genome,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 89, pp. 11307–11311, 1992.

[71] U. Hohmann, T. R. Endo, K. S. Gill, and B. S. Gill, “Comparison of genetic and physical maps of group 7 chromosomes from *Triticum aestivum* L,” *Molecular and General Genetics*, vol. 245, no. 5, pp. 644–653, 1994.

[72] R. K. Varshney, M. Prasad, J. K. Roy, M. S. Röder, H. S. Balyan, and P. K. Gupta, “Integrated physical maps of 2DL, 6BS and 7DL carrying loci for grain protein content and pre-harvest sprouting tolerance in bread wheat,” *Cereal Research Communications*, vol. 29, no. 1-2, pp. 33–40, 2001.

[73] H. Zhang, S. Nasuda, and T. R. Endo, “Identification of AFLP markers on the satellite region of chromosome 1BS in wheat,” *Genome*, vol. 43, no. 5, pp. 729–735, 2000.

[74] M. A. Rodriguez Milla and J. P. Gustafson, “Genetic and physical characterization of chromosome 4DL in wheat,” *Genome*, vol. 44, no. 5, pp. 883–892, 2001.

[75] D. Sandhu, D. Sidhu, and K. S. Gill, “Identification of expressed sequence markers for a major gene-rich region

of wheat chromosome group 1 using RNA fingerprinting-differential display," *Crop Science*, vol. 42, no. 4, pp. 1285–1290, 2002.

[76] P. Sourdille, S. Singh, T. Cadalen, et al., "Microsatellite-based deletion bin system for the establishment of genetic-physical map relationships in wheat (*Triticum aestivum* L.)," *Functional and Integrative Genomics*, vol. 4, no. 1, pp. 12–25, 2004.

[77] A. Goyal, R. Bandopadhyay, P. Sourdille, T. R. Endo, H. S. Balyan, and P. K. Gupta, "Physical molecular maps of wheat chromosomes," *Functional & Integrative Genomics*, vol. 5, no. 4, pp. 260–263, 2005.

[78] J. H. Peng and N. L. V. Lapitan, "Characterization of EST-derived microsatellites in the wheat genome and development of eSSR markers," *Functional and Integrative Genomics*, vol. 5, no. 2, pp. 80–96, 2005.

[79] A. Mohan, A. Goyal, R. Singh, H. S. Balyan, and P. K. Gupta, "Physical mapping of wheat and rye expressed sequence tag-simple sequence repeats on wheat chromosomes," *Crop Science*, vol. 47, supplement 1, pp. S3–S13, 2007.

[80] S. K. Parida, K. A. Raj Kumar, V. Dalal, N. K. Singh, and T. Mohapatra, "Unigene derived microsatellite markers for the cereal genomes," *Theoretical and Applied Genetics*, vol. 112, no. 5, pp. 808–817, 2006.

[81] K. Hill-Ambroz, C. A. Webb, A. R. Matthews, W. Li, B. S. Gill, and J. P. Fellers, "Expression analysis and physical mapping of a cDNA library of Fusarium head blight infected wheat spikes," *Crop Science*, vol. 46, supplement 1, pp. S15–S26, 2006.

[82] S. J. Goss and H. Harris, "New method for mapping genes in human chromosomes," *Nature*, vol. 255, no. 5511, pp. 680–684, 1975.

[83] D. R. Cox, M. Burmeister, E. R. Price, S. Kim, and R. M. Myers, "Radiation hybrid mapping: a somatic cell genetic method for constructing high-resolution maps of mammalian chromosomes," *Science*, vol. 250, no. 4978, pp. 245–250, 1990.

[84] V. Kalavacharla, K. Hossain, Y. Gu, et al., "High-resolution radiation hybrid map of wheat chromosome 1D," *Genetics*, vol. 173, no. 2, pp. 1089–1099, 2006.

[85] H.-Q. Ling, Y. Zhu, and B. Keller, "High-resolution mapping of the leaf rust disease resistance gene *Lr*1 in wheat and characterization of BAC clones from the *Lr*1 locus," *Theoretical and Applied Genetics*, vol. 106, no. 5, pp. 875–882, 2003.

[86] S. Cloutier, B. D. McCallum, C. Loutre, et al., "Leaf rust resistance gene *Lr1*, isolated from bread wheat (*Triticum aestivum* L.) is a member of the large *psr567* gene family," *Plant Molecular Biology*, vol. 65, no. 1-2, pp. 93–106, 2007.

[87] C. Feuillet, S. Travella, N. Stein, L. Albar, A. Nublat, and B. Keller, "Map-based isolation of the leaf rust disease resistance gene *Lr10* from the hexaploid wheat (*Triticum aestivum* L.) genome," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 25, pp. 15253–15258, 2003.

[88] L. Huang, S. A. Brooks, W. Li, J. P. Fellers, H. N. Trick, and B. S. Gill, "Map-based cloning of leaf rust resistance gene *Lr*21 from the large and polyploid genome of bread wheat," *Genetics*, vol. 164, no. 2, pp. 655–664, 2003.

[89] L. Yan, A. Loukoianov, G. Tranquilli, M. Helguera, T. Fahima, and J. Dubcovsky, "Positional cloning of the wheat vernalization gene *VRN1*," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 10, pp. 6263–6268, 2003.

[90] L. Yan, A. Loukoianov, A. Blechl, et al., "The wheat *VRN2* gene is a flowering repressor down-regulated by vernalization," *Science*, vol. 303, no. 5664, pp. 1640–1644, 2004.

[91] L. Yan, D. Fu, C. Li, et al., "The wheat and barley vernalization gene *VRN3* is an orthologue of FT," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 51, pp. 19581–19586, 2006.

[92] K. J. Simons, J. P. Fellers, H. N. Trick, et al., "Molecular characterization of the major wheat domestication gene Q," *Genetics*, vol. 172, no. 1, pp. 547–555, 2006.

[93] J. D. Faris, J. P. Fellers, S. A. Brooks, and B. S. Gill, "A bacterial artificial chromosome contig spanning the major domestication locus *Q* in wheat and identification of a candidate gene," *Genetics*, vol. 164, no. 1, pp. 311–321, 2003.

[94] N. Yahiaoui, P. Srichumpa, R. Dudler, and B. Keller, "Genome analysis at different ploidy levels allows cloning of the powdery mildew resistance gene *Pm3b* from hexaploid wheat," *Plant Journal*, vol. 37, no. 4, pp. 528–538, 2004.

[95] S. Brunner, P. Srichumpa, N. Yahiaoui, and B. Keller, "Positional cloning and evolution of powdery mildew resistance gene at *Pm3* locus of hexaploid wheat," in *Proceedings of the Plant & Animal Genome XIII Conference*, p. 73, Town & Country Convention Center, San Diego, Calif, USA, January 2005.

[96] C. Uauy, A. Distelfeld, T. Fahima, A. Blechl, and J. Dubcovsky, "A NAC gene regulating senescence improves grain protein, zinc, and iron content in wheat," *Science*, vol. 314, no. 5803, pp. 1298–1301, 2006.

[97] A. Distelfeld, C. Uauy, S. Olmos, A. R. Schlatter, J. Dubcovsky, and T. Fahima, "Microcolinearity between a 2-cM region encompassing the grain protein content locus *Gpc-6B1* on wheat chromosome 6B and a 350-kb region on rice chromosome 2," *Functional & Integrative Genomics*, vol. 4, no. 1, pp. 59–66, 2004.

[98] S. Liu, M. O. Pumphery, X. Zhang, et al., "Towards positional cloning of Qfhs.ndsu-3BS, a major QTL for Fusarium head blight resistance in wheat," in *Proceedings of the Plant & Animal Genome XIII Conference*, p. 71, Town & Country Convention Center, San Diego, Calif, USA, January 2005.

[99] P. Ling, X. Chen, D. Q. Le, and K. G. Campbell, "Towards cloning of the *Yr*5 gene for resistance to wheat stripe rust resistance," in *Proceedings of the Plant & Animal Genomes XIII Conference*, Town & Country Convention Center, San Diego, Calif, USA, January 2005.

[100] T. Schnurbusch, N. C. Collins, R. F. Eastwood, T. Sutton, S. P. Jefferies, and P. Langridge, "Fine mapping and targeted SNP survey using rice-wheat gene colinearity in the region of the *Bo1* boron toxicity tolerance locus of bread wheat," *Theoretical and Applied Genetics*, vol. 115, no. 4, pp. 451–461, 2007.

[101] H.-J. Lu, J. P. Fellers, T. L. Friesen, S. W. Meinhardt, and J. D. Faris, "Genomic analysis and marker development for the *Tsn1* locus in wheat using bin-mapped ESTs and flanking BAC contigs," *Theoretical and Applied Genetics*, vol. 112, no. 6, pp. 1132–1142, 2006.

[102] S. Griffiths, R. Sharp, T. N. Foote, et al., "Molecular characterization of *Ph*1 as a major chromosome pairing locus in polyploid wheat," *Nature*, vol. 439, no. 7077, pp. 749–752, 2006.

[103] R. S. Kota, W. Spielmeyer, R. A. McIntosh, and E. S. Lagudah, "Fine genetic mapping fails to dissociate durable stem rust resistance gene *Sr*2 from pseudo-black chaff in

common wheat (*Triticum aestivum* L.)," *Theoretical and Applied Genetics*, vol. 112, no. 3, pp. 492–499, 2006.

[104] R. B. Flavell and D. B. Smith, "The role of homoeologous group 1 chromosomes in the control of rRNA genes in wheat," *Biochemical Genetics*, vol. 12, no. 4, pp. 271–279, 1974.

[105] W. L. Gerlach and W. J. Peacock, "Chromosomal locations of highly repeated DNA sequences in wheat," *Heredity*, vol. 44, no. 2, pp. 269–276, 1980.

[106] W. L. Gerlach, E. S. Dennis, and W. J. Peacock, "Molecular cytogenetics of wheat," in *Cytogenetics of Crop Plant*, M. S. Swaminathan, P. K. Gupta, and U. Sinha, Eds., pp. 191–212, MacMillan, Bomby, India, 1983.

[107] Y. Mukai, Y. Nakahara, and M. Yamamoto, "Simultaneous discrimination of the three genomes in hexaploid wheat by multicolor fluorescence *in situ* hybridization using total genomic and highly repeated DNA probes," *Genome*, vol. 36, no. 3, pp. 489–494, 1993.

[108] C. Pedersen and P. Langridge, "Identification of the entire chromosome complement of bread wheat by two-colour FISH," *Genome*, vol. 40, no. 5, pp. 589–593, 1997.

[109] E. D. Badaeva, A. V. Amosova, O. V. Muravenko, et al., "Genome differentiation in *Aegilops*. 3. Evolution of the D-genome cluster," *Plant Systematics and Evolution*, vol. 231, no. 1–4, pp. 163–190, 2002.

[110] P. Zhang, W. Li, B. Friebe, and B. S. Gill, "Simultaneous painting of three genomes in hexaploid wheat by BAC-FISH," *Genome*, vol. 47, no. 5, pp. 979–987, 2004.

[111] Y. Mukai, T. R. Endo, and B. S. Gill, "Physical mapping of the 5S rRNA multigene family in common wheat," *Journal of Heredity*, vol. 81, no. 4, pp. 290–295, 1990.

[112] Y. Mukai, T. R. Endo, and B. S. Gill, "Pysical mapping of the 18S.26S rRNA multigene family in common wheat: identification of a new locus," *Chromosoma*, vol. 100, no. 2, pp. 71–78, 1991.

[113] X.-F. Ma, K. Ross, and J. P. Gustafson, "Physical mapping of restriction fragment length polymorphism (RFLP) markers in homoeologous groups 1 and 3 chromosomes of wheat by *in situ* hybridization," *Genome*, vol. 44, no. 3, pp. 401–412, 2001.

[114] S. Rahman, A. Regina, Z. Li, et al., "Comparison of starch-branching enzyme genes reveals evolutionary relationships among isoforms. Characterization of a gene for starch-branching enzyme IIa from the wheat D genome donor *Aegilops tauschii*," *Plant Physiology*, vol. 125, no. 3, pp. 1314–1324, 2001.

[115] Z. Li, F. Sun, S. Xu, et al., "The structural organisation of the genes encoding class II starch synthase of wheat and barley and the evolution of the genes encoding starch syuthases in plants," *Functional & Integrative Genomics*, vol. 3, no. 1-2, pp. 76–85, 2003.

[116] K.-M. Turnbull, M. Turner, Y. Mukai, et al., "The organization of genes tightly linked to the Ha locus in *Aegilops tauschii*, the D-genome donor to wheat," *Genome*, vol. 46, no. 2, pp. 330–338, 2003.

[117] Y. Mukai and B. S. Gill, "Detection of barley chromatin added to wheat by genomic *in situ* hybridization," *Genome*, vol. 34, no. 3, pp. 448–452, 1991.

[118] T. Schwarzacher, K. Anamthawat-Jónsson, G. E. Harrison, et al., "Genomic *in situ* hybridization to identify alien chromosomes and chromosome segments in wheat," *Theoretical and Applied Genetics*, vol. 84, no. 7-8, pp. 778–786, 1992.

[119] M. Biagetti, F. Vitellozzi, and C. Ceoloni, "Physical mapping of wheat-*Aegilops longissima* breakpoints in mildew-resistant recombinant lines using FISH with highly repeated and low-copy DNA probes," *Genome*, vol. 42, no. 5, pp. 1013–1019, 1999.

[120] M. Yamamoto and Y. Mukai, "High-resolution mapping in wheat and rye by FISH on extended DNA fibres," in *Proceedings of the 9th International Wheat Genetics Symposium*, A. E. Slinkard, Ed., vol. 1, pp. 12–16, Saskatoon, Canada, August 1998.

[121] M. Yamamoto and Y. Mukai, "High-resolution physical mapping of the secalin-1 locus of rye on extended DNA fibers," *Cytogenetic and Genome Research*, vol. 109, no. 1–3, pp. 79–82, 2005.

[122] U. C. Lavania, M. Yamamoto, and Y. Mukai, "Extended chromatin and DNA fibers from active plant nuclei for high-resolution FISH," *Journal of Histochemistry & Cytochemistry*, vol. 51, no. 10, pp. 1249–1253, 2003.

[123] K.-N. Fukui, G. Suzuki, E. S. Lagudah, et al., "Physical arrangement of retrotransposon-related repeats in centromeric regions of wheat," *Plant & Cell Physiology*, vol. 42, no. 2, pp. 189–196, 2001.

[124] M. Valárik, J. Bartoš, P. Kovářová, M. Kubaláková, J. H. de Jong, and J. Doležel, "High-resolution FISH on super-stretched flow-sorted plant chromosomes," *Plant Journal*, vol. 37, no. 6, pp. 940–950, 2004.

[125] S. A. Jackson, P. Zhang, W. P. Chen, et al., "High-resolution structural analysis of biolistic transgene integration into the genome of wheat," *Theoretical and Applied Genetics*, vol. 103, no. 1, pp. 56–62, 2001.

[126] P. Zhang, B. Friebe, and B. Gill, "Potential and limitations of BAC-FISH mapping in wheat," in *Proceedings of the Plant, Animal & Microbe Genomes X Conference*, p. 272, Town & Country Convention Center, San Diego, Calif, USA, January 2002.

[127] D. Papa, C. A. Miller, G. R. Anderson, et al., "FISH physical mapping of DNA sequences associated with RWA resistance in wheat and barley," in *Proceedings of the Plant & Animal Genome VIII Conference*, p. 36, Town & Country Hotel, San Diego, Calif, USA, January 2000.

[128] P. Zhang, W. Li, J. Fellers, B. Friebe, and B. S. Gill, "BAC-FISH in wheat identifies chromosome landmarks consisting of different types of transposable elements," *Chromosoma*, vol. 112, no. 6, pp. 288–299, 2004.

[129] X. Q. Huang, H. Cöster, M. W. Ganal, and M. S. Röder, "Advanced backcross QTL analysis for the identification of quantitative trait loci alleles from wild relatives of wheat (*Triticum aestivum* L.)," *Theoretical and Applied Genetics*, vol. 106, no. 8, pp. 1379–1389, 2003.

[130] X. Q. Huang, L. X. Wang, M. X. Xu, and M. S. Röder, "Microsatellite mapping of the powdery mildew resistance gene *Pm5e* in common wheat (*Triticum aestivum* L.)," *Theoretical and Applied Genetics*, vol. 106, no. 5, pp. 858–865, 2003.

[131] G. P. Yan, X. M. Chen, R. F. Line, and C. R. Wellings, "Resistance gene-analog polymorphism markers co-segregating with the *Yr5* gene for resistance to wheat stripe rust," *Theoretical and Applied Genetics*, vol. 106, no. 4, pp. 636–643, 2003.

[132] P. K. Gupta, S. Rustgi, S. Sharma, R. Singh, N. Kumar, and H. S. Balyan, "Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat,"

Molecular Genetics and Genomics, vol. 270, no. 4, pp. 315–323, 2003.

[133] L. F. Gao, J. Tang, H. Li, and J. Jia, "Analysis of microsatellites in major crops assessed by computational and experimental approaches," *Molecular Breeding*, vol. 12, no. 3, pp. 245–261, 2003.

[134] R. Bandopadhyay, S. Sharma, S. Rustgi, et al., "DNA polymorphism among 18 species of *Triticum-Aegilops* complex using wheat EST-SSRs," *Plant Science*, vol. 166, no. 2, pp. 349–356, 2004.

[135] R. K. Varshney, R. Sigmund, A. Börner, et al., "Interspecific transferability and comparative mapping of barley EST-SSR markers in wheat, rye and rice," *Plant Science*, vol. 168, no. 1, pp. 195–202, 2005.

[136] J.-K. Yu, M. La Rota, R. V. Kantety, and M. E. Sorrells, "EST derived SSR markers for comparative mapping in wheat and rice," *Molecular Genetics and Genomics*, vol. 271, no. 6, pp. 742–751, 2004.

[137] L. Y. Zhang, M. Bernard, P. Leroy, C. Feuillet, and P. Sourdille, "High transferability of bread wheat EST-derived SSRs to other cereals," *Theoretical and Applied Genetics*, vol. 111, no. 4, pp. 677–687, 2005.

[138] J. Tang, L. Gao, Y. Cao, and J. Jia, "Homologous analysis of SSR-ESTs and transferability of wheat SSR-EST markers across barley, rice and maize," *Euphytica*, vol. 151, no. 1, pp. 87–93, 2006.

[139] K. Chabane, O. Abdalla, H. Sayed, and J. Valkoun, "Assessment of EST-microsatellites markers for discrimination and genetic diversity in bread and durum wheat landraces from Afghanistan," *Genetic Resources and Crop Evolution*, vol. 54, no. 5, pp. 1073–1080, 2007.

[140] W. Zhang, S. Chao, E. D. Akhunov, et al., "Discovery of SNPs for wheat homoeologous group 5 and polymorphism among US adapted wheat germplasm," in *Proceedings of the Plant & Animal Genome XI Conference*, p. 184, San Diego, Calif, USA, January 2007.

[141] C. Ravel, S. Praud, A. Murigneux, et al., "Single-nucleotide polymorphism frequency in a set of selected lines of bread wheat (*Triticum aestivum* L.)," *Genome*, vol. 49, no. 9, pp. 1131–1139, 2006.

[142] J. Janda, J. Bartoš, J. Šafář, et al., "Construction of a subgenomic BAC library specific for chromosomes 1D, 4D and 6D of hexaploid wheat," *Theoretical and Applied Genetics*, vol. 109, no. 7, pp. 1337–1345, 2004.

[143] J. Šafář, J. Bartoš, J. Janda, et al., "Dissecting large and complex genomes: flow sorting and BAC cloning of individual chromosomes from bread wheat," *The Plant Journal*, vol. 39, no. 6, pp. 960–968, 2004.

[144] J. Janda, J. Šafář, M. Kubaláková, et al., "Advanced resources for plant genomics: a BAC library specific for the short arm of wheat chromosome 1B," *The Plant Journal*, vol. 47, no. 6, pp. 977–986, 2006.

[145] T. Wicker, N. Stein, L. Albar, C. Feuillet, E. Schlagenhauf, and B. Keller, "Analysis of a contiguous 211 kb sequence in diploid wheat (*Triticum monococcum* L.) reveals multiple mechanisms of genome evolution," *The Plant Journal*, vol. 26, no. 3, pp. 307–316, 2001.

[146] S. A. Brooks, L. Huang, B. S. Gill, and J. P. Fellers, "Analysis of 106 kb of contiguous DNA sequence from the D genome of wheat reveals high gene density and a complex arrangement of genes related to disease resistance," *Genome*, vol. 45, no. 5, pp. 963–972, 2002.

[147] E. Paux, D. Roger, E. Badaeva, et al., "Characterizing the composition and evolution of homoeologous genomes in hexaploid wheat through BAC-end sequencing on chromosome 3B," *The Plant Journal*, vol. 48, no. 3, pp. 463–474, 2006.

[148] D. Sandhu, J. A. Champoux, S. N. Bondareva, and K. S. Gill, "Identification and physical localization of useful genes and markers to a major gene-rich region on wheat group 1S chromosomes," *Genetics*, vol. 157, no. 4, pp. 1735–1747, 2001.

[149] M. Erayman, D. Sandhu, D. Sidhu, M. Dilbirligi, P. S. Baenziger, and K. S. Gill, "Demarcating the gene-rich regions of the wheat genome," *Nucleic Acids Research*, vol. 32, no. 12, pp. 3546–3565, 2004.

[150] K. S. Gill, "Structural organization of the wheat genome," in *Frontiers of Wheat Bioscience: The 100th Memorial Issue of Wheat Information Service*, K. Tsunewaki, Ed., pp. 151–167, Kihara Memorial Yokohama Foundation for the Advancement of Life Sciences, Yokohama, Japan, 2005.

[151] D. Sidhu and K. S. Gill, "Distribution of genes and recombination in wheat and other eukaryotes," *Plant Cell, Tissue and Organ Culture*, vol. 79, no. 3, pp. 257–270, 2005.

[152] A. Barakat, N. Carels, and G. Bernardi, "The distribution of genes in the genomes of Gramineae," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 94, no. 13, pp. 6857–6861, 1997.

[153] C. Feuillet and B. Keller, "High gene density is conserved at syntenic loci of small and large grass genomes," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96, no. 14, pp. 8265–8270, 1999.

[154] D. Sandhu and K. S. Gill, "Gene-containing regions of wheat and the other grass genomes," *Plant Physiology*, vol. 128, no. 3, pp. 803–811, 2002.

[155] D. Sandhu and K. S. Gill, "Structural and functional organization of the '1S0.8 gene-rich region' in the *Triticeae*," *Plant Molecular Biology*, vol. 48, no. 5-6, pp. 791–804, 2002.

[156] P. K. Gupta, P. L. Kulwal, and S. Rustgi, "Wheat cytogenetics in the genomics era and its relevance to breeding," *Cytogenetic and Genome Research*, vol. 109, no. 1–3, pp. 315–327, 2005.

[157] M. L. Wang, A. R. Leitch, T. Schwarzacher, J. S. Heslop-Harrison, and G. Moore, "Construction of a chromosome-enriched *HpaII* library from flow-sorted wheat chromosomes," *Nucleic Acids Research*, vol. 20, no. 8, pp. 1897–1901, 1992.

[158] J.-H. Lee, K. Arumuganathan, Y. Yen, S. Kaeppler, H. Kaeppler, and P. S. Baenziger, "Root tip cell cycle synchronization and metaphase-chromosome isolation suitable for flow sorting in common wheat (*Triticum aestivum* L.)," *Genome*, vol. 40, no. 5, pp. 633–638, 1997.

[159] J. Vrána, M. Kubalákova, H. Simková, J. Cíhalíková, M. A. Lysák, and J. Dolezel, "Flow sorting of mitotic chromosomes in common wheat (*Triticum aestivum* L.)," *Genetics*, vol. 156, no. 4, pp. 2033–2041, 2000.

[160] K. S. Gill, K. Arumuganathan, and J.-H. Lee, "Isolating individual wheat (*Triticum aestivum*) chromosome arms by flow cytometric analysis of ditelosomic lines," *Theoretical and Applied Genetics*, vol. 98, no. 8, pp. 1248–1252, 1999.

[161] J. M. Vega, S. Abbo, M. Feldman, and A. A. Levy, "Chromosome painting in plants: in situ hybridization with a DNA probe from a specific microdissected chromosome arm of common wheat," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 91, no. 25, pp. 12041–12045, 1994.

[162] B. Chalhoub, H. Belcram, and M. Caboche, "Efficient cloning of plant genomes into bacterial artificial chromosome (BAC) libraries with larger and more uniform insert size," *Plant Biotechnology Journal*, vol. 2, no. 3, pp. 181–188, 2004.

[163] J. Doležel, M. Kubaláková, J. Bartoš, and J. Macas, "Flow cytogenetics and plant genome mapping," *Chromosome Research*, vol. 12, no. 1, pp. 77–91, 2004.

[164] M. Kubaláková, J. Vrána, J. Číhalíková, H. Šimková, and J. Doležel, "Flow karyotyping and chromosome sorting in bread wheat (*Triticum aestivum* L.)," *Theoretical and Applied Genetics*, vol. 104, no. 8, pp. 1362–1372, 2002.

[165] J. Doležel, M. Kubaláková, P. Suchankova, et al., "Flow cytogenetic analysis of the wheat genome," in *Frontiers of Wheat Bioscience: The 100th Memorial Issue of Wheat Information Service*, K. Tsunewaki, Ed., pp. 3–15, Yokohama Publishers, Yokohama, Japan, 2005.

[166] B. S. Gill, "International genome research on wheat (IGROW)," in *Proceedings of the National Wheat Workers Workshop*, Kansas City, Mo, USA, February 2004.

[167] P. K. Gupta, "Ultrafast and low-cost DNA sequencing methods for applied genomics research," *Proceedings of the National Academy of Sciences, India*. In press.

[168] I. D. Wilson, G. L. A. Barker, R. W. Beswick, et al., "A transcriptomics resource for wheat functional genomics," *Plant Biotechnology Journal*, vol. 2, no. 6, pp. 495–506, 2004.

[169] I. D. Wilson, G. L. Barker, C. Lu, et al., "Alteration of the embryo transcriptome of hexaploid winter wheat (*Triticum aestivum* cv. Mercia) during maturation and germination," *Functional and Integrative Genomics*, vol. 5, no. 3, pp. 144–154, 2005.

[170] R. Poole, G. Barker, I. D. Wilson, J. A. Coghill, and K. J. Edwards, "Measuring global gene expression in polyploidy; a cautionary note from allohexaploid wheat," *Functional & Integrative Genomics*, vol. 7, no. 3, pp. 207–219, 2007.

[171] C. A. McCartney, D. J. Somers, D. G. Humphreys, et al., "Mapping quantitative trait loci controlling agronomic traits in the spring wheat cross RL4452 × 'AC Domain'," *Genome*, vol. 48, no. 5, pp. 870–883, 2005.

[172] C. A. McCartney, D. J. Somers, O. Lukow, et al., "QTL analysis of quality traits in the spring wheat cross RL4452 × 'AC domain'," *Plant Breeding*, vol. 125, no. 6, pp. 565–575, 2006.

[173] D. Fu, C. Uauy, A. Blechl, and J. Dubcovsky, "RNA interference for wheat functional gene analysis," *Transgenic Research*, vol. 16, no. 6, pp. 689–701, 2007.

[174] A. Salleh, "Gene silencing yields high-fibre wheat," February 2006, ABC Science online.

[175] K. Mochida, Y. Yamazaki, and Y. Ogihara, "Discrimination of homoeologous gene expression in hexaploid wheat by SNP analysis of contigs grouped from a large number of expressed sequence tags," *Molecular Genetics and Genomics*, vol. 270, no. 5, pp. 371–377, 2003.

[176] P. Schweizer, J. Pokorny, P. Schulze-Lefert, and R. Dudler, "Double-stranded RNA interferes with gene function at the single-cell level in cereals," *The Plant Journal*, vol. 24, no. 6, pp. 895–903, 2000.

[177] A. B. Christensen, H. Thordal-Christensen, G. Zimmermann, et al., "The Germinlike protein GLP4 exhibits superoxide dismutase activity and is an important component of quantitative resistance in wheat and barley," *Molecular Plant-Microbe Interactions*, vol. 17, no. 1, pp. 109–117, 2004.

[178] A. Loukoianov, L. Yan, A. Blechl, A. Sanchez, and J. Dubcovsky, "Regulation of *VRN-1* vernalization genes in normal and transgenic polyploid wheat," *Plant Physiology*, vol. 138, no. 4, pp. 2364–2373, 2005.

[179] A. Regina, A. Bird, D. Topping, et al., "High-amylose wheat generated by RNA interference improves indices of large-bowel health in rats," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 10, pp. 3546–3551, 2006.

[180] S. Travella, T. E. Klimm, and B. Keller, "RNA interference-based gene silencing as an efficient tool for functional genomics in hexaploid bread wheat," *Plant Physiology*, vol. 142, no. 1, pp. 6–20, 2006.

[181] G. Yao, J. Zhang, L. Yang, et al., "Genetic mapping of two powdery mildew resistance genes in einkorn (*Triticum monococcum* L.) accessions," *Theoretical and Applied Genetics*, vol. 114, no. 2, pp. 351–358, 2007.

[182] S. Henikoff, B. J. Till, and L. Comai, "TILLING. Traditional mutagenesis meets functional genomics," *Plant Physiology*, vol. 135, no. 2, pp. 630–636, 2004.

[183] B. J. Till, S. H. Reynolds, E. A. Greene, et al., "Large-scale discovery of induced point mutations with high-throughput TILLING," *Genome Research*, vol. 13, no. 3, pp. 524–530, 2003.

[184] A. J. Slade and V. C. Knauf, "TILLING moves beyond functional genomics into crop improvement," *Transgenic Research*, vol. 14, no. 2, pp. 109–115, 2005.

[185] A. J. Slade, S. I. Fuerstenberg, D. Loeffler, M. N. Steine, and D. Facciotti, "A reverse genetic, nontransgenic approach to wheat crop improvement by TILLING," *Nature Biotechnology*, vol. 23, no. 1, pp. 75–81, 2005.

[186] K. M. Devos and M. D. Gale, "Genome relationships: the grass model in current research," *The Plant Cell*, vol. 12, no. 5, pp. 637–646, 2000.

[187] T. Wicker, N. Yahiaoui, and B. Keller, "Contrasting rates of evolution in *Pm3* loci from three wheat species and rice," *Genetics*, vol. 177, no. 2, pp. 1207–1216, 2007.

[188] N. Huo, Y. Q. Gu, G. R. Lazo, et al., "Construction and characterization of two BAC libraries from Brachypodium distachyon, a new model for grass genomics," *Genome*, vol. 49, no. 9, pp. 1099–1108, 2006.

[189] M. D. Gale, J. E. Flintham, and K. M. Devos, "Cereal comparative genetics and preharvest sprouting," *Euphytica*, vol. 126, no. 1, pp. 21–25, 2002.

[190] C. Feuillet, A. Penger, K. Gellner, A. Mast, and B. Keller, "Molecular evolution of receptor-like kinase genes in hexaploid wheat. Independent evolution of orthologs after polyploidization and mechanisms of local rearrangements at paralogous loci," *Plant Physiology*, vol. 125, no. 3, pp. 1304–1313, 2001.

[191] N. Chantret, A. Cenci, F. Sabot, O. Anderson, and J. Dubcovsky, "Sequencing of the *Triticum monococcum Hardness* locus reveals good microcolinearity with rice," *Molecular genetics and genomics*, vol. 271, no. 4, pp. 377–386, 2004.

[192] E. K. Khlestkina, T. A. Pshenichnikova, M. S. Röder, E. A. Salina, V. S. Arbuzova, and A. Börner, "Comparative mapping of genes for glume colouration and pubescence in hexaploid wheat (*Triticum aestivum* L.)," *Theoretical and Applied Genetics*, vol. 113, no. 5, pp. 801–807, 2006.

[193] K. M. Devos, M. D. Atkinson, C. N. Chinoy, C. J. Liu, and M. D. Gale, "RFLP-based genetic map of the homoeologous group 3 chromosomes of wheat and rye," *Theoretical and Applied Genetics*, vol. 83, no. 8, pp. 931–939, 1992.

[194] N. Stein, C. Feuillet, T. Wicker, E. Schlagenhauf, and B. Keller, "Subgenome chromosome walking in wheat: a 450-kb physical contig in *Triticum monococcum* L. spans the *Lr10* resistance locus in hexaploid wheat (*Triticum aestivum* L.)," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 24, pp. 13436–13441, 2000.

[195] T. Wicker, N. Yahiaoui, R. Guyot, et al., "Rapid genome divergence at orthologous low molecular weight glutenin loci of the A and A^m genomes of wheat," *The Plant Cell*, vol. 15, no. 5, pp. 1186–1197, 2003.

[196] E. Isidore, B. Scherrer, B. Chalhoub, C. Feuillet, and B. Keller, "Ancient haplotypes resulting from extensive molecular rearrangements in the wheat A genome have been maintained in species of three different ploidy levels," *Genome Research*, vol. 15, no. 4, pp. 526–536, 2005.

[197] Y. Q. Gu, J. Salse, D. Coleman-Derr, et al., "Types and rates of sequence evolution at the high-molecular-weight glutenin locus in hexaploid wheat and its ancestral genomes," *Genetics*, vol. 174, no. 3, pp. 1493–1504, 2006.

[198] R. Guyot, N. Yahiaoui, C. Feuillet, and B. Keller, "In silico comparative analysis reveals a mosaic conservation of genes within a novel colinear region in wheat chromosome 1AS and rice chromosome 5S," *Functional & Integrative Genomics*, vol. 4, no. 1, pp. 47–58, 2004.

[199] M. E. Sorrells, M. La Rota, C. E. Bermudez-Kandianis, et al., "Comparative DNA sequence analysis of wheat and rice genomes," *Genome Research*, vol. 13, no. 8, pp. 1818–1827, 2003.

[200] N. K. Singh, S. Raghuvanshi, S. K. Srivastava, et al., "Sequence analysis of the long arm of rice chromosome 11 for rice-wheat synteny," *Functional and Integrative Genomics*, vol. 4, no. 2, pp. 102–117, 2004.

[201] J. Draper, L. A. J. Mur, G. Jenkins, et al., "*Brachypodium distachyon*. A new model system for functional genomics in grasses," *Plant Physiology*, vol. 127, no. 4, pp. 1539–1555, 2001.

[202] R. Hasterok, A. Marasek, I. S. Donnison, et al., "Alignment of the genomes of *Brachypodium distachyon* and temperate cereals and grasses using bacterial artificial chromosome landing with fluorescence in *situ* hybridization," *Genetics*, vol. 173, no. 1, pp. 349–362, 2006.

[203] J. P. Vogel, Y. Q. Gu, P. Twigg, et al., "EST sequencing and phylogenetic analysis of the model grass *Brachypodium distachyon*," *Theoretical and Applied Genetics*, vol. 113, no. 2, pp. 186–195, 2006.

[204] E. Bossolini, T. Wicker, P. A. Knobel, and B. Keller, "Comparison of orthologous loci from small grass genomes *Brachypodium* and rice: implications for wheat genomics and grass genome annotation," *The Plant Journal*, vol. 49, no. 4, pp. 704–717, 2007.

[205] Y. Xie, Z. Ni, Y. Yao, Y. Yin, Q. Zhang, and Q. Sun, "Analysis of differential cytosine methylation during seed development in wheat," in *Proceedings of the Plant Genomics in China VIII*, p. 60, Shanghai, China, August 2007.

[206] N. Shitsukawa, C. Tahira, K.-I. Kassai, et al., "Genetic and epigenetic alteration among three homoeologous genes of a class E MADS box gene in hexaploid wheat," *Plant Cell*, vol. 19, no. 6, pp. 1723–1737, 2007.

[207] Y. Nemoto, M. Kisaka, T. Fuse, M. Yano, and Y. Ogihara, "Characterization and functional analysis of three wheat genes with homology to the *CONSTANS* flowering time gene in transgenic rice," *The Plant Journal*, vol. 36, no. 1, pp. 82–93, 2003.

[208] N. Chantret, J. Salse, F. Sabot, et al., "Molecular basis of evolutionary events that shaped the hardness locus in diploid and polyploid wheat species (*Triticum and Aegilops*)," *The Plant Cell*, vol. 17, no. 4, pp. 1033–1045, 2005.

[209] T. Nomura, A. Ishihara, R. C. Yanagita, T. R. Endo, and H. Iwamura, "Three genomes differentially contribute to the biosynthesis of benzoxazinones in hexaploid wheat," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 45, pp. 16490–16495, 2005.

[210] L. Comai, "Genetic and epigenetic interactions in allopolyploid plants," *Plant Molecular Biology*, vol. 43, no. 2-3, pp. 387–399, 2000.

[211] Z. J. Chen and Z. Ni, "Mechanisms of genomic rearrangements and gene expression changes in plant polyploids," *BioEssays*, vol. 28, no. 3, pp. 240–252, 2006.

[212] M. Feldman, B. Liu, G. Segal, S. Abbo, A. A. Levy, and J. M. Vega, "Rapid elimination of low-copy DNA sequences in polyploid wheat: a possible mechanism for differentiation of homoeologous chromosomes," *Genetics*, vol. 147, no. 3, pp. 1381–1387, 1997.

[213] B. Liu, J. M. Vega, G. Segal, S. Abbo, M. Rodova, and M. Feldman, "Rapid genomic changes in newly synthesized amphiploids of *Triticum* and *Aegilops*—I: changes in low-copy noncoding DNA sequences," *Genome*, vol. 41, no. 2, pp. 272–277, 1998.

[214] B. Liu, J. M. Vega, and M. Feldman, "Rapid genomic changes in newly synthesized amphiploids of *Triticum* and *Aegilops*—II: changes in low-copy coding DNA sequences," *Genome*, vol. 41, no. 4, pp. 535–542, 1998.

[215] L. Z. Xiong, C. G. Xu, M. A. S. Maroof, and Q. Zhang, "Patterns of cytosine methylation in an elite rice hybrid and its parental lines, detected by a methylation-sensitive amplification polymorphism technique," *Molecular and General Genetics*, vol. 261, no. 3, pp. 439–446, 1999.

[216] H. Shaked, K. Kashkush, H. Özkan, M. Feldman, and A. A. Levy, "Sequence elimination and cytosine methylation are rapid and reproducible responses of the genome to wide hybridization and allopolyploidy in wheat," *Plant Cell*, vol. 13, no. 8, pp. 1749–1759, 2001.

[217] K. Kashkush, M. Feldman, and A. A. Levy, "Gene loss, silencing and activation in a newly synthesized wheat allotetraploid," *Genetics*, vol. 160, no. 4, pp. 1651–1659, 2002.

[218] A. A. Levy and M. Feldman, "Genetic and epigenetic reprogramming of the wheat genome upon allopolyploidization," *Biological Journal of the Linnean Society*, vol. 82, no. 4, pp. 607–613, 2004.

[219] P. He, B. R. Friebe, B. S. Gill, and J.-M. Zhou, "Allopolyploidy alters gene expression in the highly stable hexaploid wheat," *Plant Molecular Biology*, vol. 52, no. 2, pp. 401–414, 2003.

[220] P. L. Kulwal, R. Singh, H. S. Balyan, and P. K. Gupta, "Genetic basis of pre-harvest sprouting tolerance using single-locus and two-locus QTL analyses in bread wheat," *Functional & Integrative Genomics*, vol. 4, no. 2, pp. 94–101, 2004.

[221] P. L. Kulwal, N. Kumar, A. Kumar, R. K. Gupta, H. S. Balyan, and P. K. Gupta, "Gene networks in hexaploid wheat: interacting quantitative trait loci for grain protein content," *Functional & Integrative Genomics*, vol. 5, no. 4, pp. 254–259, 2005.

[222] N. Kumar, P. L. Kulwal, H. S. Balyan, and P. K. Gupta, "QTL mapping for yield and yield contributing traits in two mapping populations of bread wheat," *Molecular Breeding*, vol. 19, no. 2, pp. 163–177, 2007.

[223] P. Langridge, E. S. Lagudah, T. A. Holton, R. Appels, P. J. Sharp, and K. J. Chalmers, "Trends in genetic and genome analyses in wheat: a review," *Australian Journal of Agricultural Research*, vol. 52, no. 11-12, pp. 1043–1077, 2001.

[224] A. Jahoor, L. Eriksen, and G. Backes, "QTLs and genes for disease resistance in barley and wheat," in *Cereal Genomics*, P. K. Gupta and R. K. Varshney, Eds., pp. 199–251, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2004.

[225] R. Tuberosa and S. Salvi, "QTLs and genes for tolerance to abiotic stresses in cereals," in *Cereal Genomics*, P. K. Gupta and R. K. Varshney, Eds., pp. 253–315, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2004.

[226] P. K. Gupta, S. Rustgi, and N. Kumar, "Genetic and molecular basis of grain size and grain number and its relevance to grain productivity in higher plants," *Genome*, vol. 49, no. 6, pp. 565–571, 2006.

[227] W. Li and B. S. Gill, "Genomics for cereal improvement," in *Cereal Genomics*, P. K. Gupta and R. K. Varshney, Eds., pp. 585–634, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2004.

[228] S. D. Tanksley and J. C. Nelson, "Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines," *Theoretical and Applied Genetics*, vol. 92, no. 2, pp. 191–203, 1996.

[229] X. Q. Huang, H. Kempf, M. W. Canal, and M. S. Röder, "Advanced backcross QTL analysis in progenies derived from a cross between a German elite winter wheat variety and a synthetic wheat (*Triticum aestivum* L.)," *Theoretical and Applied Genetics*, vol. 109, no. 5, pp. 933–943, 2004.

[230] A. Kunert, A. A. Naz, O. Dedeck, K. Pillen, and J. León, "AB-QTL analysis in winter wheat—I: synthetic hexaploid wheat (*T. turgidum* ssp. *dicoccoides* × *T. tauschii*) as a source of favourable alleles for milling and baking quality traits," *Theoretical and Applied Genetics*, vol. 115, no. 5, pp. 683–695, 2007.

[231] N. Amiour, M. Merlino, P. Leroy, and G. Branlard, "Chromosome mapping and identification of amphiphilic proteins of hexaploid wheat kernels," *Theoretical and Applied Genetics*, vol. 108, no. 1, pp. 62–72, 2003.

[232] R. B. Flavell, M. D. Bennett, A. G. Seal, and J. Hutchinson, "Chromosome structure and organisation," in *Wheat Breeding, Its Scientific Basis*, F. G. H. Lupton, Ed., pp. 211–268, Chapman & Hall, London, UK, 1987.

[233] G Kimber, "The B genome of wheat: the present status," in *Cytogenetics of Crop Plants*, M. S. Swaminathan, P. K. Gupta, and U. Sinha, Eds., pp. 213–224, Macmillan, Delhi, India, 1983.

[234] G. Kimber and E. R. Sears, "Evolution in the genus *Triticum* and the origin of cultivated wheat," in *Wheat and Wheat Improvement*, E. G. Heyne, Ed., pp. 154–164, American Society of Agronomy, Madison, Wis, USA, 1987.

[235] M. Feldman, F. G. H. Lupton, and T. E. Miller, "Wheats," in *Evolution of Crops*, J. Smartt and N. W. Simmonds, Eds., pp. 184–192, Longman Scientific, London, UK, 2nd edition, 1995.

[236] B. S. Gill and B. Friebe, "Cytogenetics, phylogeny and evolution of cultivated wheats," in *Bread Wheat, Improvement and Production*, B. C. Curtis, S. Rajaram, and H. G. Macpherson, Eds., Plant Production and Protection Series 30, FAO, Rome, Italy, 2002.

[237] Y. Yen, P. S. Baenziger, and R. Morris, "Genomic constitution of bread wheat: current status," in *Methods of Genome Analysis in Plants*, P. P. Jauhar, Ed., pp. 359–373, CRC Press, Boca Raton, Fla, USA, 1996.

[238] A. A. Levy and M. Feldman, "The impact of polyploidy on grass genome evolution," *Plant Physiology*, vol. 130, no. 4, pp. 1587–1593, 2002.

[239] K. S. Caldwell, J. Dvorak, E. S. Lagudah, et al., "Sequence polymorphism in polyploid wheat and their D-genome diploid ancestor," *Genetics*, vol. 167, no. 2, pp. 941–947, 2004.

[240] S. Huang, A. Sirikhachornkit, X. Su, et al., "Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the *Triticum/Aegilops* complex and the evolutionary history of polyploid wheat," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 12, pp. 8133–8138, 2002.

[241] B. Maestra and T. Naranjo, "Homoeologous relationships of *Aegilops speltoides* chromosomes to bread wheat," *Theoretical and Applied Genetics*, vol. 97, no. 1-2, pp. 181–186, 1998.

[242] E. Nevo, A. B. Korol, A. Beiles, and T. Fahima, *Evolution of Wild Emmer and Wheat Improvement*, Springer, Berlin, Germany, 2002.

[243] N. K. Blake, B. R. Lehfeldt, M. Lavin, and L. E. Talbert, "Phylogenetic reconstruction based on low copy DNA sequence data in an allopolyploid: the B genome of wheat," *Genome*, vol. 42, no. 2, pp. 351–360, 1999.

[244] B. Kilian, H. Özkan, O. Deusch, et al., "Independent wheat B and G genome origins in outcrossing *Aegilops* progenitor haplotypes," *Molecular Biology and Evolution*, vol. 24, no. 1, pp. 217–227, 2007.

[245] L. E. Talbert and N. K. Blake, "Comparative DNA sequence analysis and the origin of wheat," in *Proceedings of the Plant & Animal Genomes VIII Conference*, Town & Country Convention Center, San Diego, Calif, USA, January 2000.

[246] F. Sabot, B. Laubin, L. Amilhat, P. Leroy, P. Sourdille, and M. Bernard, "Evolution history of the *Triticum* sp. through the study of transposable elements," in *Proceedings of the Plant & Animal Genome XII Conference*, p. 421, Town & Country Convention Center, San Diego, Calif, USA, January 2004.

[247] A. H. Schulman, P. K. Gupta, and R. K. Varshney, "Organization of retrotransposons and microsatellites in cereal genomes," in *Cereal Genomics*, P. K. Gupta and R. K. Varshney, Eds., pp. 83–118, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2004.

[248] G. Schachermayr, H. Siedler, M. D. Gale, H. Winzeler, M. Winzeler, and B. Keller, "Identification and localization of molecular markers linked to the *Lr9* leaf rust resistance gene of wheat," *Theoretical and Applied Genetics*, vol. 88, no. 1, pp. 110–115, 1994.

[249] C. Feuillet, M. Messmer, G. Schachermayr, and B. Keller, "Genetic and physical characterization of the *LR1* leaf rust resistance locus in wheat (*Triticum aestivum* L.)," *Molecular and General Genetics*, vol. 248, no. 5, pp. 553–562, 1995.

[250] G. M. Schachermayr, M. M. Messmer, C. Feuillet, H. Winzeler, M. Winzeler, and B. Keller, "Identification of molecular markers linked to the *Agropyron elongatum*-derived leaf rust resistance gene *Lr24* in wheat," *Theoretical and Applied Genetics*, vol. 90, no. 7-8, pp. 982–990, 1995.

[251] G. Schachermayr, C. Feuillet, and B. Keller, "Molecular markers for the detection of the wheat leaf rust resistance gene *Lr10* in diverse genetic backgrounds," *Molecular Breeding*, vol. 3, no. 1, pp. 65–74, 1997.

[252] S. Naik, K. S. Gill, V. S. Prakasa Rao, et al., "Identification of a STS marker linked to the *Aegilops speltoides*-derived leaf

rust resistance gene *Lr28* in wheat," *Theoretical and Applied Genetics*, vol. 97, no. 4, pp. 535–540, 1998.

[253] F. Sacco, E. Y. Suárez, and T. Naranjo, "Mapping of the leaf rust resistance gene *Lr3* on chromosome 6B of Sinvalocho MA wheat," *Genome*, vol. 41, no. 5, pp. 686–690, 1998.

[254] R. Seyfarth, C. Feuillet, G. Schachermayr, M. Winzeler, and B. Keller, "Development of a molecular marker for the adult plant leaf rust resistance gene *Lr35* in wheat," *Theoretical and Applied Genetics*, vol. 99, no. 3-4, pp. 554–560, 1999.

[255] M. Helguera, I. A. Khan, and J. Dubcovsky, "Development of PCR markers for the wheat leaf rust resistance gene *Lr*47," *Theoretical and Applied Genetics*, vol. 100, no. 7, pp. 1137–1143, 2000.

[256] M. Aghaee-Sarbarzeh, H. Singh, and H. S. Dhaliwal, "A microsatellite marker linked to leaf rust resistance transferred from *Aegilops triuncialis* into hexaploid wheat," *Plant Breeding*, vol. 120, no. 3, pp. 259–261, 2001.

[257] R. Prins, J. Z. Groenewald, G. F. Marais, J. W. Snape, and R. M. D. Koebner, "AFLP and STS tagging of *Lr19*, a gene conferring resistance to leaf rust in wheat," *Theoretical and Applied Genetics*, vol. 103, no. 4, pp. 618–624, 2001.

[258] W. J. Raupp, S. Singh, G.L. Brown-Guedira, and B. S. Gill, "Cytogenetic and molecular mapping of the leaf rust resistance gene *Lr39* in wheat," *Theoretical and Applied Genetics*, vol. 102, no. 2-3, pp. 347–352, 2001.

[259] S. Seah, H. Bariana, J. Jahier, K. Sivasithamparam, and E. S. Lagudah, "The introgressed segment carrying rust resistance genes *Yr17*, *Lr37* and *Sr38* in wheat can be assayed by a cloned disease resistance gene-like sequence," *Theoretical and Applied Genetics*, vol. 102, no. 4, pp. 600–605, 2001.

[260] C. Neu, N. Stein, and B. Keller, "Genetic mapping of the *Lr20-Pm1* resistance locus reveals suppressed recombination on chromosome arm 7AL in hexaploid wheat," *Genome*, vol. 45, no. 4, pp. 737–744, 2002.

[261] D. P. Cherukuri, S. K. Gupta, A. Charpe, et al., "Identification of a molecular marker linked to an *Agropyron elongatum*-derived gene *Lr19* for leaf rust resistance in wheat," *Plant Breeding*, vol. 122, no. 3, pp. 204–208, 2003.

[262] H.-Q. Ling, J. Qiu, R. P. Singh, and B. Keller, "Identification and genetic characterization of an *Aegilops tauschii* ortholog of the wheat leaf rust disease resistance gene *Lr*1," *Theoretical and Applied Genetics*, vol. 109, no. 6, pp. 1133–1138, 2004.

[263] D. P. Cherukuri, S. K. Gupta, A. Charpe, et al., "Molecular mapping of *Aegilops speltoides* derived leaf rust resistance gene *Lr28* in wheat," *Euphytica*, vol. 143, no. 1-2, pp. 19–26, 2005.

[264] W. Spielmeyer, R. A. McIntosh, J. Kolmer, and E. S. Lagudah, "Powdery mildew resistance and *Lr34/Yr18* genes for durable resistance to leaf and stripe rust cosegregate at a locus on the short arm of chromosome 7D of wheat," *Theoretical and Applied Genetics*, vol. 111, no. 4, pp. 731–735, 2005.

[265] C. W. Hiebert, J. B. Thomas, and B. D. McCallum, "Locating the broad-spectrum wheat leaf rust resistance gene *Lr*52(*LrW*) to chromosome 5B by a new cytogenetic method," *Theoretical and Applied Genetics*, vol. 110, no. 8, pp. 1453–1457, 2005.

[266] S. K. Gupta, A. Charpe, K. V. Prabhu, and Q. M. R. Haque, "Identification and validation of molecular markers linked to the leaf rust resistance gene *Lr*19 in wheat," *Theoretical and Applied Genetics*, vol. 113, no. 6, pp. 1027–1036, 2006.

[267] E. Bossolini, S. G. Krattinger, and B. Keller, "Development of simple sequence repeat markers specific for the *Lr34* resistance region of wheat using sequence information from rice and *Aegilops tauschii*," *Theoretical and Applied Genetics*, vol. 113, no. 6, pp. 1049–1062, 2006.

[268] C. W. Hiebert, J. B. Thomas, D. J. Somers, B. D. McCallum, and S. L. Fox, "Microsatellite mapping of adult-plant leaf rust resistance gene *Lr22a* in wheat," *Theoretical and Applied Genetics*, vol. 115, no. 6, pp. 877–884, 2007.

[269] J.-W. Qiu, A. C. Schürch, N. Yahiaoui, et al., "Physical mapping and identification of a candidate for the leaf rust resistance gene *Lr1* of wheat," *Theoretical and Applied Genetics*, vol. 115, no. 2, pp. 159–168, 2007.

[270] D. E. Obert, A. K. Fritz, J. L. Moran, S. Singh, J. C. Rudd, and M. A. Menz, "Identification and molecular tagging of a gene from PI 289824 conferring resistance to leaf rust (*Puccinia triticina*) in wheat," *Theoretical and Applied Genetics*, vol. 110, no. 8, pp. 1439–1444, 2005.

[271] T. Schnurbusch, S. Paillard, A. Schori, et al., "Dissection of quantitative and durable leaf rust resistance in Swiss winter wheat reveals a major resistance QTL in the *Lr34* chromosomal region," *Theoretical and Applied Genetics*, vol. 108, no. 3, pp. 477–484, 2004.

[272] I. N. Leonova, L. I. Laikova, O. M. Popova, O. Unger, A. Börner, and M. S. Röder, "Detection of quantitative trait loci for leaf rust resistance in wheat—*T. timopheevii/T. tauschii* introgression lines," *Euphytica*, vol. 155, no. 1-2, pp. 79–86, 2007.

[273] R. P. Singh, J. C. Nelson, and M. E. Sorrells, "Mapping *Yr28* and other genes for resistance to stripe rust in wheat," *Crop Science*, vol. 40, no. 4, pp. 1148–1155, 2000.

[274] G. L. Sun, T. Fahima, A. B. Korol, et al., "Identification of molecular markers linked to the *Yr15* stripe rust resistance gene of wheat originated in wild emmer wheat, *Triticum dicoccoides*," *Theoretical and Applied Genetics*, vol. 95, no. 4, pp. 622–628, 1997.

[275] J. H. Peng, T. Fahima, M. S. Röder, et al., "Microsatellite tagging of the stripe-rust resistance gene *YrH52* derived from wild emmer wheat, *Triticum dicoccoides*, and suggestive negative crossover interference on chromosome 1B," *Theoretical and Applied Genetics*, vol. 98, no. 6-7, pp. 862–872, 1999.

[276] A. Börner, M. S. Röder, O. Unger, and A. Meinel, "The detection and molecular mapping of a major gene for non-specific adult-plant disease resistance against stripe rust (*Puccinia striiformis*) in wheat," *Theoretical and Applied Genetics*, vol. 100, no. 7, pp. 1095–1099, 2000.

[277] J. H. Peng, T. Fahima, M. S. Röder, et al., "High-density molecular map of chromosome region harboring stripe-rust resistance genes *YrH52* and *Yr15* derived from wild emmer wheat, *Triticum dicoccoides*," *Genetica*, vol. 109, no. 3, pp. 199–210, 2001.

[278] Z. X. Shi, X. M. Chen, R. F. Line, H. Leung, and C. R. Wellings, "Development of resistance gene analog polymorphism markers for the *Yr9* gene resistance to wheat stripe rust," *Genome*, vol. 44, no. 4, pp. 509–516, 2001.

[279] J. Ma, R. Zhou, Y. Dong, L. Wang, X. Wang, and J. Jia, "Molecular mapping and detection of the yellow rust resistance gene *Yr26* in wheat transferred from *Triticum turgidum* L. using microsatellite markers," *Euphytica*, vol. 120, no. 2, pp. 219–226, 2001.

[280] L. Wang, J. Ma, R. Zhou, X. Wang, and J. Jia, "Molecular tagging of the yellow rust resistance gene *Yr10* in common wheat, P.I.178383 (*Triticum aestivum* L.)," *Euphytica*, vol. 124, no. 1, pp. 71–73, 2002.

[281] C. Uauy, J. C. Brevis, X. Chen, et al., "High-temperature adult-plant (HTAP) stripe rust resistance gene *Yr36* from *Triticum turgidum* ssp. *dicoccoides* is closely linked to the grain protein content locus Gpc-B1," *Theoretical and Applied Genetics*, vol. 112, no. 1, pp. 97–105, 2005.

[282] G. Q. Li, Z. F. Li, W. Y. Yang, et al., "Molecular mapping of stripe rust resistance gene *YrCH*42 in Chinese wheat cultivar Chuanmai 42 and its allelism with *Yr*24 and *Yr*26," *Theoretical and applied genetics*, vol. 112, no. 8, pp. 1434–1440, 2006.

[283] Z. F. Li, T. C. Zheng, Z. H. He, et al., "Molecular tagging of stripe rust resistance gene *YrZH*84 in Chinese wheat line Zhou 8425B," *Theoretical and Applied Genetics*, vol. 112, no. 6, pp. 1098–1103, 2006.

[284] H. S. Bariana, N. Parry, I. R. Barclay, et al., "Identification and characterization of stripe rust resistance gene *Yr34* in common wheat," *Theoretical and Applied Genetics*, vol. 112, no. 6, pp. 1143–1148, 2006.

[285] C. Wang, Y. Zhang, D. Han, et al., "SSR and STS markers for wheat stripe rust resistance gene *Yr26*," *Euphytica*, vol. 159, no. 3, pp. 359–366, 2008.

[286] S. Mallard, D. Gaudet, A. Aldeia, et al., "Genetic analysis of durable resistance to yellow rust in bread wheat," *Theoretical and Applied Genetics*, vol. 110, no. 8, pp. 1401–1409, 2005.

[287] M. J. Christiansen, B. Feenstra, I. M. Skovgaard, and S. B. Andersen, "Genetic analysis of resistance to yellow rust in hexaploid wheat using a mixture model for multiple crosses," *Theoretical and Applied Genetics*, vol. 112, no. 4, pp. 581–591, 2006.

[288] J. G. Paull, M. A. Pallotta, P. Langridge, and T. T. The, "RFLP markers associated with *Sr22* and recombination between chromosome 7A of bread wheat and the diploid species Triticum boeoticum," *Theoretical and Applied Genetics*, vol. 89, no. 7-8, pp. 1039–1045, 1994.

[289] W. Spielmeyer, P. J. Sharp, and E. S. Lagudah, "Identification and validation of markers linked to broad-spectrum stem rust resistance gene *Sr2* in wheat (*Triticum aestivum* L.)," *Crop Science*, vol. 43, no. 1, pp. 333–336, 2003.

[290] P. A. Cuthbert, D. J. Somers, and A. Brulé-Babel, "Mapping of *Fhb2* on chromosome 6BS: a gene controlling Fusarium head blight field resistance in bread wheat (*Triticum aestivum* L.)," *Theoretical and Applied Genetics*, vol. 114, no. 3, pp. 429–437, 2007.

[291] W. Bourdoncle and H. W. Ohm, "Quantitative trait loci for resistance to Fusarium head blight in recombinant inbred wheat lines from the cross huapei 57-2/Patterson," *Euphytica*, vol. 131, no. 1, pp. 131–136, 2003.

[292] I. A. del Blanco, R. C. Frohberg, R. W. Stack, W. A. Berzonsky, and S. F. Kianian, "Detection of QTL linked to Fusarium head blight resistance in Sumai 3-derived North Dakota bread wheat lines," *Theoretical and Applied Genetics*, vol. 106, no. 6, pp. 1027–1031, 2003.

[293] F. Lin, Z. X. Kong, H. L. Zhu, et al., "Mapping QTL associated with resistance to Fusarium head blight in the Nanda2419 × Wangshuibai population—I: type II resistance," *Theoretical and Applied Genetics*, vol. 109, no. 7, pp. 1504–1511, 2004.

[294] F. Lin, S. L. Xue, Z. Z. Zhang, et al., "Mapping QTL associated with resistance to Fusarium head blight in the Nanda2419 × Wangshuibai population—II: type I resistance," *Theoretical and Applied Genetics*, vol. 112, no. 3, pp. 528–535, 2006.

[295] S. Paillard, T. Schnurbusch, R. Tiwari, et al., "QTL analysis of resistance to Fusarium head blight in Swiss winter wheat (*Triticum aestivum* L.)," *Theoretical and Applied Genetics*, vol. 109, no. 2, pp. 323–332, 2004.

[296] J. Gilsinger, L. Kong, X. Shen, and H. Ohm, "DNA markers associated with low Fusarium head blight incidence and narrow flower opening in wheat," *Theoretical and Applied Genetics*, vol. 110, no. 7, pp. 1218–1225, 2005.

[297] G. Jia, P. Chen, G. Qin, et al., "QTLs for Fusarium head blight response in a wheat DH population of Wangshuibai/Alondra's," *Euphytica*, vol. 146, no. 3, pp. 183–191, 2005.

[298] X. Chen, J. D. Faris, J. Hu, et al., "Saturation and comparative mapping of a major Fusarium head blight resistance QTL in tetraploid wheat," *Molecular Breeding*, vol. 19, no. 2, pp. 113–124, 2007.

[299] G.-L. Jiang, Y. Dong, J. Shi, and R. W. Ward, "QTL analysis of resistance to Fusarium head blight in the novel wheat germplasm CJ 9306—II: resistance to deoxynivalenol accumulation and grain yield loss," *Theoretical and Applied Genetics*, vol. 115, no. 8, pp. 1043–1052, 2007.

[300] A. Klahr, G. Zimmermann, G. Wenzel, and V. Mohler, "Effects of environment, disease progress, plant height and heading date on the detection of QTLs for resistance to Fusarium head blight in an European winter wheat cross," *Euphytica*, vol. 154, no. 1-2, pp. 17–28, 2007.

[301] X. Shen and H. Ohm, "Molecular mapping of *Thinopyrum*-derived *Fusarium* head blight resistance in common wheat," *Molecular Breeding*, vol. 20, no. 2, pp. 131–140, 2007.

[302] W. Zhou, F. L. Kolb, G. Bai, G. Shaner, and L. L. Domier, "Genetic analysis of scab resistance QTL in wheat with microsatellite and AFLP markers," *Genome*, vol. 45, no. 4, pp. 719–727, 2002.

[303] W.-C. Zhou, F. L. Kolb, G.-H. Bai, L. L. Domier, L. K. Boze, and N. J. Smith, "Validation of a major QTL for scab resistance with SSR markers and use of marker-assisted selection in wheat," *Plant Breeding*, vol. 122, no. 1, pp. 40–46, 2003.

[304] L. Hartl, H. Weiss, U. Stephan, F. J. Zeller, and A. Jahoor, "Molecular identification of powdery mildew resistance genes in common wheat (*Triticum aestivum* L.)," *Theoretical and Applied Genetics*, vol. 90, no. 5, pp. 601–606, 1995.

[305] J. Jia, K. M. Devos, S. Chao, T. E. Miller, S. M. Reader, and M. D. Gale, "RFLP-based maps of the homoeologous group-6 chromosomes of wheat and their application in the tagging of *Pm*12, a powdery mildew resistance gene transferred from *Aegilops speltoides* to wheat," *Theoretical and Applied Genetics*, vol. 92, no. 5, pp. 559–565, 1996.

[306] Z. Liu, Q. Sun, Z. Ni, and T. Yang, "Development of SCAR markers linked to the *Pm*21 gene conferring resistance to powdery mildew in common wheat," *Plant Breeding*, vol. 118, no. 3, pp. 215–219, 1999.

[307] P. Sourdille, P. Robe, M.-H. Tixier, G. Doussinault, M.-T. Pavoinc, and M. Bernard, "Location of *Pm3g*, a powdery mildew resistance allele in wheat, by using a monosomic analysis and by identifying associated molecular markers," *Euphytica*, vol. 110, no. 3, pp. 193–198, 1999.

[308] X. Q. Huang, S. L. K. Hsam, F. J. Zeller, G. Wenzel, and V. Mohler, "Molecular mapping of the wheat powdery mildew resistance gene *Pm*24 and marker validation for molecular breeding," *Theoretical and Applied Genetics*, vol. 101, no. 3, pp. 407–414, 2000.

[309] J. K. Rong, E. Millet, J. Manisterski, and M. Feldman, "A new powdery mildew resistance gene: introgression from

wild emmer into common wheat and RFLP-based mapping," *Euphytica*, vol. 115, no. 2, pp. 121–126, 2000.

[310] W. J. Tao, D. Liu, J. Y. Liu, Y. Feng, and P. Chen, "Genetic mapping of the powdery mildew resistance gene *Pm6* in wheat by RFLP analysis," *Theoretical and Applied Genetics*, vol. 100, no. 3-4, pp. 564–568, 2000.

[311] K. Järve, H. O. Peusha, J. Tsymbalova, S. Tamm, K. M. Devos, and T. M. Enno, "Chromosomal location of a *Triticum timopheevii*-derived powdery mildew resistance gene transferred to common wheat," *Genome*, vol. 43, no. 2, pp. 377–381, 2000.

[312] V. Mohler, S. L. K. Hsam, F. J. Zeller, and G. Wenzel, "An STS marker distinguishing the rye-derived powdery mildew resistance alleles at the *Pm8/Pm17* locus of common wheat," *Plant Breeding*, vol. 120, no. 5, pp. 448–450, 2001.

[313] Y. Bougot, J. Lemoine, M. T. Pavoine, D. Barloy, and G. Doussinault, "Identification of a microsatellite marker associated with *Pm3* resistance alleles to powdery mildew in wheat," *Plant Breeding*, vol. 121, no. 4, pp. 325–329, 2002.

[314] F. J. Zeller, L. Kong, L. Hartl, V. Mohler, and S. L. K. Hsam, "Chromosomal location of genes for resistance to powdery mildew in common wheat (*Triticum aestivum* L. em Thell.) 7. Gene *Pm29* in line Pova," *Euphytica*, vol. 123, no. 2, pp. 187–194, 2002.

[315] Z. Liu, Q. Sun, Z. Ni, E. Nevo, and T. Yang, "Molecular characterization of a novel powdery mildew resistance gene *Pm*30 in wheat originating from wild emmer," *Euphytica*, vol. 123, no. 1, pp. 21–29, 2002.

[316] C. Alberto, D. Renato, T. O. Antonio, C. Carla, P. Marina, and P. Enrico, "Genetic analysis of the *Aegilops longissima* 3S chromosome carrying the *Pm13* resistance gene," *Euphytica*, vol. 130, no. 2, pp. 177–183, 2003.

[317] Z.-Q. Ma, J.-B. Wei, and S.-H. Cheng, "PCR-based markers for the powdery mildew resistance gene *Pm4a* in wheat," *Theoretical and Applied Genetics*, vol. 109, no. 1, pp. 140–145, 2004.

[318] Y. C. Qiu, R. H. Zhou, X. Y. Kong, S. S. Zhang, and J. Z. Jia, "Microsatellite mapping of a *Triticum urartu* Tum. derived powdery mildew resistance gene transferred to common wheat (*Triticum aestivum* L.)," *Theoretical and Applied Genetics*, vol. 111, no. 8, pp. 1524–1531, 2005.

[319] L. M. Miranda, J. P. Murphy, D. Marshall, and S. Leath, "*Pm34*: a new powdery mildew resistance gene transferred from *Aegilops tauschii* Coss. to common wheat (*Triticum aestivum* L.)," *Theoretical and Applied Genetics*, vol. 113, no. 8, pp. 1497–1504, 2006.

[320] Z. Zhu, R. Zhou, X. Kong, Y. Dong, and J. Jia, "Microsatellite marker identification of a *Triticum aestivum—Aegilops umbellulata* substitution line with powdery mildew resistance," *Euphytica*, vol. 150, no. 1-2, pp. 149–153, 2006.

[321] L. M. Miranda, J. P. Murphy, D. Marshall, C. Cowger, and S. Leath, "Chromosomal location of *Pm35*, a novel *Aegilops tauschii* derived powdery mildew resistance gene introgressed into common wheat (*Triticum aestivum* L.)," *Theoretical and Applied Genetics*, vol. 114, no. 8, pp. 1451–1456, 2007.

[322] G. Nematollahi, V. Mohler, G. Wenzel, F. J. Zeller, and S. L. K. Hsam, "Microsatellite mapping of powdery mildew resistance allele *Pm5d* from common wheat line IGV1-455," *Euphytica*, vol. 159, no. 3, pp. 307–313, 2008.

[323] W. Song, H. Xie, Q. Liu, et al., "Molecular identification of *Pm12*-carrying introgression lines in wheat using genomic and EST-SSR markers," *Euphytica*, vol. 158, no. 1-2, pp. 95–102, 2007.

[324] N. Chantret, P. Sourdille, M. Röder, M. Tavaud, M. Bernard, and G. Doussinault, "Location and mapping of the powdery mildew resistance gene *MlRE* and detection of a resistance QTL by bulked segregant analysis (BSA) with microsatellites in wheat," *Theoretical and Applied Genetics*, vol. 100, no. 8, pp. 1217–1224, 2000.

[325] C. Xie, Q. Sun, Z. Ni, T. Yang, E. Nevo, and T. Fahima, "Chromosomal location of a *Triticum dicoccoides*-derived powdery mildew resistance gene in common wheat by using microsatellite markers," *Theoretical and Applied Genetics*, vol. 106, no. 2, pp. 341–345, 2003.

[326] Ch. Singrün, S. L. K. Hsam, F. J. Zeller, G. Wenzel, and V. Mohler, "Localization of a novel recessive powdery mildew resistance gene from common wheat line RD30 in the terminal region of chromosome 7AL," *Theoretical and Applied Genetics*, vol. 109, no. 1, pp. 210–214, 2004.

[327] M. Keller, B. Keller, G. Schachermayr, et al., "Quantitative trait loci for resistance against powdery mildew in a segregating wheat × spelt population," *Theoretical and Applied Genetics*, vol. 98, no. 6-7, pp. 903–912, 1999.

[328] S. Liu, C. A. Griffey, and M. A. Saghai Maroof, "Identification of molecular markers associated with adult plant resistance to powdery mildew in common wheat cultivar Massey," *Crop Science*, vol. 41, no. 4, pp. 1268–1275, 2001.

[329] D. Mingeot, N. Chantret, P. V. Baret, et al., "Mapping QTL involved in adult plant resistance to powdery mildew in the winter wheat line RE714 in two susceptible genetic backgrounds," *Plant Breeding*, vol. 121, no. 2, pp. 133–140, 2002.

[330] Y. Bougot, J. Lemoine, M. T. Pavoine, et al., "A major QTL effect controlling resistance to powdery mildew in winter wheat at the adult plant stage," *Plant Breeding*, vol. 125, no. 6, pp. 550–556, 2006.

[331] D. M. Tucker, C. A. Griffey, S. Liu, G. Brown-Guedira, D. S. Marshall, and M. A. S. Maroof, "Confirmation of three quantitative trait loci conferring adult plant resistance to powdery mildew in two winter wheat populations," *Euphytica*, vol. 155, no. 1-2, pp. 1–13, 2007.

[332] A. Laroche, T. Demeke, D. A. Gaudet, B. Puchalski, M. Frick, and R. McKenzie, "Development of a PCR marker for rapid identification of the *Bt*-10 gene for common bunt resistance in wheat," *Genome*, vol. 43, no. 2, pp. 217–223, 2000.

[333] B. Fofana, D. G. Humphreys, S. Cloutier, C. A. McCartney, and D. J. Somers, "Mapping quantitative trait loci controlling common bunt resistance in a doubled haploid population derived from the spring wheat cross RL4452 × AC Domain," *Molecular Breeding*, vol. 21, no. 3, pp. 317–325, 2008.

[334] J. D. Faris, J. A. Anderson, L. J. Francl, and J. G. Jordahl, "RFLP mapping of resistance to chlorosis induction by *Pyrenophora tritici-repentis* in wheat," *Theoretical and Applied Genetics*, vol. 94, no. 1, pp. 98–103, 1997.

[335] P. K. Singh, M. Mergoum, T. B. Adhikari, S. F. Kianian, and E. M. Elias, "Chromosomal location of genes for seedling resistance to tan spot and *Stagonospora nodorum* blotch in tetraploid wheat," *Euphytica*, vol. 155, no. 1-2, pp. 27–34, 2007.

[336] W. Tadesse, M. Schmolke, S. L. K. Hsam, V. Mohler, G. Wenzel, and F. J. Zeller, "Molecular mapping of resistance genes to tan spot [*Pyrenophora tritici-repentis* race 1] in synthetic wheat lines," *Theoretical and Applied Genetics*, vol. 114, no. 5, pp. 855–862, 2007.

[337] L. S. Arraiano, A. J. Worland, C. Ellerbrook, and J. K. M. Brown, "Chromosomal location of a gene for resistance

to septoria tritici blotch (*Mycosphaerella graminicola*) in the hexaploid wheat 'Synthetic 6x'," *Theoretical and Applied Genetics*, vol. 103, no. 5, pp. 758–764, 2001.

[338] M. R. Simón, F. M. Ayala, C. A. Cordo, M. S. Röder, and A. Börner, "Molecular mapping of quantitative trait loci determining resistance to septoria tritici blotch caused by *Mycosphaerella graminicola* in wheat," *Euphytica*, vol. 138, no. 1, pp. 41–48, 2004.

[339] L. Ayala, M. Henry, M. van Ginkel, R. Singh, B. Keller, and M. Khairallah, "Identification of QTLs for BYDV tolerance in bread wheat," *Euphytica*, vol. 128, no. 2, pp. 249–259, 2002.

[340] V. Aguilar, P. Stamp, M. Winzeler, et al., "Inheritance of field resistance to *Stagonospora nodorum* leaf and glume blotch and correlations with other morphological traits in hexaploid wheat (*Triticum aestivum* L.)," *Theoretical and Applied Genetics*, vol. 111, no. 2, pp. 325–336, 2005.

[341] L. E. Talbert, P. L. Bruckner, L. Y. Smith, R. Sears, and T. J. Martin, "Development of PCR markers linked to resistance to wheat streak mosaic virus in wheat," *Theoretical and Applied Genetics*, vol. 93, no. 3, pp. 463–467, 1996.

[342] A. A. Khan, G. C. Bergstrom, J. C. Nelson, and M. E. Sorrells, "Identification of RFLP markers for resistance to wheat spindle streak mosaic bymovirus (WSSMV) disease," *Genome*, vol. 43, no. 3, pp. 477–482, 2000.

[343] W. Liu, H. Nie, S. Wang, et al., "Mapping a resistance gene in wheat cultivar Yangfu 9311 to yellow mosaic virus, using microsatellite markers," *Theoretical and Applied Genetics*, vol. 111, no. 4, pp. 651–657, 2005.

[344] R. C. de la Peña, T. D. Murray, and S. S. Jones, "Identification of an RFLP interval containing *Pch2* on chromosome 7AL in wheat," *Genome*, vol. 40, no. 2, pp. 249–252, 1997.

[345] V. Huguet-Robert, F. Dedryver, M. S. Röder, et al., "Isolation of a chromosomally engineered durum wheat line carrying the *Aegilops ventricosa Pch*1 gene for resistance to eyespot," *Genome*, vol. 44, no. 3, pp. 345–349, 2001.

[346] J. Z. Groenewald, A. S. Marais, and G. F. Marais, "Amplified fragment length polymorphism-derived microsatellite sequence linked to the *Pch*1 and *Ep-D*1 loci in common wheat," *Plant Breeding*, vol. 122, no. 1, pp. 83–85, 2003.

[347] Y. Weng and M. D. Lazar, "Amplified fragment length polymorphism- and simple sequence repeat-based molecular tagging and mapping of greenbug resistance gene *Gb3* in wheat," *Plant Breeding*, vol. 121, no. 3, pp. 218–223, 2002.

[348] E. Boyko, S. Starkey, and M. Smith, "Molecular genetic mapping of *Gby*, a new greenbug resistance gene in bread wheat," *Theoretical and Applied Genetics*, vol. 109, no. 6, pp. 1230–1236, 2004.

[349] Y. Weng, W. Li, R. N. Devkota, and J. C. Rudd, "Microsatellite markers associated with two *Aegilops tauschii*-derived greenbug resistance loci in wheat," *Theoretical and Applied Genetics*, vol. 110, no. 3, pp. 462–469, 2005.

[350] L. C. Zhu, C. M. Smith, A. Fritz, E. Boyko, P. Voothuluru, and B. S. Gill, "Inheritance and molecular mapping of new greenbug resistance genes in wheat germplasms derived from *Aegilops tauschii*," *Theoretical and Applied Genetics*, vol. 111, no. 5, pp. 831–837, 2005.

[351] Z.-Q. Ma, B. S. Gill, M. E. Sorrells, and S. D. Tanksley, "RELP markers linked to two Hessian fly-resistance genes in wheat (*Triticum aestivum* L.) from *Triticum tauschii* (coss.) Schmal," *Theoretical and Applied Genetics*, vol. 85, no. 6-7, pp. 750–754, 1993.

[352] I. Dweikat, H. W. Ohm, S. Mackenzie, F. Patterson, S. Cambron, and R. Ratcliffe, "Association of a DNA marker with Hessian fly resistance gene *H9* in wheat," *Theoretical and Applied Genetics*, vol. 89, no. 7-8, pp. 964–968, 1994.

[353] I. Dweikat, H. W. Ohm, F. Patterson, and S. Cambron, "Identification of RAPD markers for 11 Hessian fly resistance genes in wheat," *Theoretical and Applied Genetics*, vol. 94, no. 3-4, pp. 419–423, 1997.

[354] Y. W. Seo, J. W. Johnson, and R. L. Jarret, "A molecular marker associated with the *H21* Hessian fly resistance gene in wheat," *Molecular Breeding*, vol. 3, no. 3, pp. 177–181, 1997.

[355] I. Dweikat, W. Zhang, and H. W. Ohm, "Development of STS markers linked to Hessian fly resistance gene *H6* in wheat," *Theoretical and Applied Genetics*, vol. 105, no. 5, pp. 766–770, 2002.

[356] X. M. Liu, B. S. Gill, and M.-S. Chen, "Hessian fly resistance gene *H*13 is mapped to a distal cluster of resistance genes in chromosome 6DS of wheat," *Theoretical and Applied Genetics*, vol. 111, no. 2, pp. 243–249, 2005.

[357] T. Wang, S. S. Xu, M. O. Harris, J. Hu, L. Liu, and X. Cai, "Genetic characterization and molecular mapping of Hessian fly resistance genes derived from *Aegilops tauschii* in synthetic wheat," *Theoretical and Applied Genetics*, vol. 113, no. 4, pp. 611–618, 2006.

[358] H. X. Zhao, X. M. Liu, and M.-S. Chen, "*H22*, a major resistance gene to the Hessian fly (*Mayetiola destructor*), is mapped to the distal region of wheat chromosome 1DS," *Theoretical and Applied Genetics*, vol. 113, no. 8, pp. 1491–1496, 2006.

[359] L. Kong, S. E. Cambron, and H. W. Ohm, "Hessian fly resistance genes *H*16 and *H*17 are mapped to a resistance gene cluster in the distal region of chromosome 1AS in wheat," *Molecular Breeding*, vol. 21, no. 2, pp. 183–194, 2008.

[360] X. M. Liu, C. M. Smith, B. S. Gill, and V. Tolmay, "Microsatellite markers linked to six Russian wheat aphid resistance genes in wheat," *Theoretical and Applied Genetics*, vol. 102, no. 4, pp. 504–510, 2001.

[361] C. A. Miller, A. Altinkut, and N. L. V. Lapitan, "A microsatellite marker for tagging *Dn2*, a wheat gene conferring resistance to the Russian wheat aphid," *Crop Science*, vol. 41, no. 5, pp. 1584–1589, 2001.

[362] X. M. Liu, C. M. Smith, and B. S. Gill, "Identification of microsatellite markers linked to Russian wheat aphid resistance genes *Dn4* and *Dn6*," *Theoretical and Applied Genetics*, vol. 104, no. 6-7, pp. 1042–1048, 2002.

[363] K. J. Williams, J. M. Fisher, and P. Langridge, "Identification of RFLP markers linked to the cereal cyst nematode resistance gene (*Cre*) in wheat," *Theoretical and Applied Genetics*, vol. 89, no. 7-8, pp. 927–930, 1994.

[364] J. Jahier, P. Abelard, A. M. Tanguy, et al., "The *Aegilops ventricosa* segment on chromosome 2AS of the wheat cultivar 'VPM1' carries the cereal cyst nematode resistance gene *Cre*5," *Plant Breeding*, vol. 120, no. 2, pp. 125–128, 2001.

[365] F. C. Ogbonnaya, S. Seah, A. Delibes, et al., "Molecular-genetic characterisation of a new nematode resistance gene in wheat," *Theoretical and Applied Genetics*, vol. 102, no. 4, pp. 623–629, 2001.

[366] D. Barloy, J. Lemoine, F. Dredryver, and J. Jahier, "Molecular markers linked to the *Aegilops variabilis*-derived root-knot nematode resistance gene *Rkn-mn1* in wheat," *Plant Breeding*, vol. 119, no. 2, pp. 169–172, 2000.

[367] K. J. Williams, S. P. Taylor, P. Bogacki, M. Pallotta, H. S. Bariana, and H. Wallwork, "Mapping of the root lesion nematode (*Pratylenchus neglectus*) resistance gene *Rlnn1* in

wheat," *Theoretical and Applied Genetics*, vol. 104, no. 5, pp. 874–879, 2002.

[368] K. Kato, W. Nakamura, T. Tabiki, H. Miura, and S. Sawada, "Detection of loci controlling seed dormancy on group 4 chromosomes of wheat and comparative mapping with rice and barley genomes," *Theoretical and Applied Genetics*, vol. 102, no. 6-7, pp. 980–985, 2001.

[369] J. Flintham, R. Adlam, M. Bassoi, M. Holdsworth, and M. D. Gale, "Mapping genes for resistance to sprouting damage in wheat," *Euphytica*, vol. 126, no. 1, pp. 39–45, 2002.

[370] H. Miura, N. Sato, K. Kato, and Y. Amano, "Detection of chromosomes carrying genes for seed dormancy of wheat using the backcross reciprocal monosomic method," *Plant Breeding*, vol. 121, no. 5, pp. 394–399, 2002.

[371] M. Osa, K. Kato, M. Mori, C. Shindo, A. Torada, and H. Miura, "Mapping QTLs for seed dormancy and the *Vp1* homologue on chromosome 3A in wheat," *Theoretical and Applied Genetics*, vol. 106, no. 8, pp. 1491–1496, 2003.

[372] P. L. Kulwal, N. Kumar, A. Gaur, et al., "Mapping of a major QTL for pre-harvest sprouting tolerance on chromosome 3A in bread wheat," *Theoretical and Applied Genetics*, vol. 111, no. 6, pp. 1052–1059, 2005.

[373] D. Mares, K. Mrva, J. Cheong, et al., "A QTL located on chromosome 4A associated with dormancy in white- and red-grained wheats of diverse origin," *Theoretical and Applied Genetics*, vol. 111, no. 7, pp. 1357–1364, 2005.

[374] L. R. Joppa, C. Du, G. E. Hart, and G. A. Hareland, "Mapping gene(s) for grain protein in tetraploid wheat (*Triticum turgidum* L.) using a population of recombinant inbred chromosome lines," *Crop Science*, vol. 37, no. 5, pp. 1586–1589, 1997.

[375] M. Prasad, N. Kumar, P. L. Kulwal, et al., "QTL analysis for grain protein content using SSR markers and validation studies using NILs in bread wheat," *Theoretical and Applied Genetics*, vol. 106, no. 4, pp. 659–667, 2003.

[376] A. Blanco, R. Simeone, and A. Gadaleta, "Detection of QTLs for grain protein content in durum wheat," *Theoretical and Applied Genetics*, vol. 112, no. 7, pp. 1195–1204, 2006.

[377] G. D. Parker, K. J. Chalmers, A. J. Rathjen, and P. Langridge, "Mapping loci associated with flour colour in wheat (*Triticum aestivum* L.)," *Theoretical and Applied Genetics*, vol. 97, no. 1-2, pp. 238–245, 1998.

[378] G. D. Parker, K. J. Chalmers, A. J. Rathjen, and P. Langridge, "Mapping loci associated with milling yield in wheat (*Triticum aestivum* L.)," *Molecular Breeding*, vol. 5, no. 6, pp. 561–568, 1999.

[379] M. R. Perretant, T. Cadalen, G. Charmet, et al., "QTL analysis of bread-making quality in wheat using a doubled haploid population," *Theoretical and Applied Genetics*, vol. 100, no. 8, pp. 1167–1175, 2000.

[380] G. Charmet, N. Robert, G. Branlard, L. Linossier, P. Martre, and E. Triboï, "Genetic analysis of dry matter and nitrogen accumulation and protein composition in wheat kernels," *Theoretical and Applied Genetics*, vol. 111, no. 3, pp. 540–550, 2005.

[381] W. Ma, R. Appels, F. Bekes, O. Larroque, M. K. Morell, and K. R. Gale, "Genetic characterisation of dough rheological properties in a wheat doubled haploid population: additive genetic effects and epistatic interactions," *Theoretical and Applied Genetics*, vol. 111, no. 3, pp. 410–422, 2005.

[382] M. Arbelbide and R. Bernardo, "Mixed-model QTL mapping for kernel hardness and dough strength in bread wheat," *Theoretical and Applied Genetics*, vol. 112, no. 5, pp. 885–890, 2006.

[383] O. Dobrovolskaya, V. S. Arbuzova, U. Lohwasser, M. S. Röder, and A. Börner, "Microsatellite mapping of complementary genes for purple grain colour in bread wheat (*Triticum aestivum*) L.," *Euphytica*, vol. 150, no. 3, pp. 355–364, 2006.

[384] J. C. Nelson, C. Andreescu, F. Breseghello, et al., "Quantitative trait locus analysis of wheat quality traits," *Euphytica*, vol. 149, no. 1-2, pp. 145–159, 2006.

[385] F. Chen, Z. Luo, Z. Zhang, G. Xia, and H. Min, "Variation and potential value in wheat breeding of low-molecular-weight glutenin subunit genes cloned by genomic and RT-PCR in a derivative of somatic introgression between common wheat and *Agropyron elongatum*," *Molecular Breeding*, vol. 20, no. 2, pp. 141–152, 2007.

[386] C. J. Pozniak, R. E. Knox, F. R. Clarke, and J. M. Clarke, "Identification of QTL and association of a phytoene synthase gene with endosperm colour in durum wheat," *Theoretical and Applied Genetics*, vol. 114, no. 3, pp. 525–537, 2007.

[387] T. Cadalen, P. Sourdille, G. Charmet, et al., "Molecular markers linked to genes affecting plant height in wheat using a doubled-haploid population," *Theoretical and Applied Genetics*, vol. 96, no. 6-7, pp. 933–940, 1998.

[388] V. Korzun, M. S. Röder, M. W. Ganal, A. J. Worland, and C. N. Law, "Genetic analysis of the dwarfing gene (*Rht8*) in wheat—I: molecular mapping of *Rht8* on the short arm of chromosome 2D of bread wheat (*Triticum aestivum* L.)," *Theoretical and Applied Genetics*, vol. 96, no. 8, pp. 1104–1109, 1998.

[389] M. H. Ellis, D. G. Bonnett, and G. J. Rebetzke, "A 192bp allele at the Xgwm261 locus is not always associated with the *Rht8* dwarfing gene in wheat (*Triticum aestivum* L.)," *Euphytica*, vol. 157, no. 1-2, pp. 209–214, 2007.

[390] V. Kuraparthy, S. Sood, H. S. Dhaliwal, P. Chhuneja, and B. S. Gill, "Identification and mapping of a tiller inhibition gene (*tin*3) in wheat," *Theoretical and Applied Genetics*, vol. 114, no. 2, pp. 285–294, 2007.

[391] K. Kosuge, N. Watanabe, T. Kuboyama, et al., "Cytological and microsatellite mapping of mutant genes for spherical grain and compact spikes in durum wheat," *Euphytica*, vol. 159, no. 3, pp. 289–296, 2008.

[392] K. Kato, H. Miura, and S. Sawada, "QTL mapping of genes controlling ear emergence time and plant height on chromosome 5A of wheat," *Theoretical and Applied Genetics*, vol. 98, no. 3-4, pp. 472–477, 1999.

[393] P. Sourdille, J. W. Snape, T. Cadalen, et al., "Detection of QTLs for heading time-and photoperiod response in wheat using a doubled-haploid population," *Genome*, vol. 43, no. 3, pp. 487–494, 2000.

[394] E. Hanocq, M. Niarquin, E. Heumez, M. Rousset, and J. Le Gouis, "Detection and mapping of QTL for earliness components in a bread wheat recombinant inbred lines population," *Theoretical and Applied Genetics*, vol. 110, no. 1, pp. 106–115, 2004.

[395] X. Xu, G. Bai, B. F. Carver, and G. E. Shaner, "A QTL for early heading in wheat cultivar Suwon 92," *Euphytica*, vol. 146, no. 3, pp. 233–237, 2005.

[396] E. Hanocq, A. Laperche, O. Jaminon, A.-L. Lainé, and J. Le Gouis, "Most significant genome regions involved in the control of earliness traits in bread wheat, as revealed by QTL meta-analysis," *Theoretical and Applied Genetics*, vol. 114, no. 3, pp. 569–584, 2007.

[397] K. Kato, H. Miura, and S. Sawada, "Mapping QTLs controlling grain yield and its components on chromosome 5A of wheat," *Theoretical and Applied Genetics*, vol. 101, no. 7, pp. 1114–1121, 2000.

[398] B. Narasimhamoorthy, B. S. Gill, A. K. Fritz, J. C. Nelson, and G. L. Brown-Guedira, "Advanced backcross QTL analysis of a hard winter wheat × synthetic wheat population," *Theoretical and Applied Genetics*, vol. 112, no. 5, pp. 787–796, 2006.

[399] F. M. Kirigwi, M. van Ginkel, G. Brown-Guedira, B. S. Gill, G. M. Paulsen, and A. K. Fritz, "Markers associated with a QTL for grain yield in wheat under drought," *Molecular Breeding*, vol. 20, no. 4, pp. 401–413, 2007.

[400] H. Kuchel, K. J. Williams, P. Langridge, H. A. Eagles, and S. P. Jefferies, "Genetic dissection of grain yield in bread wheat—I: QTL analysis," *Theoretical and Applied Genetics*, vol. 115, no. 8, pp. 1029–1041, 2007.

[401] Z. Ma, D. Zhao, C. Zhang, et al., "Molecular genetic analysis of five spike-related traits in wheat using RIL and immortalized F_2 populations," *Molecular Genetics and Genomics*, vol. 277, no. 1, pp. 31–42, 2007.

[402] N. Kumar, P. L. Kulwal, A. Gaur, et al., "QTL analysis for grain weight in common wheat," *Euphytica*, vol. 151, no. 2, pp. 135–144, 2006.

[403] K. Kato, H. Miura, M. Akiyama, M. Kuroshima, and S. Sawada, "RFLP mapping of the three major genes, *Vrn1*, *Q* and *B1*, on the long arm of chromosome 5A of wheat," *Euphytica*, vol. 101, no. 1, pp. 91–95, 1998.

[404] Z. S. Peng, C. Yen, and J. L. Yang, "Chromosomal location of genes for supernumerary spikelet in bread wheat," *Euphytica*, vol. 103, no. 1, pp. 109–114, 1998.

[405] E. Salina, A. Börner, I. Leonova, et al., "Microsatellite mapping of the induced sphaerococcoid mutation genes in *Triticum aestivum*," *Theoretical and Applied Genetics*, vol. 100, no. 5, pp. 686–689, 2000.

[406] L. Bullrich, M. L. Appendino, G. Tranquilli, S. Lewis, and J. Dubcovsky, "Mapping of a thermo-sensitive earliness per se gene on *Triticum monococcum* chromosome 1Am," *Theoretical and Applied Genetics*, vol. 105, no. 4, pp. 585–593, 2002.

[407] E. K. Khlestkina, E. G. Pestsova, M. S. Röder, and A. Börner, "Molecular mapping, phenotypic expression and geographical distribution of genes determining anthocyanin pigmentation of coleoptiles in wheat (*Triticum aestivum* L.)," *Theoretical and Applied Genetics*, vol. 104, no. 4, pp. 632–637, 2002.

[408] Q. H. Xing, Z. G. Ru, C. J. Zhou, et al., "Genetic analysis, molecular tagging and mapping of the thermo-sensitive genic male-sterile gene (*wtms1*) in wheat," *Theoretical and Applied Genetics*, vol. 107, no. 8, pp. 1500–1504, 2003.

[409] C.-G. Chu, J. D. Faris, T. L. Friesen, and S. S. Xu, "Molecular mapping of hybrid necrosis genes *Ne1* and *Ne2* in hexaploid wheat using microsatellite markers," *Theoretical and Applied Genetics*, vol. 112, no. 7, pp. 1374–1381, 2006.

[410] O. Dobrovolskaya, T. A. Pshenichnikova, V. S. Arbuzova, U. Lohwasser, M. S. Röder, and A. Börner, "Molecular mapping of genes determining hairy leaf character in common wheat with respect to other species of the *Triticeae*," *Euphytica*, vol. 155, no. 3, pp. 285–293, 2007.

[411] S. Houshmand, R. E. Knox, F. R. Clarke, and J. M. Clarke, "Microsatellite markers flanking a stem solidness gene on chromosome 3BL in durum wheat," *Molecular Breeding*, vol. 20, no. 3, pp. 261–270, 2007.

[412] M. Keller, Ch. Karutz, J. E. Schmid, et al., "Quantitative trait loci for lodging resistance in a segregating wheat × spelt population," *Theoretical and Applied Genetics*, vol. 98, no. 6-7, pp. 1171–1182, 1999.

[413] L. Hai, H. Guo, S. Xiao, et al., "Quantitative trait loci (QTL) of stem strength and related traits in a doubled-haploid population of wheat (*Triticum aestivum* L.)," *Euphytica*, vol. 141, no. 1-2, pp. 1–9, 2005.

[414] V. J. Nalam, M. I. Vales, C. J. W. Watson, S. F. Kianian, and O. Riera-Lizarazu, "Map-based analysis of genes affecting the brittle rachis character in tetraploid wheat (*Triticum turgidum* L.)," *Theoretical and Applied Genetics*, vol. 112, no. 2, pp. 373–381, 2006.

[415] G. J. Rebetzke, M. H. Ellis, D. G. Bonnett, and R. A. Richards, "Molecular mapping of genes for coleoptile growth in bread wheat (*Triticum aestivum* L.)," *Theoretical and Applied Genetics*, vol. 114, no. 7, pp. 1173–1183, 2007.

[416] G. Zhang and M. Mergoum, "Molecular mapping of kernel shattering and its association with *Fusarium* head blight resistance in a Sumai3 derived population," in *Theoretical and Applied Genetics*, vol. 115, pp. 757–766, October 2007.

[417] K. Kato, S. Kidou, H. Miura, and S. Sawada, "Molecular cloning of the wheat *CK2α* gene and detection of its linkage with *Vrn-A1* on chromosome 5A," *Theoretical and Applied Genetics*, vol. 104, no. 6-7, pp. 1071–1077, 2002.

[418] Q. Liu, Z. Ni, H. Peng, W. Song, Z. Liu, and Q. Sun, "Molecular mapping of a dominant non-glaucousness gene from synthetic hexaploid wheat (*Triticum aestivum* L.): molecular mapping of non-glaucousness gene in wheat," *Euphytica*, vol. 155, no. 1-2, pp. 71–78, 2007.

[419] A. Carrera, V. Echenique, W. Zhang, et al., "A deletion at the *Lpx-B1* locus is associated with low lipoxygenase activity and improved pasta color in durum wheat (*Triticum turgidum ssp. durum*)," *Journal of Cereal Science*, vol. 45, no. 1, pp. 67–77, 2007.

[420] X. Y. He, Z. H. He, L. P. Zhang, et al., "Allelic variation of *polyphenol oxidase (PPO)* genes located on chromosomes 2A and 2D and development of functional markers for the *PPO* genes in common wheat," *Theoretical and Applied Genetics*, vol. 115, no. 1, pp. 47–58, 2007.

[421] S. Nakamura, T. Komatsuda, and H. Miura, "Mapping diploid wheat homologues of *Arabidopsis* seed ABA signaling genes and QTLs for seed dormancy," *Theoretical and Applied Genetics*, vol. 114, no. 7, pp. 1129–1139, 2007.

[422] R. Raman, H. Raman, and P. Martin, "Functional gene markers for polyphenol oxidase locus in bread wheat (*Triticum aestivum* L.)," *Molecular Breeding*, vol. 19, no. 4, pp. 315–328, 2007.

[423] D.-L. Yang, R.-L. Jing, X.-P. Chang, and W. Li, "Identification of quantitative trait loci and environmental interactions for accumulation and remobilization of water-soluble carbohydrates in wheat (*Triticum aestivum* L.) stems," *Genetics*, vol. 176, no. 1, pp. 571–584, 2007.

[424] V. Mohler, R. Lukman, S. Ortiz-Islas, et al., "Genetic and physical mapping of photoperiod insensitive gene *Ppd-B1* in common wheat," *Euphytica*, vol. 138, no. 1, pp. 33–40, 2004.

[425] H. Raman, R. Raman, R. Wood, and P. Martin, "Repetitive indel markers within the *ALMT1* gene conditioning aluminium tolerance in wheat (*Triticum aestivum* L.)," *Molecular Breeding*, vol. 18, no. 2, pp. 171–183, 2006.

[426] L.-L. Zhou, G.-H. Bai, H.-X. Ma, and B. F. Carver, "Quantitative trait loci for aluminum resistance in wheat," *Molecular Breeding*, vol. 19, no. 2, pp. 153–161, 2007.

[427] S. P. Jefferies, M. A. Pallotta, J. G. Paull, et al., "Mapping and validation of chromosome regions conferring boron

toxicity tolerance in wheat (*Triticum aestivum*)," *Theoretical and Applied Genetics*, vol. 101, no. 5-6, pp. 767–777, 2000.

[428] B. Tóth, G. Galiba, E. Fehér, J. Sutka, and J. W. Snape, "Mapping genes affecting flowering time and frost resistance on chromosome 5B of wheat," *Theoretical and Applied Genetics*, vol. 107, no. 3, pp. 509–514, 2003.

[429] L. Ma, E. Zhou, N. Huo, R. Zhou, G. Wang, and J. Jia, "Genetic analysis of salt tolerance in a recombinant inbred population of wheat (*Triticum aestivum* L.)," *Euphytica*, vol. 153, no. 1-2, pp. 109–117, 2007.

[430] J. Peng, Y. Ronin, T. Fahima, et al., "Domestication quantitative trait loci in *Triticum dicoccoides*, the progenitor of wheat," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 5, pp. 2489–2494, 2003.

[431] C. Pozzi, L. Rossini, A. Vecchietti, and F. Salamini, "Gene and genome changes during domestication of cereals," in *Cereal Genomics*, P. K. Gupta, R. K. Varshney, et al., Eds., pp. 165–198, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2004.

[432] J. Dubcovsky and J. Dvorak, "Genome plasticity a key factor in the success of polyploid wheat under domestication," *Science*, vol. 316, no. 5833, pp. 1862–1866, 2007.

[433] C. Uauy, J. C. Brevis, and J. Dubcovsky, "The high grain protein content gene Gpc-B1 accelerates senescence and has pleiotropic effects on protein content in wheat," *Journal of Experimental Botany*, vol. 57, no. 11, pp. 2785–2794, 2006.

[434] S. A. Flint-Garcia, A.-C. Thuillet, J. Yu, et al., "Maize association population: a high-resolution platform for quantitative trait locus dissection," *The Plant Journal*, vol. 44, no. 6, pp. 1054–1064, 2005.

[435] J. Yu and E. S. Buckler, "Genetic association mapping and genome organization of maize," *Current Opinion in Biotechnology*, vol. 17, no. 2, pp. 155–160, 2006.

[436] F. Breseghello and M. E. Sorrells, "Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars," *Genetics*, vol. 172, no. 2, pp. 1165–1177, 2006.

[437] C. Ravel, S. Praud, and A. Murigneux, "Identification of *Glu-B1-1* as a candidate gene for the quantity of high-molecular-weight glutenin in bread wheat (*Triticum aestivum* L.) by means of an association study," *Theoretical and Applied Genetics*, vol. 112, no. 4, pp. 738–743, 2006.

[438] J. Crossa, J. Burgueño, S. Dreisigacker, et al., "Association analysis of historical bread wheat germplasm using additive genetic covariance of relatives and population structure," *Genetics*, vol. 177, no. 3, pp. 1889–1913, 2007.

[439] L. Tommasini, T. Schnurbusch, D. Fossati, F. Mascher, and B. Keller, "Association mapping of *Stagonospora nodorum* blotch resistance in modern European winter wheat varieties," *Theoretical and Applied Genetics*, vol. 115, no. 5, pp. 697–708, 2007.

[440] J. Dubcovsky, "Marker-assisted selection in public breeding programs: the wheat experience," *Crop Science*, vol. 44, no. 6, pp. 1895–1898, 2004.

[441] M. E. Sorrells, "Application of new knowledge, technologies, and strategies to wheat improvement," *Euphytica*, vol. 157, no. 3, pp. 299–306, 2007.

[442] H. A. Eagles, H. S. Bariana, F. C. Ogbonnaya, et al., "Implementation of markers in Australian wheat breeding," *Australian Journal of Agricultural Research*, vol. 52, no. 11-12, pp. 1349–1356, 2001.

[443] F. C. Ogbonnaya, N. C. Subrahmanyam, O. Moullet, et al., "Diagnostic DNA markers for cereal cyst nematode resistance in bread wheat," *Australian Journal of Agricultural Research*, vol. 52, no. 11-12, pp. 1367–1374, 2001.

[444] S. Landjeva, V. Korzun, and A. Börner, "Molecular markers: actual and potential contributions to wheat genome characterization and breeding," *Euphytica*, vol. 156, no. 3, pp. 271–296, 2007.

[445] H. Kuchel, G. Ye, R. Fox, and S. Jefferies, "Genetic and economic analysis of a targeted marker-assisted wheat breeding strategy," *Molecular Breeding*, vol. 16, no. 1, pp. 67–78, 2005.

[446] H. Kuchel, R. Fox, J. Reinheimer, et al., "The successful application of a marker-assisted wheat breeding strategy," *Molecular Breeding*, vol. 20, no. 4, pp. 295–308, 2007.

[447] H. M. William, R. Trethowan, and E. M. Crosby-Galvan, "Wheat breeding assisted by markers: CIMMYT's experience," *Euphytica*, vol. 157, no. 3, pp. 307–319, 2007.

[448] C. Lange and J. C. Whittaker, "On prediction of genetic values in marker-assisted selection," *Genetics*, vol. 159, no. 3, pp. 1375–1381, 2001.

[449] N. Radovanovic and S. Cloutier, "Gene-assisted selection for high molecular weight glutenin subunits in wheat doubled haploid breeding programs," *Molecular Breeding*, vol. 12, no. 1, pp. 51–59, 2003.

[450] C. M. Bowman, C. J. Howe, and T. A. Dyer, "Molecular mechanisms contributing to the evolution of (wheat) chloroplast genomes," in *Proceedings of the 7th International Wheat Genetics Symposium*, T. E. Miller and R. M. D. Koebner, Eds., pp. 69–73, Cambridge, UK, July 1988.

[451] Y. Ogihara, "Genome science of polyploid wheat," in *Frontiers of Wheat Bioscience. The 100th Memorial Issue of Wheat Information Service*, K. Tsunewaki, Ed., pp. 169–184, Kihara Memorial Yokohama Foundation for the Advancement of Life Sciences, Yokohama, Japan, 2005.

[452] K. Tsunewaki, "Plasmon differentiation in *Triticum* and *Aegilops* revealed by cytoplasmic effects on the wheat genome manifestation," in *Proceedings of the US-Japan Symposium on Classical and Molecular Cytogenetic Analysis of Cereal Genomes*, W. J. Raupp and B. S. Gill, Eds., pp. 38–48, Kansas Agricultural Experiment Station, Manhattan, NY, USA, 1995.

[453] K. J. Newton, "Plant mitochondrial genomes: organization, expression and variation," *Annual Review of Plant Physiology and Plant Molecular Biology*, vol. 39, pp. 503–532, 1988.

Hindawi Publishing Corporation
International Journal of Plant Genomics
Volume 2008, Article ID 486258, 13 pages
doi:10.1155/2008/486258

Review Article

Barley Genomics: An Overview

Nese Sreenivasulu, Andreas Graner, and Ulrich Wobus

Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK), Corrensstraße 3, 06466 Gatersleben, Germany

Correspondence should be addressed to Nese Sreenivasulu, srinivas@ipk-gatersleben.de

Received 15 November 2007; Accepted 8 February 2008

Recommended by P. Gupta

Barley (*Hordeum vulgare*), first domesticated in the Near East, is a well-studied crop in terms of genetics, genomics, and breeding and qualifies as a model plant for *Triticeae* research. Recent advances made in barley genomics mainly include the following: (i) rapid accumulation of EST sequence data, (ii) growing number of studies on transcriptome, proteome, and metabolome, (iii) new modeling techniques, (iv) availability of genome-wide knockout collections as well as efficient transformation techniques, and (v) the recently started genome sequencing effort. These developments pave the way for a comprehensive functional analysis and understanding of gene expression networks linked to agronomically important traits. Here, we selectively review important technological developments in barley genomics and related fields and discuss the relevance for understanding genotype-phenotype relationships by using approaches such as genetical genomics and association studies. High-throughput genotyping platforms that have recently become available will allow the construction of high-density genetic maps that will further promote marker-assisted selection as well as physical map construction. Systems biology approaches will further enhance our knowledge and largely increase our abilities to design refined breeding strategies on the basis of detailed molecular physiological knowledge.

1. INTRODUCTION

In the 21st century, cereals continue to constitute the most important crops with an annual output of 2 billion tons (according to FAO in 2006; http://www.fao.org). In today's worldwide production, barley ranks fourth among cereals and is preferentially used as feed grain, as a raw material for beer production and, to a smaller extent, as food. Initially, barley was domesticated in the fertile crescent of the Neolithic Near East over 10 000 years ago [1]. In the subsequent millennia, farmers continuously adapted local populations to their needs, leading to a great variety of landraces. About 100 years ago, these formed the basis for the development of modern cultivars by cross breeding. During this time, grain yield was more than doubled with an estimated genetic contribution to this increase of about 30–50% [2]. However, to meet the future challenges imposed by a changing environment, to feed a growing world population, and to provide renewable resources to satisfy the soaring demand for energy, genomics-based technologies have to be efficiently implemented to study the genetic basis of plant performance and to isolate agronomically important genes from the genetic diversity present in the gene pool of barley. A broad spectrum of resources has been developed during the last two decades to facilitate the systematic analysis of the barley genome. These include a large number of mapped molecular markers, comprehensive EST collections, BAC libraries, mutant collections, DNA arrays, and enabling technologies such as the large scale production of doubled haploids and efficient transformation protocols. Advances made in barley genomics and recent efforts made towards physical map construction and sequencing of the barley gene space (http://barleygenome.org) will largely contribute to a comprehensive understanding of gene functions in the context of agronomical important phenotypes (refer to Figure 1 and Table 1). Recently, the techniques and methods employed in cereal genomics have been reviewed [3–8]. In this overview, we have tried to summarize progress in structural and functional genomics of barley and put emphasis on important agronomical aspects such as grain yield, seed quality traits, and implications for malting quality improvement.

Table 1: Barley genomic resources.

Databases	Website URL	Application
Barley Genetic Stocks	http://ace.untamo.net/cgi-bin/ace/searches/basic	Provides information on the morphological & genetic background of barley mutants and contains information on 736 barley translocation and duplication lines.
US Barley Germplasm	http://barleyworld.org/northamericanbarley/germplasm.php	Contains information on barley recombinant chromosome substitution lines and North American barley mapping populations.
EBDB	http://pgrc-35.ipk-gatersleben.de/portal/page/portal/PG_BICGH/P_BICGH/P_BICGH_RESOURCES/P_BICGHI_RESEBDB	The European Barley Database (EBDB) contains passport and evaluation data of 155,000 barley accessions including the international barley core collection.
ICARDA Barley varieties	http://www.icarda.cgiar.org/Crops_Varieties.htm#Barley	Provides an index of barley variety releases from ICARDA, 1977–2005.
Barley TILLING	http://www.scri.ac.uk/research/genetics/BarleyTILLING	A reverse genetics platform, which can be screened for 8,600 barley (cultivar "Optic") EMS mutagenized lines.
CR-EST	http://pgrc.ipk-gatersleben.de/cr-est/index.php	Barley EST database containing sequences, functional annotation and clustering information of more than 232,000 ESTs.
HarvEST	http://harvest.ucr.edu	Barley EST database containing unigene sequences and the oligo design of Barley1 Affymetrix array. It also includes a 1000 barley SNP loci genetic map showing syntenic information with rice.
HvGI	http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?gudb=barley	This *Hordeum vulgare* Gene Index provides functional annotation information, 70-mer oligo predictions and in silico gene expression data for 50,000 unigenes.
NCBI Barley genome view	http://www.ncbi.nlm.nih.gov/mapview map_search.cgi?taxid=4513	Provides an overview about the available genomic and genomic survey sequences (GSS) of barley.
IBSC	http://barleygenome.org	Activities of the International Barley Genome Sequencing Consortium (IBSC) are highlighted.
Barley genome	http://phymap.ucdavis.edu:8080/barley	Barley physical mapping database and available BAC clones together with the accompaning ESTs.
Barley physical map	http://pgrc.ipk-gatersleben.de/kuenzel/barleymap.html	Barley translocation breakpoints integrated into the Igri/Franka-derived RFLP linkage map.
Barley genomics	http://barleygenomics.wsu.edu	Contains information about barley molecular markers, genetic maps, BACs and mutants.
Barley DB	http://ukcrop.net/perl/ace/search/BarleyDB	Contains information about barley germ plasm, molecular markers, genetic maps and BACs.
Gramene	http://www.gramene.org	Provides an overview of comparative maps of cereals including available updated molecular markers and maps of barley.
GrainGenes	http://wheat.pw.usda.gov/GG2/index.shtml	Triticeae database provides an overview about available maps, genetic markers, QTLs and gene expression data.

Table 1: Continued.

Databases	Website URL	Application
Barley dbEST SSRs	http://www.genome.clemson.edu/projects/barley/ssr.dbest.html	15,182 barley simple sequence repeats (SSR) were predicted using the available 328,724 dbEST dataset.
Barley SNP database	http://bioinf.scri.ac.uk/barley_snpdb	Barley SNP linkage map.
Barley RFLP database	http://pgrc.ipk-gatersleben.de/rflp/rflp.html	Contains data of mapped barley RFLP-markers from IPK.
Barley DArT	http://www.triticarte.com/content/barley_diversity_analysis.html	High density consensus map of barley DArT markers linking to existing SSR, RFLP and STS loci.
BarleyBase	http://www.plexdb.org/plex.php?database=Barley	An online dataset for storing and visualizing gene expression data of the Barley 1 GeneChip Affymetrix array.
BDC-GED	http://pgrc.ipk-gatersleben.de/seeds	Contains barley developing caryopses gene expression data.

2. BARLEY ESTS, BACS, AND PHYSICAL MAPS—A SPRINGBOARD FOR THE EXPLORATION OF THE GENOME

The seven barley chromosomes represent the basic genome of all Triticeae species. Still, the large genome (~5500 MB), of which 80% is composed of repetitive DNA is presently not amenable to whole genome sequencing. Therefore, large scale sequencing programs for the development of expressed sequence tags (ESTs) from various cDNA libraries have been initiated. The progress made in the last 5 years resulted in the generation of 437,713 ESTs covering different cDNA libraries from various stages of plant development and tissues challenged with abiotic and biotic stresses (http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html, September 14th 2007 release). Alignment of these ESTs led to the identification of a representative set of 50,453 unigenes with 23,176 tentative consensi and 27,094 singletons (http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?gudb=barley), representing possibly about 75% of all genes in the barley genome. An earlier estimate of the barley gene content based on 110,000 ESTs led to the prediction of around 30 000 unique genes [9]. This number might be an under representation due to the low EST coverage. The same EST data set, which was generated from different tissues covering the plant's life cycle, was analyzed to gain insight into differential gene expression programs in diverse plant tissues by in silico expression studies [9]. In this way, comprehensive analysis of extensive EST resources generated from large genomes provides snap shots of the transcriptome aiding in gene discovery. This also allows identifying coregulated metabolic and regulatory networks [10, 11] and helps to establish high-density molecular maps [12–14] which form the basis for comparative genomic studies, trait mapping, and map-based gene isolation. Thus, in large genome cereal species like barley, EST sequences facilitate a comprehensive overview of gene content and represent a resource to study the evolution and organization of a genome. Regarding the latter, EST-derived information remains limited as it fails to provide, for instance, regulatory information, since promoters and full length sequences are not available.

Physical maps represent an important link to connect the genetic level to the sequence level. Similar to genetic maps, physical maps are available at different levels of resolution. Wheat-barley addition lines are a useful resource to rapidly assign ESTs to an entire chromosome or to a chromosome arm [15]. Using this resource, 1787 genes present on the Barley 1 GeneChip could be assigned to the six different chromosomes of barley (365 genes to 2H, 271 to 3H, 265 to 4H, 323 to5H, 194 to 6H, and 369 to 7H) [16]. At a higher resolution, a physical map of all the seven barley chromosomes has been prepared by mapping DNA markers derived from both genomic as well as gene-based sequences relative to the translocation breakpoints of individual chromosomes that had been isolated using microdissection techniques [17]. The resulting map is of particular value, as it can be directly aligned to the genetic map of barley by common markers and thus allows for the estimation of the ratio between genetic and physical distances. An alternative approach has been described by Masoudi-Nejad et al. [18]. Here the presence of a wheat gametocidal chromosome in a wheat barley addition line was exploited to select 90 progeny lines that carried differently sized fragments of barley chromosome 7H. These were subsequently used to determine the physical order and distance of markers located on barley chromosome 7H.

During the past several years, core public resources have been established by generating "bacterial artificial chromosome" (BAC) libraries from different barley cultivars: "Morex" ([19]; 313,344 clones), "Cebada Capa" ([20]; 177,000 clones) and "Haruno Nijo" (http://www.intl-pag.org/10/abstracts/PAGX_P393.html). Based on fluorescence in situ hybridization (FISH) techniques karyotype landmarks were derived for barley, which could be used in future to place the BAC clones onto the physical map [21]. This map shows that the genetic linkage maps are well covered with markers among all chromosomes. At the same time, the physical maps reveal large areas of the barley genome that have yet to be mapped. These unmapped areas mainly

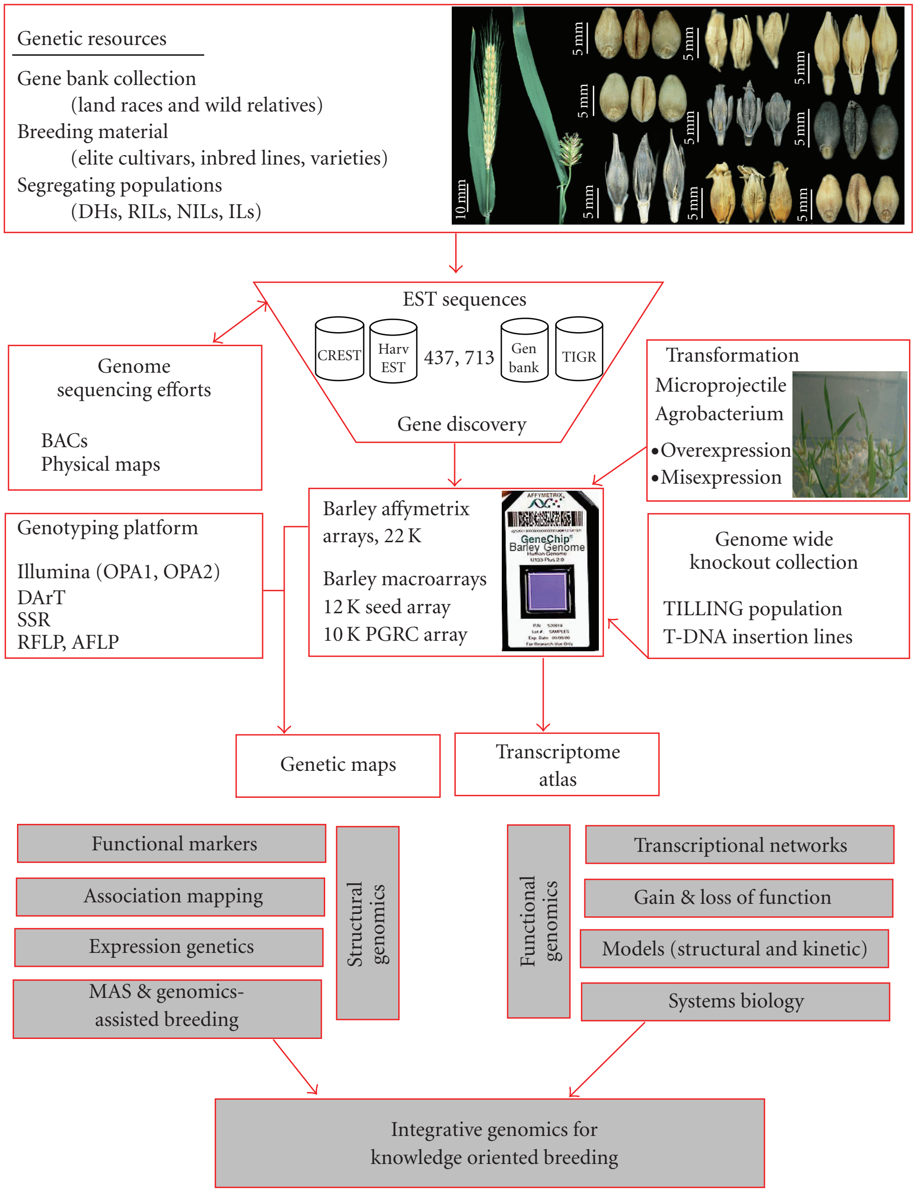


Figure 1: Schematic representation of structural and functional genomic aspects of barley.

consist of heterochromatin and show very low recombination rates [17]. In accordance with these findings, there is increasing evidence that genes are not randomly distributed across the barley genome butconfined to a gene space, which mainly covers the distal parts of the chromosomes. Experimental evidence for the existence of a gene space has been gained from screening a barley BAC library with EST-derived probes, which showed a significant nonrandom distribution across the BAC clones [22]. More direct evidence has been reached on the sequence level for barley and other Triticeae species. Although up to now only a limited amount of sequence data is available, the average density of annotated genes is much higher than that expected for a random distribution across the genome. The disproportionate gene number found is probably due to the preferential selection of gene containing BACs for sequence analysis. Within single BACs, there is considerable variation ranging, in case of barley, from 1 gene in 12 kb up to 1 gene in 220 kb (for review see [23]). Thus even the gene space itself seems to be characterized by a highly variable distribution of genes against the backdrop of noncoding, mainly repetitive DNA.

The existence of a gene space also opens up new opportunities to focus analyses on gene-rich regions only. Recently, international efforts have been gearing up to utilize the extensive barley EST resources for BAC anchoring and genetic mapping. An elegant approach of screening of the Morex BAC library using EST-derived, pooled “overgo” probes [24] resulted in the identification of gene containing BACs. Upon fingerprinting of a subset of 21 161 clones, 2262 contigs could be assembled covering approximately 9.4% of the barley genome. Furthermore, a database has been set up to search screening results of BAC libraries as well as to provide an integrative view of data from the existing barley genetic and physical maps (http://www.genome.clemson.edu). The identified BAC-based gene-rich regions of the genome have been selected as a genomic reference from cultivar Morex to initiate sequencing of all gene-containing regions of the barley genome by an international effort coordinated through the International Barley Sequencing Consortium (IBSC, http://barleygenome.org).

3. A BARLEY TRANSCRIPTOME ATLAS

Despite the lack of a barley genome sequence, functional genomics efforts have been initiated by taking advantage of the available EST sequence information generated by multinational coordinated efforts (see above). As a first step, efforts were made to derive functional assignments of the available barley unigene set by annotation transfer from homologous sequences relying on the available plant whole genome sequences and by identifying common motifs from Interpro. As a result, several ontology structures such as MIPS [9] and MAPMAN functional categories (N. Sreenivasulu, unpublished data; http://mapman.mpimp-golm.mpg.de/index.shtml) were developed. Such computational methods also yielded putative regulatory networks as well as metabolic pathway interaction networks, but still about half of the genes have to be classified as "unknown."

The available barley EST unigene resources played a profound role in developing several platforms for transcriptome analysis including cDNA-based macroarrays [11, 25], microarrays [26], and oligonucleotide-based affymetrix arrays [10, 27]. Other profiling techniques used in barley include cDNA-AFLP [28], SAGE (Serial Analysis of Gene Expression) [29, 30], and iGentifier. The latter method combines elements of tag sequencing such as SAGE and fragment display [31]. By successfully applying these techniques, barley transcriptome data have been collected from grain development [11, 25, 32], grain germination [33, 34], at least 15 different tissues/organs covering different growth stages [10], and abiotic [26, 35–37] as well as biotic stress responses [38, 39]. The new insights gained from transcriptome analysis of host-pathogen studies have lately been reviewed by Wise et al. [40]. These large scale gene expression data sets serve as baseline experiments to generate a barley transcriptome atlas. Also, an online Plant Expression Database (PLEXdb), previously known as BarleyBase (http://www.plexdb.org/plex.php?database=Barley) has been created to store, visualize, and statistically analyze Barley 1 GeneChip data [41].

While transcriptomics have brought about substantial progress in elucidating biochemical pathways of barley seed metabolism (see reviews [5, 42]), very recent findings shed light on the interplay of many cellular and metabolic events that are coordinated by a complex regulatory network during barley seed development [10, 11, 25]. Studying expression data of nearly 12 000 seed-expressed genes revealed, for instance, the participation of tissue-specific signaling networks controlling ABA-mediated starch accumulation (via SNF1 kinase and a set of transcription factors) in the endosperm and participation of ABA-responsive genes in establishing embryo desiccation tolerance [11]. CpG methylation found in the promoters of prolamin box-binding factor and B-hordein genes suppresses transcript levels during the prestorage until the intermediate phase of grain development. This process coincides with the coexpression of methyltransferases, core histones and DNA-unwinding ATPases [43]. Thus storage protein gene expression may be regulated by CpG methylation. Using a *lys 3a* mutant, it has been shown that demethylation of the B-hordein promoter does not occur in the mutant, hence transcripts encoding storage proteins such as B-hordeins and C-hordeins are almost absent in the developing endosperm of this mutant [44]. Transcriptome profiling of barley embryos using the 22K affymetrix Barley 1 GeneChip revealed activation of developmentally distinct defense related gene sets including coregulated phenylpropanoid and phytoalexin related genes around 20 days after flowering (DAF), followed by upregulation of antioxidant and pathogen related gene sets around 37 DAF [45]. The knowledge obtained on metabolic processes of seed quality traits could eventually be used to develop superior varieties by genetic engineering or by marker-assisted selection in conventional breeding programs.

Transcriptome analysis has also been carried out during barley grain germination at tissue-specific levels [10, 46]. Using cDNA array technology gene expression was analyzed in germinating seed samples, collected from ten different barley genotypes showing differential malting response [46]. Based on six different malting quality parameters related to hydrolytic events connected to protein, starch, and cell wall degradation 19 candidate genes were identified, whose transcript abundance showed a significant correlation with some of the malting quality parameters. White et al. [30] analyzed seven different SAGE libraries derived from malted grains and identified 100 most abundant transcripts showing differential responses during eight different time points during malting. These transcripts are related to stress and defense response, hydrolytic processes and translational events. The list of candidate genes identified in the two studies [30, 46] was further validated by a genetical genomics approach in which gene expression studies were conducted with populations segregating for malting traits [34, 47].

4. FUNCTIONAL GENOMICS APPROACHES IN BARLEY

A major aim of functional genomic studies is to understand the metabolic and regulatory networks within the structural and functional context of cells, tissues, and organs often changing with time. Hence in this review, we update the functional genomic resources available (Table 1) to study gene functions in barley using reverse genetics approaches and highlight the initial success achieved through genetic engineering based on the manipulation of individual genes.

4.1. Reverse genetics

To determine gene-function relationships, large scale genome-wide reverse genetics approaches have been developed in barley (see [48] for review) which includes both nontransgenic technology platforms such as TILLING (targeting induced local lesions in genomes) [49] and insertional mutagenesis systems based on transgenic technology [50–54]. Thus, the Scottish Crop Research Institute generated a large M_2 TILLING population in the barley cultivar "Optic" with leaf material and seeds from 20 000 plants freeze dried and archived [49]. EMS induced mutations were scored at various growth stages under different conditions and documented [49, 55]. Mutant phenotypes, candidate genes, and observed DNA sequence variations can be queried in an SCRI mutant database (http://germinate.scri.ac.uk/barley/mutants/index.php?option=com_wrapper&Itemid=35). In a more recent attempt, IPK developed a TILLING population of 10 000 M2 plants in the cultivar 'Barke" (N. Stein, personnel communication). Similarly, a collection of 5000 M3 mutants of the cultivar "Morex" is provided by the University of Bologna (http://www.intl-pag.org/13/abstracts/PAG13_P081.html).

To aid functional gene analysis, insertional mutagenesis approaches were followed in barley during the last decade (i) to create loss-of-function mutations by the insertion of transposable elements into a gene of interest [50–53] and (ii) use activation tagging (the random genomic insertion of either promoter or enhancer sequences) to generate dominant gain-of-function mutations [54, 56]. Insertion lines have been generated by creating transgenic plants carrying *Ac* and *Ds* elements, and crossed them to induce *Ds* transposition [50–52]. *Ds* elements were preferentially found in genic regions and exhibited a high-remobilization frequency [52, 53]. Such *Ds* launch pads, represented by barley lines with each harboring a single copy *DS* insertion at a well-defined position in the genome, will be valuable for future targeted gene tagging. Similarly, dominant overexpression phenotypes [54, 56] will help to study gene functions in the large barley genome where loss-of-function mutations often may not cause phenotypes because of gene redundancy.

4.2. Transgenic barley and its potential applications

In order to functionally characterize candidate genes identified in functional genomic studies, it was mandatory to establish a stable and efficient genetic transformation technique in barley. In contrast to the biolistic gene transfer technique [57], a more efficient *Agrobacterium* mediated barley genetic transformation method based on immature embryos was developed in spring barley [58]. In a recent attempt to further improve this technology, Kumlehn et al. [59] developed a transformation method for winter barley based upon the infection with *Agrobacterium* of androgenic pollen cultures. By this approach, homozygous double haploid plants could be immediately obtained at high frequency through chromosome doubling.

During the last decade, systematic efforts were made for genetic engineering of barley to improve seed quality traits including those related to malting (reviewed in [60]). Malting improvement has been addressed by altering the expression of hydrolytic enzymes related to the degradation of storage products such as starch (α and ß-amylases, [61, 62]) and cell wall components. In another approach, several enzymes such as xylanase, glucanase, endo-, and exoprotease were over expressed in transgenic barley grains and preferably the enzyme mix necessary for malting process are provided by transgenic seeds [63].

Protein engineering has been used to produce thermostable 1, 3; 1, 4ß-glucanases in transgenic barley grains [64–66]. Such grains can be used to enhance the feed quality of barley for poultry [67, 68]. In a similar approach, a hybrid cellulase gene driven by the endosperm specific rice GluB-1 promoter was expressed and produced the enzyme up to 1.5% of total grain protein [69]. In addition, functions of key genes involved in determining seed quality traits related to storage product accumulation were tested. For instance, antisense downregulation of limit dextrinase inhibitor showed reduced amylose over amylopectin levels and eventually reduced total starch [70]. Also overexpression of wheat thioredoxin *h* in the endosperm of transgenic barley grain leads to increased activity of the starch debranching enzyme limit dextrinase [71, 72]. Further, a powerful approach of antisense oligodeoxynucleotide inhibition has been used to reveal sugar signaling networks. Short stretches of 12–25 nucleotide long single-strand sequences have been delivered to barley leaf cells to block the effect of SUSIBA2, a key transcriptional activator involved in plant sugar signaling [73]. Recently, this approach has been successfully implemented to deliver antisense oligodeoxynucleotides to barley seed endosperm to suppress sugar related signaling genes [74]. HvGAMYB, a transcription factor initially identified in aleurone and shown to be upregulated by gibberellin, has been shown to be expressed also in barley anthers. The overexpressing HvGAMYB transgenic lines show reduced anther size with a male sterility phenotype [75]. Our laboratory has recently characterized a new protein called Jekyll, which is preferentially expressed in barley grain nucellar projection tissue [76]. Its downregulation decelerates autolysis of nurse tissue. As a result, proliferation of endosperm nuclei is impaired and less starch is finally accumulated in the endosperm [77].

4.3. Towards systems biology

With respect to applied aspects in crop plants, a comprehensive knowledge of cellular and functional complexity as related to key agronomic traits could be revealed using a systems biology approach. With this in mind, a number of tools and databases were developed at our institute (Leibniz Institute of Plant Genetics and Crop Plant Research/IPK) to store, analyze, and display the data derived from multiparallel-OMICs profiling studies at transcript, metabolite, and protein/enzyme level with the aim to eventually gain insight into the organization of function-related networks in barley [78, 79]. These include CR-EST [78] (it provides access to clustering and annotation data of IPK EST projects), Meta-All [79], and MetaCrop [80] (they allow to access curated metabolic pathway information and kinetic reactions of crop plants), VANTED [81] (for visualization and analysis of metabolic and regulatory networks), HiT-MDS [32] (for screening of coexpressed genes and validation of cluster centroids) as well as barley MapMan and PageMan [http://mapman.mpimp-golm.mpg.de; to index and visualize overrepresented functional categories and detailed metabolic pathway charts from throughput transcriptome data]. With the focus of using the "developing seed" as model for systems biology studies, we investigated transcriptional and metabolic networks during grain development [11, 25, 82], developed 3D models of the developing barley grain [83], implemented magnetic resonance-based techniques to establish 4D models as a framework to store different sets of data in their spatiotemporal context [84], visualized the spatial distribution of specific biochemical compounds by noninvasive NMR-based imaging methods [85] and established kinetic models of primary metabolism ([86] and E. Grafahrend-Belau and B. Junker, unpublished data) as already worked out for potato [87]. In addition, a proteomic platform has been successfully established to study barley grain development [88, 89]. The emerging model (largely qualitative) explaining how the barley grain develops and functions has to be further validated especially by the creation and analysis of different lines of transgenic plants with perturbations at putative key metabolic and/or regulatory sites (see Figure 1).

5. FUNCTIONAL MOLECULAR MARKERS AND THEIR POTENTIAL APPLICATIONS IN THE AREA OF APPLIED GENOMICS

5.1. Marker development and marker-assisted selection (MAS)

Almost two decades ago, RFLP markers were employed to develop the first comprehensive molecular marker maps in barley [90–92]. Using those RFLP maps, a series of agronomic traits and characters including many quality traits and resistance against several diseases have been mapped (for review see [93, 94]). Later, the availability of large numbers of ESTs facilitated the systematic development of functional markers, for example, by extracting ESTs containing simple sequence repeat (SSR) motifs using appropriate software tools [95]. Although EST-based SSR markers have been shown to be less polymorphic than their genomic counterparts, this drawback is more than compensated for by the ease of their development. Also, the availability of ESTs from multiple-genotypes/cultivars of barley provides the possibility to identify sequence polymorphisms (mainly single-nucleotide polymorphisms and small InDels) in the corresponding EST alignments. These in turn can be exploited for the development of markers [96, 97]. Kota et al. [98] developed the computer algorithm SNiPping for discovery of functional markers through browsing EST assemblies in barley. Also an SNP2CAPS program has been published to facilitate the computational conversion of SNP markers into CAPS markers [99]. Information generated from the diverse mapping projects was further enhanced by the development of consensus maps [14, 100–102]. These provide integrative genetic information by featuring high marker densities. Although the gel-based genotyping platforms offer the best quality marker systems, their low throughput encouraged researchers to explore high-throughput technologies that can simultaneously assay thousands of markers based on single nucleotide polymorphisms (SNP). Most recently, genome-wide scans using SNP-based genotyping platforms such as Illumina GoldenGate BeadArrays [103] and the diversity arrays technology (DArT), which do not require any sequence information [104] have been successfully established in barley. Although DArTs are not systematically interrogating expressed sequences, the choice of appropriate enzymes facilitates their enriched representation. Based on DArT technology, a high-density consensus map has recently been established [105]. A number of recent studies also reported the use of the affymetrix Barley 1 GeneChip [27] for identifying single-feature polymorphisms (SFPs), which cover not only SNPs but also indels and polymorphisms generated due to alternative splicing and polyadenylation [34, 106].

An important application of the above discussed functional markers is marker-assisted selection (MAS). MAS is based on linking the DNA polymorphisms revealed by marker analysis with agronomical traits allowing for their rapid selection in routine breeding programs. MAS can be performed already at juvenile growth stages and before flowering, and thus provides breeders with the opportunity to implement faster back-crossing strategies and allele enrichment in complex crosses, which eventually reduces the time and costs required for the development of improved varieties. Despite its inherent advantages, the application of MAS in barley up to now has mainly been restricted to monogenic traits such as disease resistances. Here, one of the most widespread examples is the marker assisted selection of the *rym4* gene giving resistance to the barley yellow mosaic virus complex. For this gene, several closely linked and easily scorable markers have been developed [107, 108]. More recently, cloning of the gene facilitated the exploitation of functional polymorphisms within the coding region of the resistance gene to differentiate between alleles [109]. Using MAS, several genes providing full resistance could be readily combined in complex crosses without time consuming progeny tests in the greenhouse or in the field (e.g., [110, 111]).

MAS for quantitative traits suffers from two major limitations. (i) Compared to monogenic traits, quantitative traits are characterized by lower heritabilities impairing their accurate scoring and entailing a less accurately defined genetic position of the corresponding quantitative trait locus (QTL). As a result, large chromosomal fragment needs to be selected for, resulting in the meiotic transfer of many potentially undesired genes. Meiotic purification of a QTL into a "mendelian" locus, showing monogenic inheritance, provides a solution to this problem. The feasibility of down-tracking a QTL to a single gene has been initially demonstrated in tomato and requires the stepwise size reduction of a QTL fragment and its conversion into a near isogenic line by repeated backcrossing (for review see [112]). In barley, this approach has been successfully employed to isolate the bot1 gene underlying a major QTL conferring boron tolerance

[113]. (ii) Many of QTL alleles escape detection, when transferred into a different genetic background. The reasons for the "disappearance of QTLs" include epistatic interactions, QTL x environment effects, the allelic states of the parental lines or the small contribution of a single QTL to the overall variance. As a result, only few common QTLs were detected, when the results of mapping studies that were performed in different crosses were compared [114].

Although the number of successful examples for applying MAS in barley breeding is still rather limited (see reviews by [114, 115], the recent implementation of high-throughput genotyping platforms (Illumina, DArT, and SFP identification by using Barley 1 GeneChip affymetrix array) in barley will significantly increase the identification of marker trait associations, and the subsequent identification of potential candidate genes. Finally, this will allow to treat QTLs as monogenic traits and thus spur their marker assisted manipulation in breeding programs. In combination with a wide range of mapping populations developed for specific agronomic traits, this comprehensive resource of markers now allows the identification of polymorphisms in functionally defined sequences [12, 34, 105, 106]. Functional markers will also be useful for (i) association studies based on linkage disequilibrium, (ii) detection of *cis* and *trans*-acting regulators either based on genetical genomics studies using well-defined mapping populations or by investigating allelic imbalance [116], (iii) identification of alleles influencing agronomically important traits using TILLING/EcoTilling approaches (EcoTilling is a means to determine the extent of natural variation in selected genes), and (iv) genomics-assisted breeding (see Figure 1).

5.2. Linkage disequilibrium-based association studies

Linkage disequilibrium is the nonrandom distribution of alleles in a sample population and forms the basis for the construction of genetic maps and the localization of genetic loci for a variety of traits. The principles leading to LD apply to both biparental mapping populations (F2, RILs, etc.) and natural populations. Therefore, LD mapping is the method of choice for genetic analysis in organisms like humans and animals, where experimental populations are either not available or difficult to establish [117].

Because of its inherent advantages, LD mapping approaches are increasingly being applied for plant species, in particular maize. Due to the outbreeding character of this species, LD extends only over a few kb and thus leads to a high-genetic resolution, up to the level of individual candidate genes that can be associated with a given trait (see recent reviews [118, 119]). The use of association genetic analyses in inbreeding species such as barley has been limited so far. However, recent studies have shown that LD extends over much longer genetic distances in barley than in maize. A European germplasm collection of 146 two-rowed spring barley cultivars was used to carry out LD mapping of yield traits using 236 AFLP markers [120]. Associated markers were identified that are located in similar regions where QTLs for yield had been found in barley [93, 121, 122]. A systematic survey of 953 gene bank accessions representing a broad spectrum of the genetic diversity in barley genetic resources revealed that LD extends up to 50 cM but is highly dependent on population structure [120, 123]. On the one hand, the high level of LD in barley is due to the inbreeding mating type of this species; on the other hand, the selection of germplasm plays an important role. Analysis of a germplasm collection of European cultivars, land races, and wild barley accession from the Fertile Crescent region provided hints that the level of LD decreases from cultivars to landraces to wild barley [124]. Similarly, Morrell et al. [125] reported low levels of LD in wild barley by examining LD within and between 18 genes from 25 accessions. Local differences in LD have been observed at the grain hardness locus comprising four closely linked genes (*hinb*, *hina*, GSP, PG2). Here, a high level of LD was observed in the intergenic region between *hinb-1* and *hina* probably due to transposable elements present in this region, which influence the local recombination rate [114]. By assaying 1524 genome-wide SNPs in elite northwest European barley using the Illumina GoldenGate BeadArray platform Rostoks et al. [103] concluded that whole-genome association scans can be exploited for trait mapping in barley. This was further exemplified by the identification of a marker that showed an association with the winter habit and which could be tracked to a cluster of CBF (C-repeat/DRE-binding factor) gene homologs. In a recent whole genome LD-mapping approach, Steffenson et al. [126] used 318 wild barley accessions to perform association mapping studies using DArT markers to identify rust resistance genes. In addition, LD analysis has been performed based on haplotypes derived from 131 accessions by covering 83 SNPs within 132 kb around the gene HveIF4E, which confers resistance to barley yellow mosaic virus. The authors identified three haplogroups discriminating between the alleles *rym4* and *rym5* [127]. Taken together, the above mentioned association studies provide starting points for a more systematic analysis of agronomic traits. These may be selected from the vast ex situ gene bank collections available for this crop. Alone at the IPK gene bank some 20 000 different barley accessions represent an ample cross section of the genetic diversity present in this species. However, in order to fully exploit the potential of LD-based association analysis in this species, populations have to be carefully selected to minimize the confounding effects of population structure. This is particularly evident in modern barley germplasm, which is frequently structured into spring and winter as well as 2-rowed and 6-rowed types, forming distinct subpopulations (e.g., [95]). If these effects are not adequately accounted for during association analysis, the risk of detecting spurious associations increases.

5.3. Genetical genomics studies

The genetical genomics strategy was first outlined by Jansen and Nap [128]. It combines gene expression studies with genetic linkage analysis. Differentially expressed genes (but also proteins and metabolites) involved in metabolic and regulatory pathways and identified by high-throughput technologies are treated as phenotypes, and genetic variants that influence gene expression are identified in genetically related lines. This strategy has been successfully applied also in plant systems, and relevant data were reviewed elsewhere [128–130]. Here, we will focus on the latest development in expression QTL (eQTL) mapping in barley. Using the Barley 1 GeneChip affymetrix array SFP genotyping has been performed in 35 recombinant lines of a Steptoe × Morex doubled-haploid population, enabling eQTL studies [34]. Using a high-throughput SFP genotyping platform, genome-wide linkage analysis has been performed based on 22 000 transcript data collected from 139 DH lines (Steptoe × Morex). The most significant eQTLs derived from germinating barley grain are linked to *cis* regulation [47]. Using the same mapping population, a serine carboxypeptidase 1 eQTL has been mapped on chromosome 3H to the same region where a QTL for the malting quality trait "diastatic power" has been mapped [131]. In another study, instead of a segregating population a set of 47 BC3 DH introgression lines was employed (wild barley [*H. spontaneum*] is introgressed in the genetic background of the elite line "Brenda" [*H. vulgare*]) in order to understand gene expression networks controlling seed traits. Initially, this BC3 DH population was used to identify QTLs for yield and yield components [132]. In further experiments, expression data from nearly 12 000 genes interrogated by using a barley seed specific array were used to calculate eQTLs (C. Pietsch et al., unpublished). Although such initial studies provide evidence that genetical genomics is a promising concept which assists to expose gene-trait relationships, an extensive exploration of the technology needs the full barley genome sequence and improved high-throughput genotyping information.

6. OUTLOOOK

In recent years, we experienced a dramatic development of new tools and technologies for genome research and a concomitantly dramatic increase in data leading to a much improved and advanced knowledge base. Barley research gained a lot of momentum from this development but the nonavailability of a whole genome sequence is still a serious limitation. However, due to consortial efforts (see above) and the rapidly developing sequencing technologies that are relevant for even complex genomes like that of barley [133] this limitation will be largely overcome, hopefully within the next five years. High-throughput transcriptome analysis techniques have already provided numerous new insights in transcriptional networks. They will, together with rapidly improving protein and metabolite profiling techniques and in combination with new genetic analysis concepts such as genetical genomics and association genetics, improve our knowledge on the relationship between the genetic and the phenotypic architecture of agronomic traits and thus create a basis for knowledge-based molecular breeding [134]. As a next step systems biology approaches are emerging, which attempt to model complex cellular or organismic functions in response to changing internal and external factors [135]. Until now molecular markers have had limited success in barley breeding programs, but due to recent advancement of barley genomics a stronger impact on breeding strategies is expected. For instance, marker technologies together with double haploid production have almost halved the time of variety development in Australian wheat and barley breeding programs [136]. However, new whole-genome breeding strategies have to be developed to make full use of the ever increasing knowledge about crop plant genomes and their behavior.

ACKNOWLEDGMENTS

The authors acknowledge the support of the Federal Ministry of Education and Research (BMBF) for supporting barley research activities at IPK within the GABI funding program. The authors apologize for being unable to cite all relevant papers.

REFERENCES

[1] F. Salamini, H. Özkan, A. Brandolini, R. Schäfer-Pregl, and W. Martin, "Genetics and geography of wild cereal domestication in the near east," *Nature Reviews Genetics*, vol. 3, no. 6, pp. 429–441, 2002.

[2] W. H. Schuster, "Welchen Beitrag leistet die Pflanzenzüchtung zurLeistungssteigerung von Kulturpflanzenarten," *Pflanzenbauwissenschaften*, vol. 1, pp. 9–18, 1997.

[3] A. Rafalski, "Applications of single nucleotide polymorphisms in crop genetics," *Current Opinion in Plant Biology*, vol. 5, no. 2, pp. 94–100, 2002.

[4] R. K. Varshney, D. A. Hoisington, and A. K. Tyagi, "Advances in cereal genomics and applications in crop breeding," *Trends in Biotechnology*, vol. 24, no. 11, pp. 490–499, 2006.

[5] U. Wobus and N. Sreenivasulu, "Genomics approaches for the improvement of cereals," in *European Training and Networking Activity, Plant Genomics and Bioinformatics Expression Micro Arrays and Beyond—A Course Book*, J. Freitag, Ed., pp. 146–155, National Institute of Biology, Ljubljana Slovenia, 2006.

[6] M. Bagge, X. Xia, and T. Lübberstedt, "Functional markers in wheat," *Current Opinion in Plant Biology*, vol. 10, no. 2, pp. 211–216, 2007.

[7] A. S. Milligan, S. Lopato, and P. Langridge, "Functional genomics of seed development in cereals," in *Cereal Genomics*, K. P. Gupta and R. K. Varshney, Eds., pp. 447–481, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2004.

[8] R. K. Varshney, P. Langridge, and A. Graner, "Application of genomics to molecular breeding of wheat and barley," *Advances of Genetics*, vol. 58, pp. 121–155, 2007.

[9] H. Zhang, N. Sreenivasulu, W. Weschke, et al., "Large-scale analysis of the barley transcriptome based on expressed sequence tags," *The Plant Journal*, vol. 40, no. 2, pp. 276–290, 2004.

[10] A. Druka, G. Muehlbauer, I. Druka, et al., "An atlas of gene expression from seed to seed through barley development," *Functional & Integrative Genomics*, vol. 6, no. 3, pp. 202–211, 2006.

[11] N. Sreenivasulu, V. Radchuk, M. Strickert, O. Miersch, W. Weschke, and U. Wobus, "Gene expression patterns reveal tissue-specific signaling networks controlling programmed cell death and ABA- regulated maturation in developing barley seeds," *The Plant Journal*, vol. 47, no. 2, pp. 310–327, 2006.

[12] N. Rostoks, S. Mudie, L. Cardle, et al., "Genome-wide SNP discovery and linkage analysis in barley based on genes responsive to abiotic stress," *Molecular Genetics and Genomics*, vol. 274, no. 5, pp. 515–527, 2005.

[13] P. R. Hearnden, P. J. Eckermann, G. L. McMichael, M. J. Hayden, J. K. Eglinton, and K. J. Chalmers, "A genetic map of 1,000 SSR and DArT markers in a wide barley cross," *Theoretical and Applied Genetics*, vol. 115, no. 3, pp. 383–391, 2007.

[14] N. Stein, M. Prasad, U. Scholz, et al., "A 1,000-loci transcript map of the barley genome: new anchoring points for integrative grass genomics," *Theoretical and Applied Genetics*, vol. 114, no. 5, pp. 823–839, 2007.

[15] A. K. M. R. Islam, K. W. Shepherd, and D. H. B. Sparrow, "Isolation and characterization of euplasmic wheat-barley chromosome addition lines," *Heredity*, vol. 46, pp. 161–174, 1981.

[16] S. Cho, D. F. Garvin, and G. J. Muehlbauer, "Transcriptome analysis and physical mapping of barley genes in wheat-barley chromosome addition lines," *Genetics*, vol. 172, no. 2, pp. 1277–1285, 2006.

[17] G. Künzel, L. Korzun, and A. Meister, "Cytologically integrated physical restriction fragment length polymorphism maps for the barley genome based on translocation breakpoints," *Genetics*, vol. 154, no. 1, pp. 397–412, 2000.

[18] A. Masoudi-Nejad, S. Nasuda, M.-T. Bihoreau, R. Waugh, and T. R. Endo, "An alternative to radiation hybrid mapping for large-scale genome analysis in barley," *Molecular Genetics and Genomics*, vol. 274, no. 6, pp. 589–594, 2005.

[19] Y. Yu, J. P. Tomkins, R. Waugh, et al., "A bacterial artificial chromosome library for barley (*Hordeum vulgare* L.) and the identification of clones containing putative resistance genes," *Theoretical and Applied Genetics*, vol. 101, no. 7, pp. 1093–1099, 2000.

[20] E. Isidore, B. Scherrer, A. Bellec, et al., "Direct targeting and rapid isolation of BAC clones spanning a defined chromosome region," *Functional & Integrative Genomics*, vol. 5, no. 2, pp. 97–103, 2005.

[21] J. L. Stephens, S. E. Brown, N. L. V. Lapitan, and D. L. Knudson, "Physical mapping of barley genes using an ultrasensitive fluorescence in situ hybridization technique," *Genome*, vol. 47, no. 1, pp. 179–189, 2004.

[22] R. K. Varshney, I. Grosse, U. Hähnel, et al., "Genetic mapping and BAC assignment of EST-derived SSR markers shows non-uniform distribution of genes in the barley genome," *Theoretical and Applied Genetics*, vol. 113, no. 2, pp. 239–250, 2006.

[23] N. Stein, "Triticeae genomics: advances in sequence analysis of large genome cereal crops," *Chromosome Research*, vol. 15, no. 1, pp. 21–31, 2007.

[24] K. Madishetty, P. Condamine, J. T. Svensson, E. Rodriguez, and T. J. Close, "An improved method to identify BAC clones using pooled overgos," *Nucleic Acids Research*, vol. 35, no. 1, p. e5, 2007.

[25] N. Sreenivasulu, L. Altschmied, V. Radchuk, S. Gubatz, U. Wobus, and W. Weschke, "Transcript profiles and deduced changes of metabolic pathways in maternal and filial tissues of developing barley grains," *The Plant Journal*, vol. 37, no. 4, pp. 539–553, 2004.

[26] Z. N. Ozturk, V. Talamé, M. Deyholos, et al., "Monitoring large-scale changes in transcript abundance in drought- and salt-stressed barley," *Plant Molecular Biology*, vol. 48, no. 5-6, pp. 551–573, 2002.

[27] T. J. Close, S. I. Wanamaker, R. A. Caldo, et al., "A new resource for cereal genomics: 22K barley genechip comes of age," *Plant Physiology*, vol. 134, no. 3, pp. 960–968, 2004.

[28] J. Leymarie, E. Bruneaux, S. Gibot-Leclerc, and F. Corbineau, "Identification of transcripts potentially involved in barley seed germination and dormancy using cDNA-AFLP," *Journal of Experimental Botany*, vol. 58, no. 3, pp. 425–437, 2007.

[29] A. F. M. Ibrahim, P. E. Hedley, L. Cardle, et al., "A comparative analysis of transcript abundance using SAGE and Affymetrix arrays," *Functional & Integrative Genomics*, vol. 5, no. 3, pp. 163–174, 2005.

[30] J. White, T. Pacey-Miller, A. Crawford, et al., "Abundant transcripts of malting barley identified by serial analysis of gene expression (SAGE)," *Plant Biotechnology Journal*, vol. 4, no. 3, pp. 289–301, 2006.

[31] A. Fischer, A. Lenhard, H. Tronecker, et al., " *iGentifier*: indexing and large-scale profiling of unknown transcriptomes," *Nucleic Acids Research*, vol. 35, no. 14, pp. 4640–4648, 2007.

[32] M. Strickert, N. Sreenivasulu, B. Usadel, and U. Seiffert, "Correlation-maximizing surrogate gene space for visual mining of gene expression patterns in developing barley endosperm tissue," *BMC Bioinformatics*, vol. 8, article 165, pp. 1–11, 2007.

[33] E. Potokina, N. Sreenivasulu, L. Altschmied, W. Michalek, and A. Graner, "Differential gene expression during seed germination in barley (*Hordeum vulgare* L.)," *Functional & Integrative Genomics*, vol. 2, no. 1-2, pp. 28–39, 2002.

[34] Z. W. Luo, E. Potokina, A. Druka, R. Wise, R. Waugh, and M. J. Kearsey, "SFP genotyping from affymetrix arrays is robust but largely detects *cis*-acting expression regulators," *Genetics*, vol. 176, no. 2, pp. 789–800, 2007.

[35] A. Ueda, A. Kathiresan, J. Bennett, and T. Takabe, "Comparative transcriptome analyses of barley and rice under salt stress," *Theoretical and Applied Genetics*, vol. 112, no. 7, pp. 1286–1294, 2006.

[36] V. Talamé, N. Z. Ozturk, H. J. Bohnert, and R. Tuberosa, "Barley transcript profiles under dehydration shock and drought stress treatments: a comparative analysis," *Journal of Experimental Botany*, vol. 58, no. 2, pp. 229–240, 2007.

[37] H. Walia, C. Wilson, P. Condamine, X. Liu, A. M. Ismail, and T. J. Close, "Large-scale expression profiling and physiological characterization of jasmonic acid-mediated adaptation of barley to salinity stress," *Plant, Cell & Environment*, vol. 30, no. 4, pp. 410–421, 2007.

[38] R. A. Caldo, D. Nettleton, and R. P. Wise, "Interaction-dependent gene expression in *Mla*-specified response to barley powdery mildew," *Plant Cell*, vol. 16, no. 9, pp. 2514–2528, 2004.

[39] T. Gjetting, P. H. Hagedorn, P. Schweizer, H. Thordal-Christensen, T. L. W. Carver, and M. F. Lyngkjær, "Single-cell transcript profiling of barley attacked by the powdery mildew fungus," *Molecular Plant-Microbe Interactions*, vol. 20, no. 3, pp. 235–246, 2007.

[40] R. P. Wise, M. J. Moscou, A. J. Bogdanove, and S. A. Whitham, "Transcript profiling in host-pathogen interactions," *Annual Review of Phytopathology*, vol. 45, pp. 329–369, 2007.

[41] L. Shen, J. Gong, R. A. Caldo, et al., "BarleyBase—an expression profiling database for plant genomics," *Nucleic Acids Research*, vol. 33, pp. D614–D618, 2005.

[42] U. Wobus, N. Sreenivasulu, L. Borisjuk, et al., "Molecular physiology and genomics of developing barley grains," in *Recent Research Developments in Plant Molecular Biology*, vol. 2, pp. 1–29, Research Signpost, Kerala, India, 2005.

[43] V. V. Radchuk, N. Sreenivasulu, R. I. Radchuk, U. Wobus, and W. Weschke, "The methylation cycle and its possible functions in barley endosperm development," *Plant Molecular Biology*, vol. 59, no. 2, pp. 289–307, 2005.

[44] M. B. Sørensen, "Methylation of B-hordein genes in barley endosperm is inversely correlated with gene activity and affected by the regulatory gene *Lys3*," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 89, no. 9, pp. 4119–4123, 1992.

[45] M. E. Nielsen, F. Lok, and H. B. Nielsen, "Distinct developmental defense activations in barley embryos identified by transcriptome profiling," *Plant Molecular Biology*, vol. 61, no. 4-5, pp. 589–601, 2006.

[46] E. Potokina, M. Caspers, M. Prasad, et al., "Functional association between malting quality trait components and cDNA array based expression patterns in barley (*Hordeum vulgare* L.)," *Molecular Breeding*, vol. 14, no. 2, pp. 153–170, 2004.

[47] E. Potokina, A. Druka, Z. Luo, R. Wise, R. Waugh, and M. J. Kearsey, "Gene expression quantitative trait locus analysis of 16 000 barley genes reveals a complex pattern of genome-wide transcriptional regulation," *The Plant Journal*, vol. 53, no. 1, pp. 90–101, 2008.

[48] R. Waugh, D. J. Leader, N. McCallum, and D. Caldwell, "Harvesting the potential of induced biological diversity," *Trends in Plant Science*, vol. 11, no. 2, pp. 71–79, 2006.

[49] D. G. Caldwell, N. McCallum, P. Shaw, G. J. Muehlbauer, D. F. Marshall, and R. Waugh, "A structured mutant population for forward and reverse genetics in Barley (*Hordeum vulgare* L.)," *The Plant Journal*, vol. 40, no. 1, pp. 143–150, 2004.

[50] T. Koprek, D. McElroy, J. Louwerse, R. Williams-Carrier, and P. G. Lemaux, "An efficient method for dispersing *Ds* elements in the barley genome as a tool for determining gene function," *The Plant Journal*, vol. 24, no. 2, pp. 253–263, 2000.

[51] L. D. Cooper, L. Marquez-Cedillo, J. Singh, et al., "Mapping *Ds* insertions in barley using a sequence-based approach," *Molecular Genetics and Genomics*, vol. 272, no. 2, pp. 181–193, 2004.

[52] J. Singh, S. Zhang, C. Chen, et al., "High-frequency *Ds* remobilization over multiple generations in barley facilitates gene tagging in large genome cereals," *Plant Molecular Biology*, vol. 62, no. 6, pp. 937–950, 2006.

[53] T. Zhao, M. Palotta, P. Langridge, et al., "Mapped *Ds*/T-DNA launch pads for functional genomics in barley," *The Plant Journal*, vol. 47, no. 5, pp. 811–826, 2006.

[54] M. A. Ayliffe, M. Pallotta, P. Langridge, and A. J. Pryor, "A barley activation tagging system," *Plant Molecular Biology*, vol. 64, no. 3, pp. 329–347, 2007.

[55] B. P. Forster, J. D. Franckowiak, U. Lundqvist, J. Lyon, I. Pitkethly, and W. T. B. Thomas, "The barley phytomer," *Annals of Botany*, vol. 100, no. 4, pp. 725–733, 2007.

[56] S. Qu, A. Desai, R. Wing, and V. Sundaresan, "A versatile transposon-based activation tag vector system for functional genomics in cereals and other monocot plants," *Plant Physiology*, vol. 146, pp. 189–199, 2008.

[57] P. B. Holm, O. Olsen, M. Schnorf, H. Brinch-Pedersen, and S. Knudsen, "Transformation of barley by microinjection into isolated zygote protoplasts," *Transgenic Research*, vol. 9, no. 1, pp. 21–32, 2000.

[58] G. Hensel, V. Valkov, J. Middlefell-Williams, and J. Kumlehn, "Efficient generation of transgenic barley: the way forward to modulate plant-microbe interactions," *Journal of Plant Physiology*, vol. 165, no. 1, pp. 71–82, 2008.

[59] J. Kumlehn, L. Serazetdinova, G. Hensel, D. Becker, and H. Loerz, "Genetic transformation of barley (*Hordeum vulgare* L.) via infection of androgenetic pollen cultures with *Agrobacterium tumefaciens*," *Plant Biotechnology Journal*, vol. 4, no. 2, pp. 251–261, 2006.

[60] D. von Wettstein, "From analysis of mutants to genetic engineering," *Annual Review of Plant Biology*, vol. 58, pp. 1–19, 2007.

[61] M. Kihara, Y. Okada, H. Kuroda, K. Saeki, N. Yoshigi, and K. Ito, "Improvement of β-amylase thermostability in transgenic barley seeds and transgene stability in progeny," *Molecular Breeding*, vol. 6, no. 5, pp. 511–517, 2000.

[62] A. Scheidig, A. Fröhlich, S. Schulze, J. R. Lloyd, and J. Kossmann, "Downregulation of a chloroplast-targeted β-amylase leads to a starch-excess phenotype in leaves," *The Plant Journal*, vol. 30, no. 5, pp. 581–591, 2002.

[63] J. Souppe and R. F. Beudeker, "Process for the production of alcoholic beverages using MaltSeed," US patent no. 6361808, 2002.

[64] L. G. Jensen, O. Olsen, O. Kops, N. Wolf, K. K. Thomsen, and D. von Wettstein, "Transgenic barley expressing a protein-engineered, thermostable (1,3-1,4)-β-glucanase during germination," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 93, no. 8, pp. 3487–3491, 1996.

[65] H. Horvath, J. Huang, O. Wong, et al., "The production of recombinant proteins in transgenic barley grains," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 4, pp. 1914–1919, 2000.

[66] A. M. Nuutila, A. Ritala, R. W. Skadsen, L. Mannonen, and V. Kauppinen, "Expression of fungal thermotolerant endo-1,4-β-glucanase in transgenic barley seeds during germination," *Plant Molecular Biology*, vol. 41, no. 6, pp. 777–783, 1999.

[67] D. von Wettstein, G. Mikhaylenko, J. A. Froseth, and C. G. Kannangara, "Improved barley broiler feed with transgenic malt containing heat-stable (1,3-1,4)-β-glucanase," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 25, pp. 13512–13517, 2000.

[68] D. von Wettstein, J. Warner, and G. G. Kannangara, "Supplements of transgenic malt or grain containing (1,3-1,4)-β-glucanase increase the nutritive value of barley-based broiler diets to that of maize," *British Poultry Science*, vol. 44, no. 3, pp. 438–449, 2003.

[69] G. P. Xue, M. Patel, J. S. Johnson, D. J. Smyth, and C. E. Vickers, "Selectable marker-free transgenic barley producing a high level of cellulase (1,4-β-glucanase) in developing grains," *Plant Cell Reports*, vol. 21, no. 11, pp. 1088–1094, 2003.

[70] Y. Stahl, S. Coates, J. H. Bryce, and P. C. Morris, "Antisense downregulation of the barley limit dextrinase inhibitor modulates starch granule size distribution, starch composition and amylopectin structure," *The Plant Journal*, vol. 39, no. 4, pp. 599–611, 2004.

[71] M.-J. Cho, J. H. Wong, C. Marx, W. Jiang, P. G. Lemaux, and B. B. Buchanan, "Overexpression of thioredoxin *h* leads to enhanced activity of starch debranching enzyme (pullulanase) in barley grain," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96, no. 25, pp. 14641–14646, 1999.

[72] J. H. Wong, Y.-B. Kim, P.-H. Ren, et al., "Transgenic barley grain overexpressing thioredoxin shows evidence that the starchy endosperm communicates with the embryo and the aleurone," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 25, pp. 16325–16330, 2002.

[73] C. Sun, A.-S. Höglund, H. Olsson, E. Mangelsen, and C. Jansson, "Antisense oligodeoxynucleotide inhibition as a potent strategy in plant biology: identification of SUSIBA2 as a

transcriptional activator in plant sugar signalling," *The Plant Journal*, vol. 44, no. 1, pp. 128–138, 2005.

[74] C. Sun, K. Ridderstråle, A.-S. Höglund, L.-G. Larsson, and C. Jansson, "Sweet delivery—sugar translocators as ports of entry for antisense oligodeoxynucleotides in plant cells," *The Plant Journal*, vol. 52, no. 6, pp. 1192–1198, 2007.

[75] F. Murray, R. Kalla, J. Jacobsen, and F. Gubler, "A role for HvGAMYB in anther development," *The Plant Journal*, vol. 33, no. 3, pp. 481–491, 2003.

[76] N. Sreenivasulu, L. Altschmied, R. Panitz, et al., "Identification of genes specifically expressed in maternal and filial tissues of barley caryopses: a cDNA array analysis," *Molecular Genetics and Genomics*, vol. 266, no. 5, pp. 758–767, 2001.

[77] V. Radchuk, L. Borisjuk, R. Radchuk, et al., "Jekyll encodes a novel protein involved in the sexual reproduction of barley," *Plant Cell*, vol. 18, no. 7, pp. 1652–1666, 2006.

[78] C. Künne, M. Lange, T. Funke, et al., "CR-EST: a resource for crop ESTs," *Nucleic Acids Research*, vol. 33, pp. D619–D621, 2005.

[79] S. Weise, I. Grosse, C. Klukas, et al., "Meta-All: a system for managing metabolic pathway information," *BMC Bioinformatics*, vol. 7, article 465, pp. 1–9, 2006.

[80] E. Grafahrend-Belau, S. Weise, D. Koschützki, U. Scholz, B. H. Junker, and F. Schreiber, "MetaCrop: a detailed database of crop plant metabolism," *Nucleic Acids Research*, vol. 36, pp. D954–D958, 2008.

[81] B. H. Junker, C. Klukas, and F. Schreiber, "VANTED: a system for advanced data analysis and visualization in the context of biological networks," *BMC Bioinformatics*, vol. 7, article 109, pp. 1–13, 2006.

[82] H. Rolletschek, W. Weschke, H. Weber, U. Wobus, and L. Borisjuk, "Energy state and its control on seed development: starch accumulation is associated with high ATP and steep oxygen gradients within barley grains," *Journal of Experimental Botany*, vol. 55, no. 401, pp. 1351–1359, 2004.

[83] S. Gubatz, V. J. Dercksen, C. Brüß, W. Weschke, and U. Wobus, "Analysis of barley (*Hordeum vulgare*) grain development using three-dimensional digital models," *The Plant Journal*, vol. 52, no. 4, pp. 779–790, 2007.

[84] M. Stark, B. Manz, A. Ehlers, et al., "Multiparametric high-resolution imaging of barley embryos by multiphoton microscopy and magnetic resonance micro-imaging," *Microscopy Research and Technique*, vol. 70, no. 5, pp. 426–432, 2007.

[85] T. Neuberger, N. Sreenivasulu, M. Rokitta, et al., "Quantitative imaging of oil storage in developing crop seeds," *Plant Biotechnology Journal*, vol. 6, no. 1, pp. 31–45, 2008.

[86] B. H. Junker, D. Koschützki, and F. Schreiber, "Kinetic modeling with the systems biology modelling environment SyBME," *Journal of Integrative Bioinformatics*, vol. 3, no. 1, Article ID 0018, 10 pages, 2006.

[87] I. Koch, B. H. Junker, and M. Heiner, "Application of Petri net theory for modelling and validation of the sucrose breakdown pathway in the potato tuber," *Bioinformatics*, vol. 21, no. 7, pp. 1219–1226, 2005.

[88] C. Finnie, S. Melchior, P. Roepstorff, and B. Svensson, "Proteome analysis of grain filling and seed maturation in barley," *Plant Physiology*, vol. 129, no. 3, pp. 1308–1319, 2002.

[89] K. Witzel, G.-K. Surabhi, G. Jyothsnakumari, C. Sudhakar, A. Matros, and H.-P. Mock, "Quantitative proteome analysis of barley seeds using ruthenium(II)-tris-(bathophenanthroline-disulphonate) staining," *Journal of Proteome Research*, vol. 6, no. 4, pp. 1325–1333, 2007.

[90] A. Graner, A. Jahoor, J. Schondelmaier, et al., "Construction of an RFLP map of barley," *Theoretical and Applied Genetics*, vol. 83, no. 2, pp. 250–256, 1991.

[91] M. Heun, A. E. Kennedy, J. A. Anderson, N. L. V. Lapitan, M. E. Sorrells, and S. D. Tanksley, "Construction of a restriction fragment length polymorphism map for barley (*Hordeum vulgare*)," *Genome*, vol. 34, no. 3, pp. 437–447, 1991.

[92] A. Kleinhofs, A. Kilian, M. A. Saghai Maroof, et al., "A molecular, isozyme and morphological map of the barley (*Hordeum vulgare*) genome," *Theoretical and Applied Genetics*, vol. 86, no. 6, pp. 705–712, 1993.

[93] P. M. Hayes, A. Castro, L. Marquez-Cedillo, et al., "Genetic diversity for quantitatively inherited agronomic and malting quality traits," in *Diversity in Barley (*Hordeum vulgare*)*, R. von Bothmer, T. van Hintum, H. Knüpffer, and K. Sato, Eds., pp. 201–226, Elsevier Science, Amsterdam, The Netherlands, 2003.

[94] W. Friedt and F. Ordon, "Molecular markers for gene pyramiding and resistance breeding in barley," in *Genomics-Assisted Crop Improvement, Vol 2: Genomics Applications in Crops*, R. Varshney and R. Tuberosa, Eds., p. 498, Springer, Berlin, Germany, 2008.

[95] T. Thiel, W. Michalek, R. K. Varshney, and A. Graner, "Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.)," *Theoretical and Applied Genetics*, vol. 106, no. 3, pp. 411–422, 2003.

[96] R. Kota, M. Wolf, W. Michalek, and A. Graner, "Application of denaturing high-performance liquid chromatography for mapping of single nucleotide polymorphisms in barley (*Hordeum vulgare* L.)," *Genome*, vol. 44, no. 4, pp. 523–528, 2001.

[97] R. Kota, R. K. Varshney, M. Prasad, H. Zhang, N. Stein, and A. Graner, "EST-derived single nucleotide polymorphism markers for assembling genetic and physical maps of the barley genome," *Functional & Integrative Genomics*. In press.

[98] R. Kota, S. Rudd, A. Facius, et al., "Snipping polymorphisms from large EST collections in barley (*Hordeum vulgare* L.)," *Molecular Genetics and Genomics*, vol. 270, no. 1, pp. 24–33, 2003.

[99] T. Thiel, R. Kota, I. Grosse, N. Stein, and A. Graner, "SNP2CAPS: a SNP and INDEL analysis tool for CAPS marker development," *Nucleic Acids Research*, vol. 32, no. 1, p. e5, 2004.

[100] A. Karakousis, J. P. Gustafson, K. J. Chalmers, A. R. Barr, and P. Langridge, "A consensus map of barley integrating SSR, RFLP, and AFLP markers," *Australian Journal of Agricultural Research*, vol. 54, no. 11-12, pp. 1173–1185, 2003.

[101] A. A. Diab, "Construction of barley consensus map showing chromosomal regions associated with economically important traits," *African Journal of Biotechnology*, vol. 5, no. 3, pp. 235–248, 2006.

[102] R. K. Varshney, T. C. Marcel, L. Ramsay, et al., "A high density barley microsatellite consensus map with 775 SSR loci," *Theoretical and Applied Genetics*, vol. 114, no. 6, pp. 1091–1103, 2007.

[103] N. Rostoks, L. Ramsay, K. MacKenzie, et al., "Recent history of artificial outcrossing facilitates whole-genome association mapping in elite inbred crop varietes," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 49, pp. 18656–18661, 2006.

[104] P. Wenzl, J. Carling, D. Kudrna, et al., "Diversity Arrays Technology (DArT) for whole-genome profiling of barley,"

Proceedings of the National Academy of Sciences of the United States of America, vol. 101, no. 26, pp. 9915–9920, 2004.

[105] P. Wenzl, H. Li, J. Carling, et al., "A high-density consensus map of barley linking DArT markers to SSR, RFLP and STS loci and agricultural traits," *BMC Genomics*, vol. 7, article 206, pp. 1–22, 2006.

[106] N. Rostoks, J. O. Borevitz, P. E. Hedley, et al., "Single-feature polymorphism discovery in the barley transcriptome," *Genome Biology*, vol. 6, no. 6, p. R54, 2005.

[107] A. Graner and E. Bauer, "RFLP mapping of the rym4 virus resistance gene in barley," *Theoretical and Applied Genetics*, vol. 86, no. 6, pp. 689–693, 1993.

[108] A. Graner, S. Streng, A. Kellermann, et al., "Molecular mapping and genetic fine-structure of the rym5 locus encoding resistance to different strains of the barley yellow mosaic virus complex," *Theoretical and Applied Genetics*, vol. 98, no. 2, pp. 285–290, 1999.

[109] N. Stein, D. Perovic, J. Kumlehn, et al., "The eukaryotic translation initiation factor 4E confers multiallelic recessive Bymovirus resistance in *Hordeum vulgare* (L.)," *The Plant Journal*, vol. 42, no. 6, pp. 912–922, 2005.

[110] J. A. Mammadov, W. S. Brooks, C. A. Griffey, and M. A. Saghai Maroof, "Validating molecular markers for barley leaf rust resistance genes Rph5 and Rph7," *Plant Breeding*, vol. 126, no. 5, pp. 458–463, 2007.

[111] K. Werner, W. Friedt, and F. Ordon, "Strategies for pyramiding resistance genes against the barley yellow mosaic virus complex (BaMMV, BaYMV, BaYMV-2)," *Molecular Breeding*, vol. 16, no. 1, pp. 45–55, 2005.

[112] S. Salvi and R. Tuberosa, "To clone or not to clone plant QTLs: present and future challenges," *Trends in Plant Science*, vol. 10, no. 6, pp. 297–304, 2005.

[113] T. Sutton, U. Baumann, J. Hayes, et al., "Boron-toxicity tolerance in barley arising from efflux transporter amplification," *Science*, vol. 318, no. 5855, pp. 1446–1449, 2007.

[114] S. J. Rae, M. Macaulay, L. Ramsay, et al., "Molecular barley breeding," *Euphytica*, vol. 158, no. 3, pp. 295–303, 2007.

[115] W. T. B. Thomas, "Prospects for molecular breeding of barley," *Annals of Applied Biology*, vol. 142, no. 1, pp. 1–12, 2003.

[116] S. Doss, E. E. Schadt, T. A. Drake, and A. J. Lusis, " *Cis*-acting expression quantitative trait loci in mice," *Genome Research*, vol. 15, no. 5, pp. 681–691, 2005.

[117] D. E. Reich, M. Cargill, S. Bolk, et al., "Linkage disequilibrium in the human genome," *Nature*, vol. 411, no. 6834, pp. 199–204, 2001.

[118] A. Rafalski and M. Morgante, "Corn and humans: recombination and linkage disequilibrium in two genomes of similar size," *Trends in Genetics*, vol. 20, no. 2, pp. 103–111, 2004.

[119] P. K. Gupta, S. Rustgi, and P. L. Kulwal, "Linkage disequilibrium and association studies in higher plants: present status and future prospects," *Plant Molecular Biology*, vol. 57, no. 4, pp. 461–485, 2005.

[120] A. T. W. Kraakman, R. E. Niks, P. M. Van den Berg, P. Stam, and F. A. Van Eeuwijk, "Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars," *Genetics*, vol. 168, no. 1, pp. 435–446, 2004.

[121] I. Romagosa, F. Han, S. E. Ullrich, P. M. Hayes, and D. M. Wesenberg, "Verification of yield QTL through realized molecular marker-assisted selection responses in a barley cross," *Molecular Breeding*, vol. 5, no. 2, pp. 143–152, 1999.

[122] J. Z. Li, X. Q. Huang, F. Heinrichs, M. W. Ganal, and M. S. Röder, "Analysis of QTLs for yield components, agronomic traits, and disease resistance in an advanced backcross population of spring barley," *Genome*, vol. 49, no. 5, pp. 454–466, 2006.

[123] L. V. Malysheva-Otto, M. W. Ganal, and M. S. Röder, "Analysis of molecular diversity, population structure and linkage disequilibrium in a worldwide survey of cultivated barley germplasm (*Hordeum vulgare* L.)," *BMC Genetics*, vol. 7, article 6, pp. 1–14, 2006.

[124] K. S. Caldwell, J. Russell, P. Langridge, and W. Powell, "Extreme population-dependent linkage disequilibrium detected in an inbreeding plant species, *Hordeum vulgare* ," *Genetics*, vol. 172, no. 1, pp. 557–567, 2006.

[125] P. L. Morrell, D. M. Toleno, K. E. Lundy, and M. T. Clegg, "Low levels of linkage disequilibrium in wild barley (*Hordeum vulgare* ssp. *spontaneum*) despite high rates of self-fertilization," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 7, pp. 2442–2447, 2005.

[126] B. J. Steffenson, P. Olivera, J. K. Roy, Y. Jin, K. P. Smith, and G. J. Muehlbauer, "A walk on the wild side: mining wild wheat and barley collections for rust resistance genes," *Australian Journal of Agricultural Research*, vol. 58, no. 6, pp. 532–544, 2007.

[127] S. Stracke, T. Presterl, N. Stein, D. Perovic, F. Ordon, and A. Graner, "Effects of introgression and recombination on haplotype structure and linkage disequilibrium surrounding a locus encoding Bymovirus resistance in barley," *Genetics*, vol. 175, no. 2, pp. 805–817, 2007.

[128] R. C. Jansen and J.-P. Nap, "Genetical genomics: the added value from segregation," *Trends in Genetics*, vol. 17, no. 7, pp. 388–391, 2001.

[129] N. Sreenivasulu, R. K. Varshney, P. B. Kavi-Kishor, and W. Weschke, "Functional genomics for tolerance to abiotic stress in cereals," in *Cereal Genomics*, P. K. Gupta and R. K. Varshney, Eds., pp. 483–514, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2004.

[130] R. K. Varshney, A. Graner, and M. E. Sorrells, "Genomics-assisted breeding for crop improvement," *Trends in Plant Science*, vol. 10, no. 12, pp. 621–630, 2005.

[131] E. Potokina, M. Prasad, L. Malysheva, M. S. Röder, and A. Graner, "Expression genetics and haplotype analysis reveal *cis* regulation of serine carboxypeptidase I (Cxp1), a candidate gene for malting quality in barley (*Hordeum vulgare* L.)," *Functional and Integrative Genomics*, vol. 6, no. 1, pp. 25–35, 2006.

[132] J. Z. Li, X. Q. Huang, F. Heinrichs, M. W. Ganal, and M. S. Röder, "Analysis of QTLs for yield, yield components, and malting quality in a BC3-DH population of spring barley," *Theoretical and Applied Genetics*, vol. 110, no. 2, pp. 356–363, 2005.

[133] T. Wicker, E. Schlagenhauf, A. Graner, T. J. Close, B. Keller, and N. Stein, "454 sequencing put to the test using the complex genome of barley," *BMC Genomics*, vol. 7, article 275, pp. 1–11, 2006.

[134] G. Hammer, M. Cooper, F. Tardieu, et al., "Models for navigating biological complexity in breeding improved crop plants," *Trends in Plant Science*, vol. 11, no. 12, pp. 587–593, 2006.

[135] X. Yin, P. C. Struik, and M. J. Kropff, "Role of crop physiology in predicting gene-to-phenotype relationships," *Trends in Plant Science*, vol. 9, no. 9, pp. 426–432, 2004.

[136] P. Langridge, "Molecular breeding of wheat and barley," in *The Wake of Double Helix: From the Green Revolution to the Gene Revolution*, R. Tuberosa, R. L. Phillips, and M. Gale, Eds., pp. 279–286, Avenue Media, Bologna, Italy, 2005.

Hindawi Publishing Corporation
International Journal of Plant Genomics
Volume 2008, Article ID 348621, 9 pages
doi:10.1155/2008/348621

Review Article

Development in Rice Genome Research Based on Accurate Genome Sequence

Takashi Matsumoto, Jianzhong Wu, Baltazar A. Antonio, and Takuji Sasaki

Division of Genome and Biodiversity Research, National Institute of Agrobiological Sciences, 2-1-2, Kannondai, Tsukuba, Ibaraki 3058602, Japan

Correspondence should be addressed to Takashi Matsumoto, mat@nias.affrc.go.jp

Received 7 September 2007; Revised 17 April 2008; Accepted 9 May 2008

Recommended by Yunbi Xu

Rice is one of the most important crops in the world. Although genetic improvement is a key technology for the acceleration of rice breeding, a lack of genome information had restricted efforts in molecular-based breeding until the completion of the high-quality rice genome sequence, which opened new opportunities for research in various areas of genomics. The syntenic relationship of the rice genome to other cereal genomes makes the rice genome invaluable for understanding how cereal genomes function. Producing an accurate genome sequence is not an easy task, and it is becoming more important as sequence deviations among, and even within, species highlight functional or evolutionary implications for comparative genomics.

1. INTRODUCTION

Food security is a major issue as we aspire toward sustainable development. In spite of continuous increases in agricultural production due to the introduction of improved crop cultivars and the wide use of affordable technologies, more than 800 million people still do not have access to sufficient food to meet their dietary needs [1]. Cereal crops are basic source of food for humankind, with 85% of total crop production represented by maize, wheat, and rice. These three crops provide more than half of the protein and energy required for daily life. However, increase of world agricultural production in 2006 was less than 1%, which was due to decrease in cereal production [2]. On the other hand, the world's population is expected to reach 9 billion by 2050 [3]. It is therefore necessary to provide food security to this growing population in the midst of global environmental problems that deprive us of much arable land and biodiversity.

Worldwide transformation of agriculture was first achieved with the Green Revolution, which led to significant increases in agricultural production. It began in the 1940s with the cultivation of a high-yielding dwarf wheat cultivar with resistance to pests and diseases. The Green Revolution for rice in the 1960s, based on the cultivar IR8, also dramatically increased rice production and helped food production to keep pace with population growth.

Now, the second Green Revolution, which will be based on genomics, is expected to pave the way for the leap in crop production. The availability of the rice genome sequence allowed the development of innovative approaches to increasing production. In the last 10 years, the basic syntenic relationships in gene content and gene order within the grass family have been established [4–6]. Therefore, the rice genome could be used as a reference genome for understanding the evolution of cereal crops and could provide a basis for their improvement [7, 8].

Among plants, only the Arabidopsis [9] and rice [10] genome sequences have been completed so far. A positionally confirmed, quality-validated genome sequence is obligatory required as a reference for the efficient use of sequence information, particularly in comparative analysis. Hence, the genome sequence derived from *Oryza sativa* ssp. *japonica* cv. Nipponbare has been recognized as a gold standard for understanding the genetics and biology of rice at the molecular level and in the breeding and genetic manipulation of cereal crops.

This chapter presents a past history of the rice genome sequencing efforts and a present endeavor for analysis of the genome sequence to clarify its structure and function.

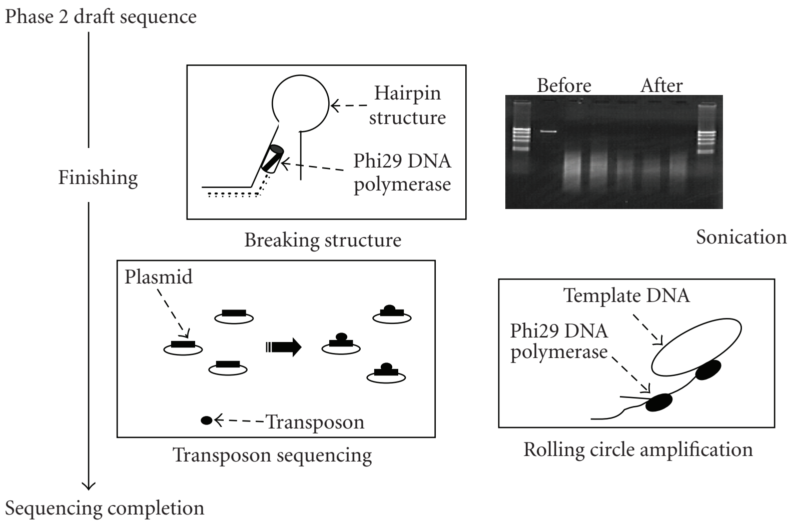


FIGURE 1: Four steps used for the finishing process to sequence completion.

Approach to the "difficult" regions whose functions are the maintenance and regulation of chromosomes—notably the centromeres and telomeres—is described. Application of the new sequencing technology toward comparative studies among genus *Oryza* is also described in the context of the rice genome as a reference.

2. GENOME SEQUENCING THROUGH INTERNATIONAL COLLABORATION

The International Rice Genome Sequencing Project (IRGSP) was established in 1997. The 10 member countries agreed to sequence a standard rice cultivar (Nipponbare), to use common resources, and to share sequencing of the 12 rice chromosomes by using a map-based clone-by-clone strategy (http://rgp.dna.affrc.go.jp/E/IRGSP/index.html). For construction of sequence-ready physical maps, two complementary approaches were used. The Rice Genome Research Program (RGP) in Japan anchored the genomic clones using expressed sequence tags/sequence-tagged sites (EST/STS) and genetic markers from the genetic and transcript maps of rice [11, 12]. The Clemson University Genomics Institute, the Arizona Genomics Institute, and the Arizona Genomics Computational Laboratory used a high-throughput bacterial artificial chromosome (BAC) fingerprint and automatic BAC contig assembly system using FPC software [13], and anchored the assembled contigs on the rice genome by hybridization-based screening [14]. The sequence-ready physical maps generated from use of these two strategies covered more than 95% of the rice genome, and 92% to 100% of each chromosome. A total of 3453 PAC/BAC clones forming the minimum tiling path were selected for sequencing. DNA from a BAC/PAC clone was purified and fragmented by sonication. The ends of 2000 subclones of each clone were sequenced with capillary sequencers and assembled using the phred/phrap assembler [15]. The genome sequences of each PAC/BAC clone at the high-throughput genomic (HTG) phase 2 category were submitted to the DNA Data Bank of Japan (DDBJ). By December 2002, almost all the clone sequences corresponding to the minimum tiling path were sequenced to at least HTG phase 2. As a result, a high-quality draft sequence representing 366 Mb of the rice genome was released in the public database [16]. Thereafter, the IRGSP continued with the arduous task of finishing: gap-filling, improving base read quality, and resolving misassemblies (Figure 1).

In December 2004, the high-quality map-based sequence of the rice genome at HTG phase 3 category was completed and released in the public domain [10]. The sequence, ca. 370 Mb in total, covered nearly 95% of the total estimated size of the genome and about 99% of the euchromatic regions. The sequence also included three centromeres, parts of the rDNA regions, and regions for various transposable elements (corresponding up to 35% in the total genome). This comprehensive, relatively accurate sequence of the rice genome, is currently considered the gold standard.

In contrast to the hierarchical clone-by-clone strategy used by the IRGSP, a whole-genome shotgun (WGS) sequencing strategy is widely used in many sequencing projects [17]. In this strategy, a high-throughput computer program to reproduce the entire genome sequence assembles millions of shotgun sequences from the total genome. This method was used in sequencing the 2.9-Gb human genome [18]. Two independent groups used the WGS strategy to sequence the rice genome. The Beijing Genome Institute assembled shotgun sequences of the *indica* line 93-11 with 4× [19] and later 6× [20] genome coverage. A private company, Syngenta (Basel, Switzerland), also used the WGS strategy to sequence the Nipponbare [21]. This WGS sequence of Nipponbare was further improved by reassembling the shotgun sequences and combining the *japonica* and *indica* (line 99-11) sequences, resulting in 433 Mb of sequence composed of 50 233 contigs of Nipponbare [20]. Nearly 99% of the rice full-length cDNAs [22] have been localized in these latest assemblies [20] of the *japonica* and *indica* genome.

The effectiveness of the WGS sequencing strategy was compared with that of the hierarchal clone-by-clone sequencing approach [23, 24]. Although WGS assembly could readily provide an overview of the genome structure with a practical level of accuracy, misassembly could result in nonhomologous, misaligned, or duplicated coverage and some mismatches even in the genic regions. Moreover, repeat sequences could not be properly assigned to their original positions in the genome. In the case of rice, which has a lot of repeat sequences, WGS sequencing is therefore not a highly reliable strategy as it creates misassembly, particularly in duplicated regions. It is therefore important to have a highly accurate map-based sequence, which can be obtained by the hierarchical clone-by-clone strategy. Today projects aiming at obtaining entire genome sequences of gramineae plants are progressing [25–29]. All the projects, either using WGS strategy or clone-by-clone strategy, regard the completed rice genome sequence as sequence reference in reconstruction of chromosome sequences, emphasizing the importance of "gold standard."

3. DECIPHERING THE GENOME THROUGH ANNOTATION

Detecting the gene-coding regions within the genome sequence is one of the most efficient ways to characterize the structure and function of the genome. RGP constructed an annotation system that facilitates gene detection of the genome sequence in a timely manner. The Rice Genome Automated Annotation System, or RiceGAAS (http://ricegaas.dna.affrc.go.jp [30]), was designed as a fully automated system for annotating rice genome sequences. It retrieves rice sequences from GenBank and analyzes them with gene prediction programs such as Genscan [31] and FgeneSH (http://www.softberry.com/berry.phtml) and with BLAST [32] for similarity to proteins, rice ESTs, and rice full-length cDNAs to generate the most accurate gene models on the basis of available information (Figure 2). A similar automatic annotation pipeline was established by TIGR (http://rice.tigr.org/tdb/e2k1/osa1/data_download.shtml), and gene models are improved with rice ESTs and transcripts [33]. Both sets of gene models are published on the Web to accelerate gene analysis. With increasing data on nucleic acids and proteins in the public databases, regular re-evaluation and update of these gene models is necessary. In this respect, one of the advantages of these full-computational approaches is that whole gene sets can be relatively easily revised.

RGP has also developed a manual annotation system to facilitate curation of the gene models by human annotators (http://rgp.dna.affrc.go.jp/genomicdata/AnnSystem.html). This pipeline directly takes the output generated from RiceGAAS for in-depth analysis with in-house editing tools. Each gene model is manually edited to improve the prediction accuracy. The gene models for each BAC or PAC clone are released to the public domain through the DDBJ/EMBL/GenBank database. All data can be accessed through the central database whole genome annotation (WhoGA) on our website at http://rgp.dna.affrc.go.jp. Initially, only the six chromosomes (1, 2, 6, 7, 8, and 9) assigned to RGP were manually curated. Recently, curation of the rest was completed, so the manual annotation of the entire genome is now available. After removal of clone overlaps, a total of 57 724 genes were predicted, including many hypothetical genes predicted by a single prediction program. Among them, 24056 gene models are supported by full-length cDNAs. All the gene models are ordered and organized in a genome browser.

Apart from these individual activities, the IRGSP conceived the establishment of the Rice Annotation Project (RAP), a community standard annotation project, in 2004. Genes were annotated at regular jamboree-style annotation meetings to facilitate the manual curation of all gene models in rice. The National Institute of Agrobiological Sciences has been leading this project, collaborating with IRGSP members and many international and Japanese laboratories. So far, three RAP meetings have been held, at which gene models, chiefly constructed by mapping full-length cDNAs on the latest rice genome assemblies, have been manually curated. This collaboration confirmed 32000 curated genes, most of which have some degree of evidence [34]. The RAP-database (RAP-DB, http://rapdb.lab.nig.ac.jp) will be further improved with the integration of other annotation and functional genomics data.

4. UNCOVERED TERRITORY—EXPLORATION OF THE MISSING REGIONS

At the time of completion of the genome in 2004, IRGSP published nearly 371 Mb of high-quality DNA sequences, leaving about 5% of its estimated 389 Mb to be sequenced [10]. These unsequenced genomic regions existed as 62 gaps, including the telomeres and centromeres in all but two out of 12 chromosomes. One of the main reasons for the presence of these gaps was that no more clones with sequence extension into gap regions could be selected from any Nipponbare genomic resources, including BAC and PAC libraries (both based on partial digestion of DNA fragments) and fosmid libraries (based on physically sheared DNA fragments), containing a total of 630 000 clones. For unknown reasons, specific genomic regions could not be cloned or maintained by using the above vectors in bacteria. In addition, a number of regions in the genome contain highly repeated sequences, making it difficult to construct a correct and complete physical map. However, analysis of sequences from these complicated genomic regions is not futile. Researchers have reported the importance of heterochromatic regions in silencing gene expression [35]. Cytological analysis has been used to define the distribution of such heterochromatin along each rice chromosome [36]. Through the IRGSP efforts, 2 of the 12 centromeres and 14 of the 24 telomeres have been completely or partly sequenced (http://rgp.dna.affrc.go.jp/E/IRGSP/index.html). Here, we focus on both regions because they play essential roles in chromosome maintenance or segregation.

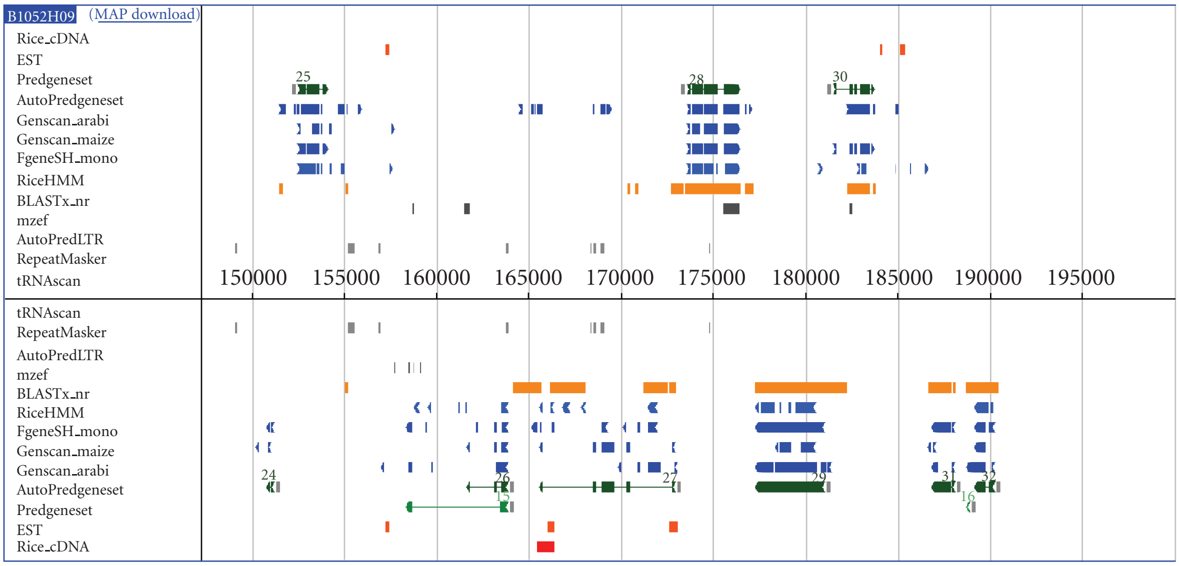


FIGURE 2: RiceGAAS annotation view, showing results from application of gene prediction software and similarity searches. Upper box: a DNA strand from left (5′) to right (3′). Lower box: from right (5′) to left (3′).

4.1. Composition and structure of rice centromeres

Because of the relatively small amount of centromeric satellite DNA in rice, significant progress has been made in genomic and molecular studies of the structures, functions, and evolution of rice centromeres. Two centromeres, derived from chromosomes 4 and 8, have been completely sequenced, revealing the complicated composition and structure of the first centromeres to have been sequenced among eukaryotes [37–39]. Repetitive sequences occupy ~60% of the whole region (~2 Mb) of the centromere of chromosome 8 (*Cen8*). The majority of copies of the 155-bp centromeric satellite repeat *CentO*, totaling 68.5-kb, occur in three large clusters in the center, separated by centromere-specific retrotransposon of rice (*CRR*) sequences. Numerous sequences of other transposable elements were also found in its surrounding region. *Cen8* contains an ~750-kb core domain that binds rice CENH3, the centromere-specific H3 histone [37]. It is surprising to find transcriptionally active genes even within the core domain of *Cen8*. A similar result was found in *Cen3*, where a much bigger region (~1881 kb) has been found to have associations with CENH3 [40]. As a chromosomal site for kinetochore assembly that plays an important role in the faithful segregation of sister chromatids during cell division, the centromere has functions that are well conserved among all higher eukaryotes. Inter- and extrachromosomal analysis of the centromeres has, however, revealed the divergence of DNA components and organization patterns even among closely related species. The amount of *CentO* satellite DNA in the centromere of individual chromosomes varies from 60 kb to 1.9 Mb in *O. sativa* [41]. The number and organization of *CentO* clusters within the core region differ markedly between *Cen4* and *Cen8* in the Nipponbare genome. *Cen8* has only three *CentO* tracts (clusters) with 442 copies of the 155-bp tandem repeat distributed within a 75-kb region, whereas *Cen4* has up to 18 tracts but only 379 copies of the repeat within a 124-kb region [38, 39] (Figure 3). *CentO* repeats, on the other hand, are absent from several wild rice species, such as *Oryza brachyantha* [42]. It would be interesting to sequence and compare the compositional and structural changes in centromeres between different *Oryza* species in the future, since in-depth analysis of the *Cen8* and *Cen4* sequences has demonstrated segmental duplication and inversion of centromeric DNA [43]. First glimpse of this analysis was performed in sequencing the centromere region of chromosome.8 from *O. brachyantha*, revealing positional shift of centromere [44].

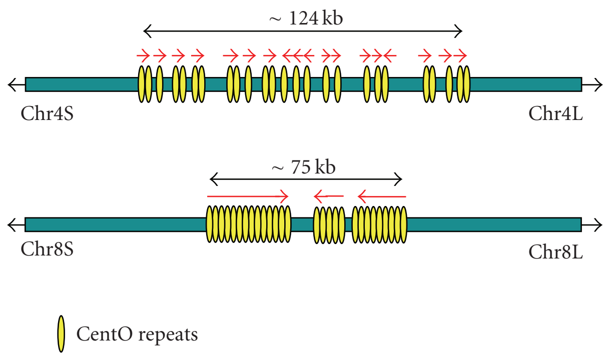


FIGURE 3: Structural comparisons of *CentO* domains between Nipponbare chromosomes 4 and 8. Yellow ovals and red arrows indicate the position of *CentO* arrays and the direction of the 155-bp tandem repeats within each array, respectively. Length of arrays ranges from 477 to 8571 bp in chromosome 4 and 7616 to 34589 bp in chromosome 8.

Rice is now becoming a model for centromere and heterochromatin research [38, 45, 46]. Further research will lead to insights into the evolutionary dynamics, processes, and molecular mechanisms of plant centromeres.

4.2. Composition and structure of rice telomeres

Like those of centromeres, the composition and structure of telomere regions in rice have also been analyzed. Telomeres form the ends of linear eukaryotic chromosomes, serving as protective caps that prevent end-to-end fusion, recombination, and degradation of chromosomal ends [47]. The telomeres of most eukaryotes consist of an array of repeats that contain similar sequences but vary in length. For example, telomere DNA has a conserved sequence of 5′-TTAGGG-3′ in humans and 5′-TTGGGG-3′ in *Tetrahymena* (a ciliate protozo) [48, 49]. The first plant telomere DNA was isolated from *Arabidopsis thaliana* and shows tandemly repeated arrays of 5′-TTTAGGG-3′ [50]. Rice telomeres consist of the same repeat [51]. Sequencing and extensive analysis of seven rice chromosomal ends revealed several basic features that could provide a platform for analyzing and understanding the telomere structures and functions. All seven rice telomeres revealed contain highly conserved TTTAGGG sequences in tandem repeats, although deletions, insertions, and substitutions of single nucleotides or inverted copies were found within the arrayed repeats, particularly in the region of the junction between the telomere and subtelomere. Fluorescent in situ hybridization and terminal restriction fragment analyses suggest that the rice telomeres are a bit longer than those of *Arabidopsis* but much shorter than those of *Nicotiana tabacum*, ranging in a length from 5 to 20 kb, thus hinting at the genetic control of telomere length in plants [52, 53]. Interestingly, variation in telomere length is observed not only among different chromosomes, but also between different species within *Oryza*; this variation should provide useful information for future studies of telomere evolution. Gene annotation in the 7 rice subtelomere regions (each within 500 kb) demonstrated that the genomic region adjacent to the chromosome terminus is gene-rich (1 gene per 5.9 kb on average). Since nearly half of these annotated genes match rice full-length cDNAs, these rice subtelomeres could be considered to have high transcriptional activity. Recently, seven new rice telomeres were partly sequenced, and their sequences have been submitted to DDBJ (Table 1; http://rgp.dna.affrc.go.jp/E/publicdata/telomere2007/index.html). Among the above 14 chromosomal ends, the telomere and subtelomere regions on the short arm of chromosome 9 show some specific compositional and structural features. Sequencing and analysis of the fosmid clone OSJNOa063K24 revealed that the telomere repeats are colocalized with the ribosomal RNA gene (rDNA) cluster [54]. Besides the telomere-specific repeat and the long rDNA array (sized in megabases), the content of repetitive sequences such as retrotransposons within the 500-kb region proximal to the centromere is relatively high, suggesting that much of the short arm of rice chromosome 9 is heterochromatic. Rice telomere reverse transcriptase has also been isolated [55]. It will be interesting to conduct future studies using rice as a model of telomere research, as has been done for centromeres, especially to reveal how telomere length (shortening or elongation) is regulated and whether the telomere repeats and structure affect the expression of genes in the subtelomere region. The sequence resources obtained from the telomere and centromere regions of rice chromosomes should thus provide an unprecedented opportunity for future study, particularly to construct an artificial chromosome for use in both molecular and applied biology in plant science.

Table 1: Mapped and sequenced rice telomeres.

Clone name	Accession no.	Copies*	Chr
OSJNOa264G09	AP008219	17	1S
OSJNOa183H18	AP006851	52	2S
OSJNOa246I10	AP008220	69	2L
OSJNOa070P15	AP009053	27	3S
OSJNOa083A04	AP009055	75	3L
OSJNOa076I12	AP009056	129	4S
OSJNOa281H13	AP009057	68	4L
OSJNOa070B13	AP009052	53	5S
OSJNOa230J22	AP006854	37	6L
OSJNOa219C16	AP008222	17	7S
OSJNOa136M23	AP008223	127	7L
OSJNOa162K02	AP008224	55	8S
OSJNOa063K24	AP009051	162	9S
OSJNOa073B23	AP009054	62	10S

*Copies of telomere-specific repeats detected from the sequenced clones.

5. GENOME SEQUENCE FOR EVOLUTIONARY GENOMICS IN RICE

Rice is believed to have been domesticated from a wild relative 0.2 Mya [56] or 0.44 Mya [57]. Asian cultivated rice (*O. sativa* L.) has two subspecies, *indica* and *japonica*. Both are important as modern crops, and there are many phenotypic variations among them, conferring adaptation to many different environmental and cultural conditions. Crossing of these subspecies has produced new cultivars of agricultural importance. Knowing the differences at the molecular level would widen the capacity for rice breeding. RGP constructed a BAC library of Kasalath, an *indica* cultivar, generating 78427 high-quality BAC end sequences from 47194 BAC clones, and mapped these end sequences on Nipponbare chromosome sequences [58]. Mapping of 12170 clones allowed the construction of 450 Kasalath BAC contigs covering 308.5 Mb. Single-nucleotide polymorphism (SNP) frequency in the BAC end sequences and corresponding Nipponbare sequences was 0.71% on average. Sequencing of part of the Kasalath genome is in progress and could in future elucidate the precise gene dynamics in evolution and domestication. Results of Kasalath BAC physical maps are shown on RGP homepage (http://rgp.dna.affrc.go.jp/E/publicdata/kasalathendmap/index.html). Figure 4 is an example of a computer-generated Kasalath

BAC physical map. BLAST searches for Kasalath BAC-end sequence screening could be performed through website (http://rgp.dna.affrc.go.jp/blast/runblast.html). The other approaches [59, 60] could identify positions of SNPs for high-density SNP markers.

It had long been a mystery how Asian rice originated from its wild progenitor, *Oryza rufipogon*. Recently, the origin has been clarified by comparison of retrotransposon [56], retroposon [61], chloroplast [62], and gene [63] sequences along the evolutionary lineages. These studies show evidence of multiple independent domestications of the two major subspecies. Further molecular studies of domestication will show how the crop and humans coevolved.

The genus *Oryza* has 23 species [64], but only two species (*O. sativa* in Asia and *O. glaberrima* in Africa) are domesticated and cultivated. This fact is remarkable given that rice grows under a wide variety of natural conditions. Consequently, many genetic resources might be waiting to be developed. Study of the wild relatives might reveal new genes for hybridization, improved yield, and sustainable production. The *Oryza* Map Alignment Project (OMAP, http://www.omap.org/index.html) of the USA and China aims at the establishment of an experimental platform to unravel and understand the evolution, physiology, and biochemistry of the genus. The Arizona Genomics Institute has constructed 12 BAC libraries from the AA (the same as sativa species) to HHKK (remote species from sativa) species genomes. Computer-based mapping and filter hybridization screening provided high-density cross-species physical maps [65, 66].

Table 2: Sequence comparison of BAC clone OR_CBa0004O24, Sanger versus Pyrosequencing.

	Low-quality threshold	
	Score 30	Score 40
Number of alignments checked	34	34
Total length of alignments	132229	132229
Total HQ bases	131759	130639
Total LQ bases	470	1590
"In/del" type discrepancy	20	20
– Sanger insertion, total	15	15
– Pyro insertion, total	5	5
LQ insertion, total	5	6
– Sanger LQ	3	3
– Pyro LQ	2	3
HQ insertion, total	15	14
– Sanger HQ insertion	12	12
– Pyro HQ insertion	3	2
"SNP" type discrepancy	60	60
– both HQ	54	47
– LQ for Sanger	5	12
– LQ for Pyro	1	1
– both LQ	0	0
Discrepancy rate (%)	0.0409	0.0359
Accuracy rate (%)	99.9591	99.9641

6. IMPACT OF NEW SEQUENCING TECHNOLOGIES

The genome sequences of the *O. sativa* and its progenitors are expected to show the extensive base substitutions and rearrangements. Therefore, it would be difficult to reconstruct the genome sequences of wild rice relatives from cultivars. As resequencing with the conventional Sanger methodology can take much time and effort, a new pyrosequencing technology was developed. Massively parallel short reads from pyrosequencing analysis [67] could sequence more than 20 million bases with much less cost and less time than with Sanger analysis. In collaboration with 454 Life Sciences and Roche Diagnostics, we compared pyrosequencer and Sanger sequence data. Eight BAC clones which include OR_CBa0076I05, OR_CBa0091G05, OR_CBa0094N06, OR_CBa0004O24, OR_CBa0063M01, OR_CBa0075G04, OR_CBa0034E23, and OR_CBa0010H05 from *O. rufipogon* IRGC105491 (AA species) were chosen from a fingerprint contig of the OMAP BAC library (OR_CBa-FPC contig 51). This contig corresponds to an 800-kb region of the short arm of Nipponbare chromosome 6 and is expected to contain two genes for rice flowering (*Hd3a* and *RFT1*). DNA of each BAC clone was purified individually and then mixed for pyrosequencing on a GS20 genome analyzer (Roche). The output from this analysis (ca. 20× coverage) contained 286639 reads. Of these, 169130 reads were mapped and 16123462 bases were aligned to the corresponding Nipponbare sequences, forming 1422 mapped contigs that cover 57.5% of the entire genomic region. The average depth was 23.39 showing deep coverage.

To compare these sequences with those from Sanger sequencing, we shotgun sequenced a BAC clone OR_CBa0004O24 and assembled it with phred/phrap software to form contigs. Each contig sequence from pyrosequencing was aligned to its corresponding Sanger sequence by BLAST alignment. Statistical results from this comparison are shown in Table 2.

Comparing only high quality (HQ sequence quality score > either 30 or 40) nucleotides gave an overall error rate of 0.0409% or 0.0359%. This means that the high-coverage reads from pyrosequencing show more than 99.95% accuracy. Researchers have pointed out that pyrosequencing is more problematic in repeats and homopolymers than Sanger technology [68, 69], but we did not observe this type of discrepancy. We also compared nucleotide sequences of *Hd3a* (one of the rice heading date QTL, corresponding to FT gene of Arabidopsis) between *O. sativa* cv. Nipponbare (by Sanger method) and *O. rufipogon* (by pyrosequencing). Only 3 SNPs and no in/del were found in exons (540 coding nt), whereas many deviations (20 SNPs, 6 indels) were found in introns; this was evolutionally reasonable. This sequence conservation might indicate that *Hd3a* is functionally important and under purifying selection.

These results show that emerging new resequencing technologies (not only pyrosequencing but also other methods [70]), when properly used in combination with current

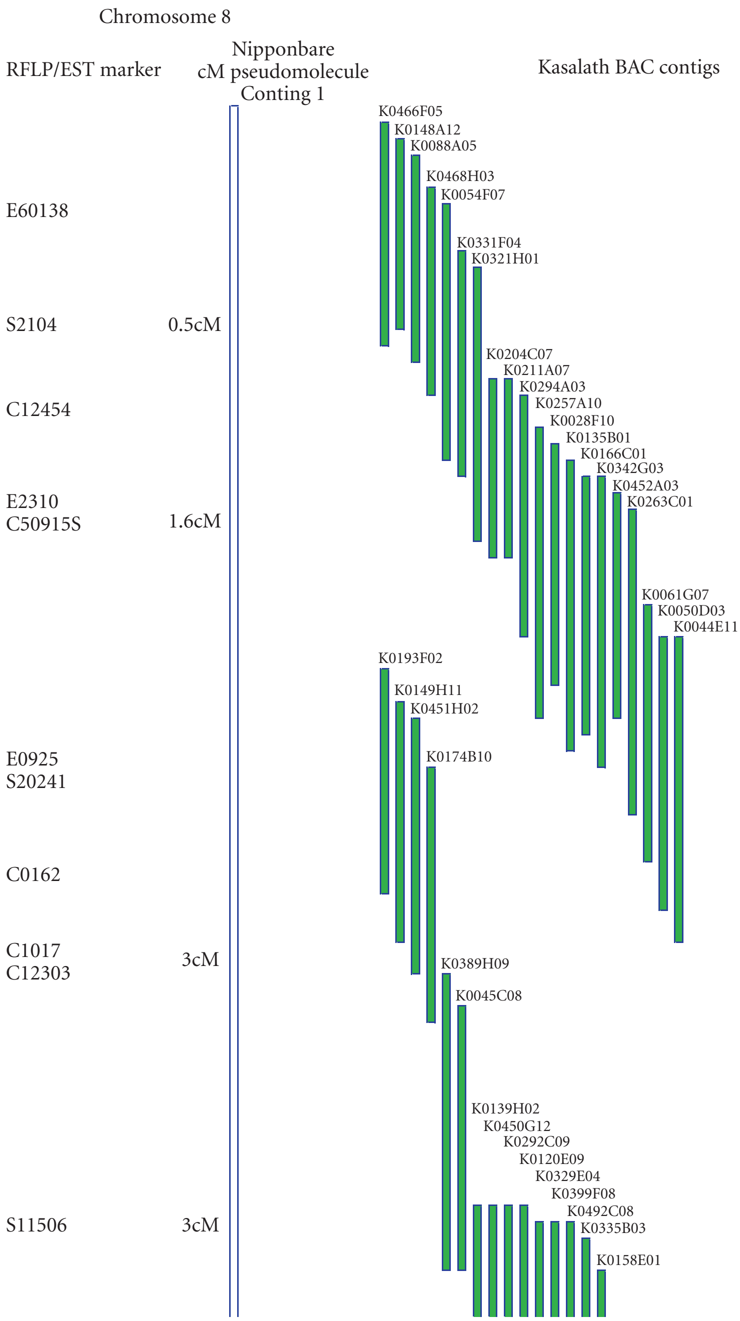


Figure 4: In silico physical map of Kasalath chromosome 8, based on the Nipponbare sequence. Green vertical bars indicate BAC clones (with K numbers) mapped against Nipponbare genome sequence (shown at left with landmarks).

methods, will revolutionize the cost and performance of rice genome resequencing and will help elucidate the evolution of the *Oryza* genomes.

7. CONCLUSION

The rice genome sequence has become available as a reference genome, providing a basis for understanding the wide range of diversity among cultivated and wild relatives of rice. The continuous efforts in generating a high-quality sequence have paved the way for clarifying the structures of genomic regions that are difficult to analyze, including centromeres and telomeres. Comparative genomics within the genus *Oryza* has also become a feasible strategy for understanding the evolutionary events that led to the development of cultivated rice. The syntenic relationships among cereal crops must be thoroughly exploited from now on. The rice genome sequence will be the most important tool in explaining the structure and function of other cereal genomes, and its use may open new opportunities for researchers to look deeper into the synteny between rice and other cereal crops, which has been maintained for some 60 million years of evolution [6]. From a more practical aspect, the rice genome sequence could be the key for developing rice-genomics-based research in order to improve crop production and food security for humankind.

ACKNOWLEDGMENTS

The authors thank all the participants of the Rice Genome Research Program (RGP) and the International Rice Genome Sequencing Project (IRGSP). This work was supported by grants from the Ministry of Agriculture, Forestry, and Fisheries of Japan (MAFF) through the Rice Genome Project, Green Technology Project, and GD 2007.

REFERENCES

[1] Director-General's message and FAO, http://www.fao.org/wfd2007/wfd_resources/dg_message.html.
[2] Faostat, http://faostat.fao.org/site/339/default.aspx.
[3] U.S. Census Bureau, http://www.census.gov/ipc/www/idb/worldpopinfo.html.
[4] M. D. Gale and K. M. Devos, "Comparative genetics in the grasses," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, no. 5, pp. 1971–1974, 1998.
[5] K. M. Devos and M. D. Gale, "Genome relationships: the grass model in current research," *The Plant Cell*, vol. 12, no. 5, pp. 637–646, 2000.
[6] K. M. Devos, "Updating the 'crop circle'," *Current Opinion in Plant Biology*, vol. 8, no. 2, pp. 155–162, 2005.
[7] M. E. Sorrells, "Cereal genomics research in the post-genomics era," in *Cereal Genomics*, P. K. Gupta and R. K. Varshney, Eds., chapter 19, pp. 559–584, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2004.
[8] R. Cooke, B. Piègu, O. Panaud, et al., "From rice to other cereals: comparative genomics," in *Rice Functional Genomics*, N. M. Upadhyaya, Ed., chapter 17, pp. 429–464, Springer, New York, NY, USA, 2007.
[9] The Arabidopsis Genome Initiative, "Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*," *Nature*, vol. 408, no. 6814, pp. 796–815, 2000.
[10] International Rice Genome Sequencing Project, "The map-based sequence of the rice genome," *Nature*, vol. 436, no. 7052, pp. 793–800, 2005.
[11] J. Wu, T. Maehara, T. Shimokawa, et al., "A comprehensive rice transcript map containing 6591 expressed sequence tag sites," *The Plant Cell*, vol. 14, no. 3, pp. 525–535, 2002.
[12] Y. Harushima, M. Yano, A. Shomura, et al., "A high-density rice genetic linkage map with 2275 markers using a single F_2 population," *Genetics*, vol. 148, no. 1, pp. 479–494, 1998.
[13] C. Soderlund, I. Longden, and R. Mott, "FPC: a system for building contigs from restriction fingerprinted clones," *Computer Applications in the Biosciences*, vol. 13, no. 5, pp. 523–535, 1997.
[14] M. Chen, G. Presting, W. B. Barbazuk, et al., "An integrated physical and genetic map of the rice genome," *The Plant Cell*, vol. 14, no. 3, pp. 537–545, 2002.
[15] B. Ewing and P. Green, "Base-calling of automated sequencer traces using phred. II. Error probabilities," *Genome Research*, vol. 8, no. 3, pp. 186–194, 1998.
[16] IRGSP, 2002 and http://rgp.dna.affrc.go.jp/E/IRGSP/Dec18_NEWS.html.
[17] J. C. Venter, H. O. Smith, and L. Hood, "A new strategy for genome sequencing," *Nature*, vol. 381, no. 6581, pp. 364–366, 1996.
[18] J. C. Venter, M. D. Adams, E. W. Myers, et al., "The sequence of the human genome," *Science*, vol. 291, no. 5507, pp. 1304–1351, 2001.
[19] J. Yu, S. Hu, J. Wang, et al., "A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*)," *Science*, vol. 296, no. 5565, pp. 79–92, 2002.
[20] J. Yu, J. Wang, W. Lin, et al., "The genomes of *Oryza sativa*: a history of duplications," *PLoS Biology*, vol. 3, no. 2, p. e38, 2005.
[21] S. A. Goff, D. Ricke, T.-H. Lan, et al., "A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*)," *Science*, vol. 296, no. 5565, pp. 92–100, 2002.
[22] The Rice Full-Length cDNA Consortium, "Collection, mapping, and annotation of over 28,000 cDNA clones from *japonica* rice," *Science*, vol. 301, no. 5631, pp. 376–379, 2003.
[23] T. Matsumoto, R. A. Wing, B. Han, and T. Sasaki, "Rice genome sequence: the foundation for understanding the genetic systems," in *Rice Functional Genomics, Challenges, Progress and Prospects*, N. M. Upadhyaya, Ed., pp. 5–20, Springer, Berlin, Germany, 2007.
[24] J. Yu, P. Ni, and G. K.-S. Wong, "Comparing the whole-genome-shotgun and map-based sequences of the rice genome," *Trends in Plant Science*, vol. 11, no. 8, pp. 387–391, 2006.
[25] E. Pennisi, "Corn genomics pops wide open," *Science*, vol. 319, no. 5868, p. 1333, 2008.
[26] phtyozome, *Sorghum bicolor*, and http://www.phytozome.net/sorghum.
[27] BrachyBase, *Brachypodium distachyon*, http://www.brachybase.org/.
[28] International Barley Sequencing Consortium: *Hordeum vulgare*, http://www.public.iastate.edu/~imagefpc/IBSC%20-Webpage/IBSC%20Template-home.html.
[29] International Wheat Genome Sequencing Consortium: *Triticum. Aestivum*, http://www.wheatgenome.org/index.php.
[30] K. Sakata, Y. Nagamura, H. Numa, et al., "RiceGAAS: an automated annotation system and database for rice genome sequence," *Nucleic Acids Research*, vol. 30, no. 1, pp. 98–102, 2002.
[31] C. Burge and S. Karlin, "Prediction of complete gene structures in human genomic DNA," *Journal of Molecular Biology*, vol. 268, no. 1, pp. 78–94, 1997.
[32] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, "Basic local alignment search tool," *Journal of Molecular Biology*, vol. 215, no. 3, pp. 403–410, 1990.
[33] B. J. Haas, A. L. Delcher, S. M. Mount, et al., "Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies," *Nucleic Acids Research*, vol. 31, no. 19, pp. 5654–5666, 2003.
[34] Rice Annotation Project, "Curated genome annotation of *Oryza sativa* ssp. *japonica* and comparative genome analysis with *Arabidopsis thaliana*," *Genome Research*, vol. 17, no. 2, pp. 175–183, 2007.
[35] P. Dimitri, N. Corradini, F. Rossi, and F. Vernì, "The paradox of functional heterochromatin," *BioEssays*, vol. 27, no. 1, pp. 29–41, 2004.
[36] Z. Cheng, C. R. Buell, R. A. Wing, M. Gu, and J. Jiang, "Toward a cytological characterization of the rice genome," *Genome Research*, vol. 11, no. 12, pp. 2133–2141, 2001.
[37] K. Nagaki, Z. Cheng, S. Ouyang, et al., "Sequencing of a rice centromere uncovers active genes," *Nature Genetics*, vol. 36, no. 2, pp. 138–145, 2004.

[38] J. Wu, H. Yamagata, M. Hayashi-Tsugane, et al., "Composition and structure of the centromeric region of rice chromosome 8," *The Plant Cell*, vol. 16, no. 4, pp. 967–976, 2004.

[39] Y. Zhang, Y. Huang, L. Zhang, et al., "Structural features of the rice chromosome 4 centromere," *Nucleic Acids Research*, vol. 32, no. 6, pp. 2023–2030, 2004.

[40] H. Yan, H. Ito, K. Nobuta, et al., "Genomic and genetic characterization of rice *Cen3* reveals extensive transcription and evolutionary implications of a complex centromere," *The Plant Cell*, vol. 18, no. 9, pp. 2123–2133, 2006.

[41] Z. Cheng, F. Dong, T. Langdon, et al., "Functional rice centromeres are marked by a satellite repeat and a centromere-specific retrotransposon," *The Plant Cell*, vol. 14, no. 8, pp. 1691–1704, 2002.

[42] H.-R. Lee, W. Zhang, T. Langdon, et al., "Chromatin immunoprecipitation cloning reveals rapid evolutionary patterns of centromeric DNA in *Oryza* species," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 33, pp. 11793–11798, 2005.

[43] H. Yan and J. Jiang, "Rice as a model for centromere and heterochromatin research," *Chromosome Research*, vol. 15, no. 1, pp. 77–84, 2007.

[44] J. Ma, R. A. Wing, J. L. Bennetzen, and S. A. Jackson, "Evolutionary history and positional shift of a rice centromere," *Genetics*, vol. 177, no. 2, pp. 1217–1220, 2007.

[45] A. Sharma and G. G. Presting, "Centromeric retrotransposon lineages predate the maize/rice divergence and differ in abundance and activity," *Molecular Genetics and Genomics*, vol. 279, no. 2, pp. 133–147, 2008.

[46] H. Mizuno, K. Ito, J. Wu, et al., "Identification and mapping of expressed genes, simple sequence repeats and transposable elements in centromeric regions of rice chromosomes," *DNA Research*, vol. 13, no. 6, pp. 267–274, 2006.

[47] J. Lingner and T. R. Cech, "Telomerase and chromosome end maintenance," *Current Opinion in Genetics & Development*, vol. 8, no. 2, pp. 226–232, 1998.

[48] R. K. Moyzis, J. M. Buckingham, L. S. Cram, et al., "A highly conserved repetitive DNA sequence, $(TTAGGG)_n$, present at the telomeres of human chromosomes," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 85, no. 18, pp. 6622–6626, 1988.

[49] C. W. Greider and E. H. Blackburn, "Identification of a specific telomere terminal transferase activity in tetrahymena extracts," *Cell*, vol. 43, no. 2, part 1, pp. 405–413, 1985.

[50] E. J. Richards and F. M. Ausubel, "Isolation of a higher eukaryotic telomere from *Arabidopsis thaliana*," *Cell*, vol. 53, no. 1, pp. 127–136, 1988.

[51] H. Mizuno, J. Wu, H. Kanamori, et al., "Sequencing and characterization of telomere and subtelomere regions on rice chromosomes 1S, 2S, 2L, 6L, 7S, 7L and 8S," *The Plant Journal*, vol. 46, no. 2, pp. 206–217, 2006.

[52] J. Fajkus, A. Kovařík, R. Královics, and M. Bezděk, "Organization of telomeric and subtelomeric chromatin in the higher plant *Nicotiana tabacum*," *Molecular and General Genetics*, vol. 247, no. 5, pp. 633–638, 1995.

[53] H. Kotani, T. Hosouchi, and H. Tsuruoka, "Structural analysis and complete physical map of *Arabidopsis thaliana* chromosome 5 including centromeric and telomeric regions," *DNA Research*, vol. 6, no. 6, pp. 381–386, 1999.

[54] M. Fujisawa, H. Yamagata, K. Kamiya, et al., "Sequence comparison of distal and proximal ribosomal DNA arrays in rice (*Oryza sativa* L.) chromosome 9S and analysis of their flanking regions," *Theoretical and Applied Genetics*, vol. 113, no. 3, pp. 419–428, 2006.

[55] K. Heller-Uszynska, W. Schnippenkoetter, and A. Kilian, "Cloning and characterization of rice (*Oryza sativa* L) telomerase reverse transcriptase, which reveals complex splicing patterns," *The Plant Journal*, vol. 31, no. 1, pp. 75–86, 2002.

[56] C. Vitte, T. Ishii, F. Lamy, D. Brar, and O. Panaud, "Genomic paleontology provides evidence for two distinct origins of Asian rice (*Oryza sativa* L.)," *Molecular Genetics and Genomics*, vol. 272, no. 5, pp. 504–511, 2004.

[57] J. Ma and J. L. Bennetzen, "Rapid recent growth and divergence of rice nuclear genomes," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 34, pp. 12404–12410, 2004.

[58] S. Katagiri, J. Wu, Y. Ito, et al., "End sequencing and chromosomal *in silico* mapping of BAC clones derived from an *indica* rice cultivar, Kasalath," *Breeding Science*, vol. 54, no. 3, pp. 273–279, 2004.

[59] C. Li, Y. Zhang, K. Ying, X. Liang, and B. Han, "Sequence variations of simple sequence repeats on chromosome-4 in two subspecies of the Asian cultivated rice," *Theoretical and Applied Genetics*, vol. 108, no. 3, pp. 392–400, 2004.

[60] F. A. Feltus, J. Wan, S. R. Schulze, J. C. Estill, N. Jiang, and A. H. Paterson, "An SNP resource for rice genetics and breeding based on subspecies *indica* and *japonica* genome alignments," *Genome Research*, vol. 14, no. 9, pp. 1812–1819, 2004.

[61] C. Cheng, R. Motohashi, S. Tsuchimoto, Y. Fukuta, H. Ohtsubo, and E. Ohtsubo, "Polyphyletic origin of cultivated rice: based on the interspersion pattern of SINEs," *Molecular Biology and Evolution*, vol. 20, no. 1, pp. 67–75, 2003.

[62] S. Kawakami, K. Ebana, T. Nishikawa, Y. Sato, D. A. Vaughan, and K. Kadowaki, "Genetic variation in the chloroplast genome suggests multiple domestication of cultivated Asian rice (*Oryza sativa* L.)," *Genome*, vol. 50, no. 2, pp. 180–187, 2007.

[63] J. P. Londo, Y.-C. Chiang, K.-H. Hung, T.-Y. Chiang, and B. A. Schaal, "Phylogeography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice, *Oryza sativa*," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 25, pp. 9578–9583, 2006.

[64] D. A. Vaughan, H. Morishima, and K. Kadowaki, "Diversity in the *Oryza* genus," *Current Opinion in Plant Biology*, vol. 6, no. 2, pp. 139–146, 2003.

[65] J. S. S. Ammiraju, M. Luo, J. L. Goicoechea, et al., "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 Research*, vol. 16, no. 1, pp. 140–147, 2006.

[66] H. Kim, B. Hurwitz, Y. Yu, et al., "Construction, alignment and analysis of twelve framework physical maps that represent the ten genome types of the genus *Oryza*," *Genome Biology*, vol. 9, no. 2, article R45, 2008.

[67] M. Margulies, M. Egholm, W. E. Altman, et al., "Genome sequencing in microfabricated high-density picolitre reactors," *Nature*, vol. 437, no. 7057, pp. 376–380, 2005.

[68] T. Wicker, E. Schlagenhauf, A. Graner, T. J. Close, B. Keller, and N. Stein, "454 sequencing put to the test using the complex genome of barley," *BMC Genomics*, vol. 7, article 275, 2006.

[69] M. J. Moore, A. Dhingra, P. S. Soltis, et al., "Rapid and accurate pyrosequencing of angiosperm plastid genomes," *BMC Plant Biology*, vol. 6, article 17, 2006.

[70] S. T. Bennett, C. Barnes, A. Cox, L. Davies, and C. Brown, "Toward the $1000 human genome," *Pharmacogenomics*, vol. 6, no. 4, pp. 373–382, 2005.

Hindawi Publishing Corporation
International Journal of Plant Genomics
Volume 2008, Article ID 524847, 25 pages
doi:10.1155/2008/524847

Review Article

Rice Molecular Breeding Laboratories in the Genomics Era: Current Status and Future Considerations

Bert C. Y. Collard,[1] Casiana M. Vera Cruz,[2] Kenneth L. McNally,[2] Parminder S. Virk,[2] and David J. Mackill[2]

[1] *Hermitage Research Station, Queensland Department of Primary Industries & Fisheries, 604 Yangan Road, Warwick, Queensland 4370, Australia*
[2] *International Rice Research Institute, DAPO Box 7777, Metro Manila, Philippines*

Correspondence should be addressed to Casiana M. Vera Cruz, c.veracruz@cgiar.org

Received 22 November 2007; Accepted 15 March 2008

Recommended by Yunbi Xu

Using DNA markers in plant breeding with marker-assisted selection (MAS) could greatly improve the precision and efficiency of selection, leading to the accelerated development of new crop varieties. The numerous examples of MAS in rice have prompted many breeding institutes to establish molecular breeding labs. The last decade has produced an enormous amount of genomics research in rice, including the identification of thousands of QTLs for agronomically important traits, the generation of large amounts of gene expression data, and cloning and characterization of new genes, including the detection of single nucleotide polymorphisms. The pinnacle of genomics research has been the completion and annotation of genome sequences for *indica* and *japonica* rice. This information—coupled with the development of new genotyping methodologies and platforms, and the development of bioinformatics databases and software tools—provides even more exciting opportunities for rice molecular breeding in the 21st century. However, the great challenge for molecular breeders is to apply genomics data in actual breeding programs. Here, we review the current status of MAS in rice, current genomics projects and promising new genotyping methodologies, and evaluate the probable impact of genomics research. We also identify critical research areas to "bridge the application gap" between QTL identification and applied breeding that need to be addressed to realize the full potential of MAS, and propose ideas and guidelines for establishing rice molecular breeding labs in the postgenome sequence era to integrate molecular breeding within the context of overall rice breeding and research programs.

1. INTRODUCTION

Rice (*Oryza sativa*) is the well-known holder of two important titles: the most important food crop in the world and a model cereal species. Rice is the staple food in many parts of the world, including many developing countries in Asia, Africa, and Latin America. The projected increase in global population to 9 billion by 2050 and predicted increases in water scarcity, decreases in arable land, the constant battle against new emerging pathogens and pests, and possible adverse effects from climate change will present great challenges for rice breeders and agricultural scientists [1–4]. Because of rice's global importance, small genome size, and genetic relatedness to other major cereals, efforts were undertaken to sequence the entire genomes of the two subspecies of rice—*indica* and *japonica*. Genome sequence drafts were completed for both subspecies in 2002 [5, 6] and a high-quality and annotated version of the *japonica* species was completed in 2005 [7], which represent landmark achievements in biological research.

One practical output from genomics research was the development of DNA markers (or molecular markers) in the late 1980s and 1990s. Marker-assisted selection (MAS)—in which DNA markers are used to infer phenotypic or genotypic data for breeding material—is widely accepted to have great potential to improve the efficiency and precision of conventional plant breeding, which may ultimately lead to the accelerated release of new crop varieties [8–13]. The potential advantages of molecular breeding demonstrated by numerous examples of MAS in rice and other crops have prompted many rice breeding and research institutes to establish biotechnology or DNA marker labs within the institute.

Genomics is the study of gene location, function, and expression. Strictly speaking, the study of gene location might be classified as molecular genetics research. However, for simplicity, we broadly define genomics as the study of genes and genomes, which includes identifying the location of genes as well as the study of gene function and regulation (expression). The beginning of the 21st century has been considered the dawn of the genomics era due to the enormous amount of genomics research in bacterial, plant, and animal species, as well as the rapid development of high-throughput equipment for whole-genome genotyping, gene expression, and genome characterization, and the establishment of advanced bioinformatics tools and databases. These rapid developments have irreversibly influenced and redefined plant breeding in the 21st century as "molecular plant breeding" or "genomics-assisted breeding" [14].

However, plant breeders and agricultural scientists face many challenges to integrate and exploit these new molecular and genomics-related technologies for more rapid and efficient variety development [15, 16]. In this article, we review the current global rice molecular breeding lab with an emphasis on recent research and the impact of rice genomics resources. We also review some current genomics research and promising new genotyping methodologies with high potential for applied outcomes. Finally, we consider the obstacles to the successful application of molecular genetics and genomics research in rice breeding programs and propose ideas on how some of these problems should be solved.

2. THE RICE MOLECULAR BREEDING LAB

2.1. View of the rice "pregenome sequence" molecular breeding lab

We arbitrarily define the "pregenome sequence molecular breeding lab" as before 2000. Although the first rice genome sequence drafts were published in 2002 and the complete genome sequence was published in 2005, sequence data were available before these publication dates so it is very difficult to exactly pinpoint the time when rice genome sequence data influenced applied rice genetics and breeding. In the early to mid-1990s, restriction fragment length polymorphism (RFLP) and random amplified polymorphic (RAPD) markers were commonly used for rice breeding research [17–21]. In Japan, RFLPs continue to be a marker system of choice [22]. Often, RFLP and RAPD markers were converted into second generation, polymerase chain reaction (PCR)-based markers called sequence tagged site (STS) markers to improve technical simplicity and reliability [23–25]. Simple sequence repeats (SSR; or "microsatellites") became the most widely used markers in cereals and rice is no exception [26–28]. In earlier reports, the principles and techniques of detecting SSR polymorphisms were called simple sequence length polymorphism (SSLP) markers [28, 29]. SSRs are highly reliable (i.e., reproducible), codominant in inheritance, highly polymorphic (compared to other markers), and generally transferable between mapping populations. The only disadvantages of SSRs are that they typically require polyacrylamide gel electrophoresis and generally give information only about a single locus per assay.

The first SSRs were reported in 1996 [30]. By 1997, there were 121 validated SSRs, which were adequate for marker-assisted evaluation of germplasm and the construction of framework linkage maps but had limited use for MAS, due to limited genome coverage [29]. By 2001, there were a total of ~500 SSRs that were developed from 57.8 Mb of publicly available rice genome data [31], which further increased the utility of these markers.

2.2. The postgenome sequence rice molecular breeding lab: opening the "treasure chest" of new rice markers

2.2.1. SSRs

Analysis of the completed rice genome sequence provided the identification of literally tens of thousands of new targets for DNA markers, especially SSRs. Using publicly available BAC and PAC clones, more than 2200 validated SSRs were released in 2002 [32]. This was soon followed by 18828 Class I (di-, tri-, tetra-repeats) SSRs that were released after the completion of the Nipponbare genome sequence in 2005 [7]. This number is by far the largest number of publicly available SSRs for any crop species. The extremely high density of SSRs (approx. 51 SSRs per Mb) will provide a considerable "tool kit" for map construction and MAS for numerous applications. Given that many labs are currently well equipped for SSR analysis, it is highly likely that SSRs will continue to be the marker of choice for years to come.

2.2.2. Single nucleotide polymorphisms (SNPs)

SNPs are the most abundant and ubiquitous type of polymorphisms in all organisms, and many researchers propose that these markers will be the marker of choice in the future [33]. In rice, SNPs can be readily identified by direct comparisons of Nipponbare and 93-11 genomes, or by sequence alignment with one or both reference sequences with available sequence data in public databases [34–36]. Recently, more SNP data have become available that have been generated by comparing partial sequences from multiple genotypes [37–39]. In some cases, DNA sequencing of target regions in specific genotypes is required. However, experimental validation of SNP-based markers is required since inaccuracies in sequence data have been reported [34, 36]. The ease with which SNPs can be identified in silico and increase in publicly available rice DNA sequence data will undoubtedly ensure that SNP-based markers will be more commonly used in the future.

It should be noted that lower levels of SNP marker polymorphism are usually detected in more closely related genotypes, which are more representative of breeders' elite germplasm (*indica* × *indica* or *japonica* × *japonica*-derived material), when compared with the *japonica-indica* reference genotypes used to determine SNP frequency. The frequency of SNPs between subspecies was reported to be from 0.68% to 0.70%, whereas it was 0.03% to 0.05% between *japonica* cultivars and 0.49% between *indic* a cultivars [35]. Interestingly, SNPs were not evenly distributed along chromosomes.

2.2.3. Indels

Insertion/deletion (indel) mutations are abundant mutations that occur in coding and noncoding regions. Indels can also be quickly identified in silico by direct comparions of *japonica* and *indica* genome sequences. The enormous number of indels between the two subspecies will provide an indispensable resource of polymorphic markers for *indica* × *japonica* populations or populations with specific introgressions [34, 35]. Either of the two rice genome reference sequences can be easily compared with other sequences for further indel identification, as was done between Nipponbare and Kasalath, a commonly used *indica* accession [40]. Like SNP-based markers, indels also need to be experimentally validated.

Introns are noncoding regions within genes and hence they "tolerate" insertion/deletion mutations compared with exons. Consequently, many indels have been identified in introns and these size polymorphisms have been exploited by the development of a new class of intron length polymorphic (ILP) markers [41]. Experimental validation of these markers indicated that the majority were reliable and codominant, and that although ILPs were designed from *indica/japonica* comparisons, they were also polymorphic between varieties within both subspecies although the level of polymorphism was lower.

2.2.4. "Custom-made" markers

The great resource for molecular breeders is the DNA sequence provided by the genome sequences since it permits markers that are tightly linked to target loci to be "custom-made" or "tailor-made" to suit the aims of MAS. The large number of custom-made markers that have already been designed or the potential for new ones to be designed is a unique feature of the rice molecular breeding lab. The number of markers that can potentially be generated using the rice genome sequence in silico is practically unlimited (Figure 1). The markers might be derived directly from the Nipponbare/93-11 sequences or used to identify corresponding EST or genomic sequences available from databases (i.e., BAC or PAC clones containing target genes that may not actually be present in reference genotypes) [42–44]. In principle, custom-made markers can be any type although they most commonly include new SSRs, indels, PCR-based SNPs, and cleaved amplified polymorphic site (CAPS) markers—which are the technically the simplest markers to be used for marker genotyping [43, 45]. It should be noted that these markers must be tested in wet-lab experiments.

Candidate gene (CG) identification can be integrated with customized marker design and development. The advantage of CG-derived markers is that they are usually more tightly linked to the gene or QTL controlling the trait. This approach has been successfully used for identifying CGs associated with disease resistance, since cloned plant disease resistance genes possess conserved domains [46, 47].

2.3. Protocols, resources, and laboratory organization

Since marker genotyping methods were first developed in the 1980s, numerous protocols and variations now exist. Many protocols have been refined and optimized specifically for the lab in which marker genotyping is conducted and will depend on budget, equipment, and personnel. One feature of rice molecular breeding labs is their diversity. Molecular breeding labs require a large initial capital investment and since many labs are based in developing countries, the equipment and resources often differ markedly from those of well-funded labs in developed countries. The cost of marker genotyping is, therefore, a critical factor for the extent of MAS in rice, and this is likely to continue to be the case for years to come given the unlikely dramatic decrease in costs.

2.3.1. DNA extraction protocols

Many general DNA extraction methods that are used in diverse plant species have been used in rice, from which it is relatively easy to extract DNA (see, e.g., [48–50]). Some methods have been specifically developed for rice [51]. The DNA extraction component is often the most time-consuming and laborious step of marker genotyping. For this reason, high-throughput methods using 96-well PCR plates have been developed [52]. The method by Xu et al. [52] does not require liquid nitrogen or freeze drying for initial grinding of leaf tissue or the use of organic solvents.

Alternative "quick and dirty" methods for DNA extractions in rice were evaluated and optimized at IRRI [53]. These methods were selected from published papers in the literature based on the time and resources required for using the protocols, as well as cost, and optimized for routine use. Two methods were selected as being the best when considering success of PCR amplification of SSRs, time, and cost [51, 54]. The modified method by Wang et al. [54] greatly reduced the time and cost for routine DNA extractions and was adapted into a 96-well plate method.

2.3.2. SSR genotyping

SSR genotyping typically requires high-resolution electrophoresis, which is performed using polyacrylamide gels or, in some cases, high-resolution agarose. The majority of labs use standard gel electrophoresis equipment and stain gels with DNA-binding stains such as ethidium bromide, safer analogs, or silver staining (for acrylamide gels only). Multiplexing refers to the combination of primer pairs in PCR (multiplex PCR) or samples during gel electrophoresis (multiplex gel loading) [55]. This has considerable potential for increasing the efficiency of marker genotyping due to savings in time and resources. Multiplex loading is simpler, since there are fewer variables and it has been successfully

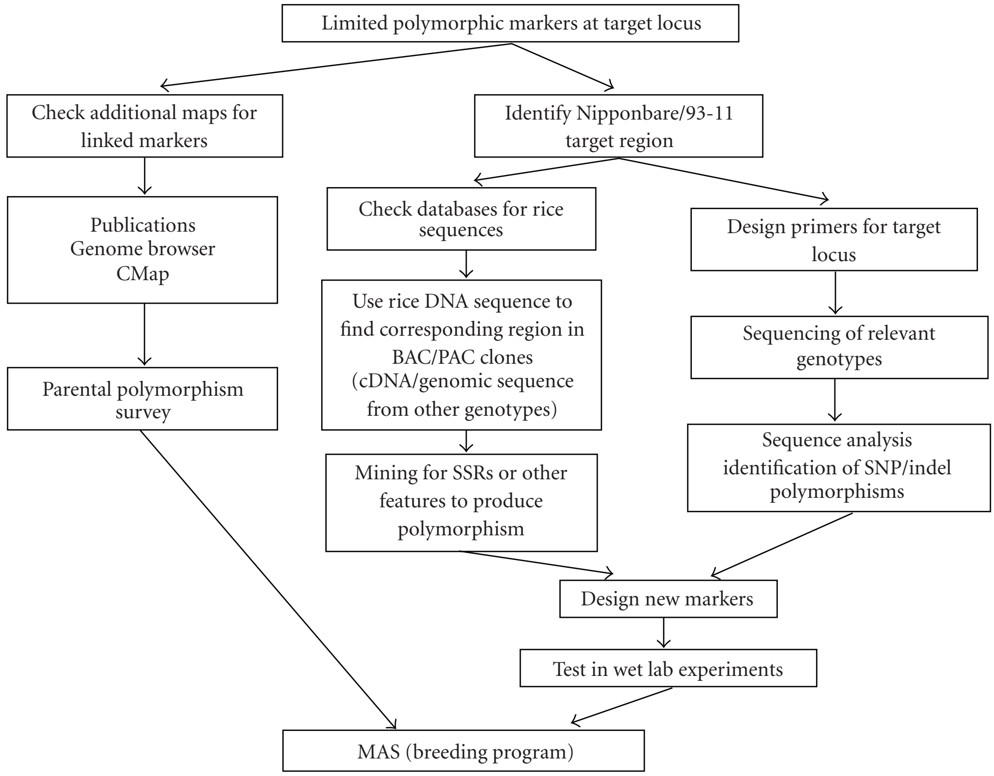


Figure 1: Process for developing custom-made rice DNA markers.

demonstrated to greatly increase genotyping efficiency [23, 28, 29].

In some labs, capillary electrophoresis systems have been established. The accuracy of marker allele determination is one of the major advantages of these platforms, since size differences of 1 bp can be discerned. Multiplex loading can also be relatively easily performed using these genotyping platforms, which use fluorescently-labeled primers in PCR [56, 57]. These platforms can also be used for DNA sequencing, highlighting their versatility. Unfortunately, the cost of consumables, the initial expense of capital equipment purchase, and possibly the reliable acquisition of consumables and technical servicing may restrict their wider-scale adoption in actual breeding stations.

2.3.3. SNP genotyping

The two simplest and most widely used methods for detecting SNP markers are PCR-based SNPs (that target SNPs by primer design) and restriction digestion of PCR amplicons, which are called cleaved amplified polymorphic site (CAPS) markers [43–45]. Komori and Nitta also used a variant of the CAPS method called derived CAPS (dCAPS), in which artificial restriction digestion sites are created in PCR amplicons. All methods use standard lab equipment [58].

Capillary electrophoresis platforms can also be used for SNP detection, based on the principle of single nucleotide primer extension (SNuPE; [35]). The high resolution of capillary electrophoresis equipment also permits small indels (say, <3 bp that are too small to be resolved on standard agarose or to be detected with acrylamide). A codominant single nucleotide length polymorphism marker (i.e., 1 bp indel) was developed from the intron region of the *Pi-ta* gene by Jiang et al. [59].

2.3.4. Indel genotyping

One attractive feature of many indels, including ILPs, is that standard agarose electrophoresis or acrylamide gel electrophesis equipment and methods used for SSR detection can be used [41]. Another attractive feature of indels that are located within genic regions is that they are gene-specific markers, so the possibility of recombination between marker and gene is eliminated.

2.3.5. Data management

It is important that molecular breeding labs have a system in place to store marker data, since they are an extremely useful resource for future breeding research. There is not a universal method for data storage—systems range from in-house Excel files to sophisticated laboratory information management systems (LIMS). We have found that standard database software is adequate for marker data storage. The development of template files and standard operating procedures for all researchers to use is more important. This information can be exploited for future genotyping activities.

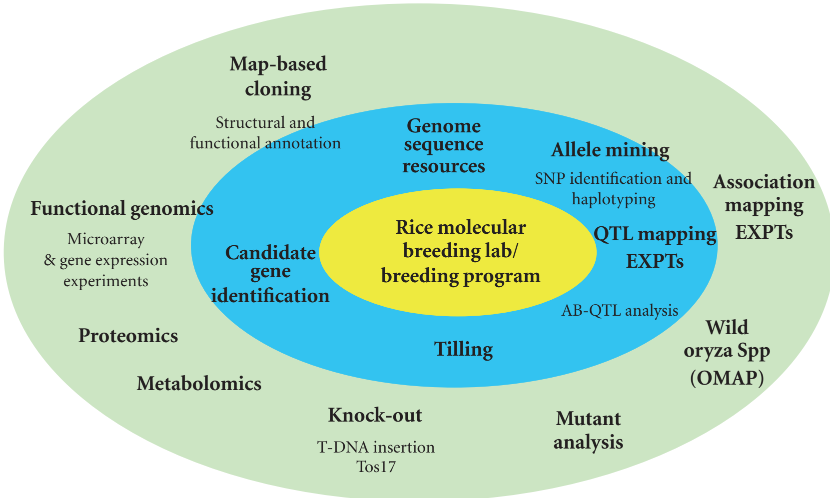


Figure 2: "Impact circle" overview of genomics and molecular research in rice. Research areas have been placed in either the inner or outer circle. Inner circle research activities are considered to provide more direct applied benefits (in terms of new markers, information for the development of new markers, or new breeding lines) to the rice molecular breeding lab/breeding program, which is located in the center. Research areas indicated in the outer circle are generally considered to provide indirect benefits to the rice molecular breeding lab/breeding program.

Careful data collation is essential to ensure that parental genotyping is not unnecessarily repeated and to determine opportunities for multiplexing.

For labs that generate large amounts of genotypic data, a more formal LIMS could be appropriated (see Figure 2); some of these systems have been recently developed for general crop species [60, 61].

2.3.6. *Rice molecular breeding Internet resources*

Markers and maps

The Internet has become a vital and convenient repository for marker and map data, and the rice molecular breeder must become familiar with these resources. There are excellent resources for published rice DNA markers that are maintained at the Gramene website [62, 63] http://www.gramene.org/ (these resources are the envy of other cereal researchers!). These web resources can be used for many applications, including obtaining SSR primer sequences, marker allele size data, and the map position of markers. Gel photos on a reference set of rice genotypes can also be obtained from this link. A large repository of published linkage maps, genes, QTLs, mutants, and references can also be searched in Gramene. The comparative map viewer (CMap) can be used to visually compare maps side by side [64].

The integrated rice genome explorer (INE; http://rgp.dna.affrc.go.jp/giot/INE.html) was developed to provide quick and simple correlations between genetic markers and EST, and physical maps with the rice genome sequence [65] are another excellent resource. These features can be viewed rapidly in the database.

"Genome browsers": the genome sequence resource for searching

The completed rice genome sequence map would be of limited use if it was not easy to search. For this purpose, user-friendly "genome browsers" (Gbrowse) have been developed. The Institute for Genomics Research (TIGR) Gbrowse resource (http://www.tigr.org/tdb/e2k1/osa1/) was designed for scientists to data-mine the rice genome [66, 67]. The rice genome sequence has been organized into "pseudomolecules" which are virtual contigs of the 12 rice chromosomes. Each gene has been designated with a locus identifier that enables specific points of reference to be identified within the pseudomolecule. This resource consists of annotated genes, identified motifs/domains within the predicted genes, a rice repeat database, identified related sequences in other plant species, and identified syntenic sequences between rice and other cereals. The TIGR Gbrowse enables structural and functional annotations to be quickly viewed. The latest version of the rice genome browser supports "tracks," which allow users to view specific features such as markers and putative genes within defined regions. Enhanced data access is available through web interfaces, FTP downloads, and a data extractor tool [68].

More recently, a genome browser was established within Gramene that enables the Nipponbare genome sequence to be quickly searched. This sequence is linked to genetic linkage maps in the Gramene database. Genome browsers are extremely user-friendly resources for assisting with basic and applied research.

2.4. *Marker-assisted selection (MAS) in rice*

MAS is the process of using DNA markers to assist in the selection of plant breeding material [11, 12, 69, 70]. Collard and Mackill [8] described three fundamental advantages of MAS compared with conventional phenotypic screening.

(i) It is generally simpler than phenotypic screening, which could save time, effort, resources, and, for some traits, money. Furthermore, MAS screening is nondestructive.

(ii) Selection can be carried out at any growth stage. Therefore, breeding lines can be screened as seedlings and undesirable plant genotypes can be quickly eliminated. This may be useful for many traits but especially for the traits that are expressed at specific developmental stages.

(iii) Single plants can be selected and their precise genotype can be determined which permits early generation selection in breeding schemes. For most traits, homozygous and heterozygous plants cannot be identified by conventional phenotypic screening. Using conventional screening methods for many traits, single-plant selection is often unreliable due to environmental effects, which can be variable.

One of the most important ways in which these advantages can be exploited by breeding programs is the more precise and efficient development of breeding lines during frequently-used breeding methods such as backcrossing, bulk, and pedigree methods [9, 13]. Target genotypes can be more effectively selected, which may enable certain traits to be "fast-tracked," potentially, resulting in quicker variety release. Markers can also be used as a replacement for phenotyping, which allows selection in off-season nurseries, making it more cost effective to grow more generations per year or to reduce the number of breeding lines that need to be tested, by the elimination of undesirable lines at early generations [13]. MAS has numerous applications in rice (Table 1). Some MAS applications represent activities that are impossible using conventional breeding methods (e.g., marker-assisted backcrossing and pyramiding). Collard and Mackill [8] emphasized the importance of exploiting the advantages of marker-assisted breeding over conventional breeding in order to maximize the impact on crop improvement.

2.4.1. Genotype identity testing

DNA markers can be used to simply and quickly identify varieties—or confirm the identity of a varietal impostor. For simple F_1 hybrids, codominant markers can be used to determine whether putative hybrids are genuine. Multiple F_1s can also easily be screened and desirable genotypes can be selected.

Seed purity or intra-variety variation can easily be tested using markers. This can be more accurate than phenotypic evaluation [71]. For the testing of hybrid rice lines, using STS and SSR markers was considerably easier than using typical "grow-out tests" that involve growing plants to maturity and evaluating purity based on morphological and floral characteristics [70, 72]. SSRs from mitochondrial genes have been targeted for the development of markers to study maternally inherited traits such as cytoplasmic male sterility or the maternal origin of rice accessions [73]. It has often been determined that relatively few well-chosen markers can provide sufficient data for varietal discrimination.

2.4.2. Genetic diversity analysis of breeding material

There have been numerous research papers on the assessment of genetic diversity in specific germplasm collections using different types of markers [74, 75]. However, in recent years, SSRs have become the marker of choice for this application (see Table 1). An example was the use of SSR markers to broaden the genetic base of U.S. rice varieties [76]. DNA markers have also been used in hybrid rice breeding in order to predict genotypes that combine to give superior hybrid vigor [77].

2.4.3. Gene surveys in parental material

The accurate evaluation of genes in breeders' germplasm is of great importance for the selection of parental lines and development of new breeding populations. Having gene information for specific target loci (deduced from markers) can be extremely useful for breeders to efficiently use germplasm. An example of this was demonstrated by Wang et al. [80, 114], who used a set of dominant allele-specific markers for surveying markers to detect the presence of the *Pi-ta* resistance gene for rice blast in a large germplasm collection ($n = 141$).

2.4.4. Marker-evaluated selection (MES)

This novel approach was used to identify genomic regions under selection (i.e., allelic shifts) of breeding populations using a modified bulk-population breeding system in target environments [99]. This approach makes no prior assumptions about traits for selection; however, selection is imposed in target environments. High-density or whole-genome marker coverage is an important prerequisite for MES. Theoretically, once specific alleles or genomic regions have been identified to be under selection, they can be combined via MAS to develop new breeding lines that are the "ideotypes" (i.e., ideal genotypes).

2.4.5. Marker-assisted backcrossing (MABC)

MABC is the process of using markers to select for target loci, minimize the length of the donor segment containing a target locus, and/or accelerate the recovery of the recurrent parent genome during backcrossing [115, 116]. These three levels of selection have been referred to as foreground, recombinant, and background selection, respectively [8]. Terms were described after Hospital and Charcosset [116], who referred to foreground selection as the selection of a target locus and background selection as the selection of the recurrent parent genome using markers on noncarrier chromosomes and also on the carrier chromosome. MABC is superior to conventional backcrossing in precision and efficiency. Background selection can greatly accelerate

Table 1: Examples of marker-assisted selection in rice. na = not applicable.

Application	Traits or germplasm	Gene/QTLs	Markers used	Reference
Early generation selection	Bacterial blight	*Xa21*	STS	[78]
Gene surveys in parental material	Blast disease, predom. Korean germplasm	*Pi5(t)*	PCR/DNA gel blot	[79]
Gene surveys in parental material	Blast disease	*Pi-z*	SSR	[42]
Gene surveys in parental material	Blast disease	*Pi-ta*	Gene-specific marker	[80]
Genetic diversity assessment	Japonica varieties for hybrid combinations	na	SSR and RAPD	[77]
Genetic diversity assessment	Indian aromatic and quality rice	na	SSR	[74]
Genetic diversity assessment	U.S. varieties	na	SSR	[76]
Genetic diversity assessment	Nepalese landraces	na	SSRs	[81]
Genetic diversity assessment	Representative wild rice in China, *Oryza rufipogon* Griff.	na	SSR	[82]
Genetic diversity assessment	Indonesian varieties and landraces	na	SSR	[83]
Genotype identity testing	Hybrid rice	na	STS and SSR	[70, 72]
MABC	Bacterial blight	*Xa21*	STS and RFLP	[84]
MABC	Bacterial blight	*Xa21*	STS and AFLP	[85]
MABC	Bacterial blight	*xa5*	STS	[86]
MABC	Deep roots	QTLs on chromosomes 1, 2, 7, and 9	RFLP and SSR	[87]
MABC	Bacterial blight	*xa5, xa13, Xa21*	STS	[88, 89]
MABC	Blast	*Pi1*	SSR and ISSR	[90]
MABC	Quality	*Waxy*	RFLP and AFLP	[91]
MABC	Bacterial blight + quality	*xa13, Xa21*	STS, SSR, and AFLP	[92]
MABC	Submergence tolerance, disease resistance, quality	*Subchr9* QTL, *Xa21*, *Bph* and blast QTLs, and quality loci	SSR and STS	[93]
MABC	Blast disease	na	SSR	[94]
MABC	Root traits and aroma	QTLs on chromosomes 2, 7, 8, 9, and 11	RFLP and SSR	[95, 96]
MABC	Heading date	QTLs for heading date (*Hd1*, *Hd4*, *Hd5*, or *Hd6*)	RFLP, STS, SSR, CAPS, dCAPs	[97]
MABC	Submergence tolerance	*Sub1* QTL	SSR	[98]
MES	Indirect selection for adaptation	na	SSRs	[99]
Pyramiding	Bacterial blight	*Xa4, xa5, Xa10*	RFLP and RAPD	[100]
Pyramiding	Bacterial blight	*xa5, xa13, Xa4, Xa21*	RFLP, STS	[17]
Pyramiding	Blast disease	*Pi1, Piz-5, Pi2, Pita*	RFLP, STS	[101]
Pyramiding	Bacterial blight	*xa5, xa13, Xa21*	STS and CAPS	[102]
Pyramiding	Bacterial blight	*xa5, xa13, Xa21*	SSR and STS	[103]
Pyramiding	Bacterial blight and waxy genes	*xa5, xa13, Xa21, Wx*	SSR, STS, and CAPS	[104]

Table 1: Continued.

Application	Traits or germplasm	Gene/QTLs	Markers used	Reference
Pyramiding	Insect resistance and bacterial blight	*Xa21* and *Bt*	STS	[59]
Pyramiding	Brown plant-hopper	*Bph1* and *Bph2*	STS	[105]
Pyramiding	Thermosensitive genetic male sterility (TGMS) genes	*tms2, tgms, tms5*	SSR	[106]
Pyramiding	Bacterial blight	*Xa7* and *Xa21*	STS	[107]
Pyramiding	Bacterial blight	*Xa4, xa5,* and *Xa21*	STS	[108]
Pyramiding	Bacterial blight	*Xa4, Xa7,* and *Xa21*	STS	[109, 110]
Pyramiding/transgene selection	Blast and bacterial blight	*Pi-z and Xa21*	STS	[111]
Pyramiding/transgene selection	Bacterial blight	*xa5, xa13, Xa21*	STS (check)	[112]
Pyramiding/transgene selection	Bacterial blight, yellow stem borer, sheath blight	*Xa21, Bt RC7* chitinase gene, *Bt*	STS	[113]

a backcrossing program compared to using conventional backcrossing [117]. Furthermore, recombinant selection can minimize the size of the donor chromosome segment, thus reducing "linkage drag"—a "universal enemy" of the plant breeder [115]. This approach has been widely used and, due to the prevalence of several rice "mega varieties," it is likely to continue being a successful approach [118].

For basic research applications, the MABC approach can be used to develop near-isogenic lines (NILs) with far greater precision than conventional backcrossing. Near-isogenic lines are valuable tools to characterize individual genes or QTLs. However, in many situations, NILs produced, using conventional backcrossing possess, many unknown donor introgressions on noncarrier chromosomes (i.e., chromosomes without target genes) and large donor chromosomal segments on the carrier chromosome. By using an MABC approach, NILs could be developed to ensure that lines are not influenced by "background" donor introgression and possess minimal donor segments flanking the target locus. We propose that NILs developed using such approaches are referred to as "precision introgression lines" (PILs). Ideally, markers with known map or physical positions should be used for PIL development.

2.4.6. Pyramiding

Pyramiding is the process of combining genes or QTLs in progeny usually arising from different parents [101, 119]. Using conventional methods, this is extremely difficult or impossible to do in early generations (e.g., F_2 or F_3) because single plants need to be screened for multiple diseases or pathogen races. Because of the importance of blast and bacterial blight, many pyramiding efforts have been directed toward breeding for resistance to these two diseases (Table 1). There is strong evidence that combining resistance genes may provide broad-spectrum resistance [88, 120–123].

Although widely used for combining disease resistance genes or QTLs, pyramiding can be used for other abiotic stress tolerance and agronomic traits. An example of pyramiding agronomic genes was the combination of three thermosensitive genetic male sterility genes [106].

2.4.7. Using transgenes

There has been much research in developing transgenic rice lines for basic and applied research applications [124]. MAS is traditionally used to screen for transformants for the transgene(s) [111]. However, with the availability of transgenics in rice for several useful traits such as resistan ce to diseases (bacterial blight, blast, sheath blight, yellow mottle virus), resistan ce to insects (stem borer, leaffolders), resistan ce to herbicide, tolerance of abiotic stress (drought, salt), nutritional traits (iron and pro-vitamin A), and photosynthetic traits [125, 126]; there is a strong interest in using transgenes in breeding. Rice breeders are excited to transfer them to successful mega varieties through conventional backcrossing or MABC. For example, transgenic rice (southern U.S. japonica-type varieties) with inherent ability to produce beta-carotene developed by Syngenta is available at IRRI and in several other national programs. However, these cultivars are not adapted to the tropical conditions in Asia, where most consumers prefer indica-type rice varieties. Therefore, at IRRI, we are introgressing the beta-carotene loci from japonica-type donor varieties into popular indica-type Asian rice varieties, using MABC. Initially, we used 3 GR1 events (GR1-146, GR1-309, and GR1-652) as donor parents, while 2 IRRI-bred mega varieties (IR64 and IR36) and a popular Bangladeshi variety (BR29) were used as recurrent parents. Subsequently, we received 6 GR2 events (GR2-E, GR2-G, GR2-L, GR2-R, GR2-T, and GR2-W). Four indica varieties, IR64, IR36, BR29, and PSB Rc 82, were used as recurrent parents. Advanced backcross progenies are available and some are ready for field testing [127–129].

3. CURRENT GENOMICS RESEARCH AND PROMISING NEW GENOTYPING METHODS

Attendance at the most recent international rice genetics conference held in Manila, Philippines (2005), indicated a mind-boggling amount of current research activities in rice genetics and genomics. These developments have been outlined in general and specific review articles (see, e.g., the excellent reviews [14, 121, 130–134]). In this section, we provide a brief overview of some of these research areas, with a focus on selected current genomics research projects that in our opinion are directed toward tangible applied molecular breeding outcomes. We also review some potentially useful and recently developed genotyping methods that could be used in breeding programs.

3.1. A brief overview of recent rice functional genomics research and annotation of the rice genome

Although the DNA sequences for Nipponbare and 93-11 are complete, rice genome sequence resources are constantly being revised and updated in terms of gene annotation [67]. There are two levels of annotation: structural annotation which refers to gene identification based on ESTs and full-length cDNA (FL-cDNA) sequences, and functional annotation which refers to the determination of gene function [132, 135]. The generation of EST libraries and FL-cDNA libraries has occurred simultaneously with genome sequencing for both *japonica* and *indica* subspecies [136, 137].

Since the actual function of the vast majority of genes remains unknown, functional annotation relies primarily on bioinformatics evidence to assign gene function [138]. To systematically and efficiently annotate the rice genome, an automated system and database called rice genome automated annotation system (RiceGAAS) was developed. This system automatically searches for rice genome sequences from GenBank, and processes them based on gene prediction and homology search programs for structural annotation. To facilitate the efficient management and retrieval of data for rice genome annotation, annotation databases such as the rice annotation project database (RAP-DB) [139] were developed.

Research in plant functional genomics provides useful data for functional annotation [140]. Reverse genetics approaches (studying the effect of gene alterations on phenotype) such as generating specific gene knockouts by RNA interference (RNAi), transfer-DNA (T-DNA), and transposon-mediated (*Ac*, *Ds*, *Ac/Ds,* and *Tos17*), and chemical/irradiated mutants have been successfully used to elucidate gene functions and determine tissue- or organ-specific gene expression (by using reporter genes) [141–146]. There are literally hundreds of thousands of mutant lines, albeit only a very small number of genotypes produced by basic research labs around the world can be screened for specific genes. Data generated by reverse genetics studies are publicly available and have been stored in curated databases such as the International Rice Information System (IRIS) [147], OryzaGenesDB [148], and EU-OSTID [149] for greater dissemination to the wider scientific community.

Microarrays have been widely adopted by plant scientists to study gene function. In rice, microarrays have been used to study processes related to yield (e.g., grain filling) and response to biotic and abiotic stresses [150–154]. Many databases have been developed to store gene expression data (reviewed in [135]). Most microarray studies have used gene-specific probes to detect gene expression and, hence, new "tiling microarrays" may study whole-genome expression, which is more informative because it is less biased [155, 156].

Although, to date, progress has been limited in rice, proteomics research also offers great promise for determining gene functions [157, 158]. In the future, it is hoped that a complete integration with proteomics and metabolomics will provide the ultimate data to elucidate not only individual gene functions but also complex pathways [135].

The generation of a deluge of genomics data has been accompanied by several integrative bioinformatics tools and databases. One notable example is called "Rice PIPELINE" which was developed for the collection and compilation of genomics data, including genome sequences, full-length cDNAs, gene expression profiles, mutant lines, and *cis* elements from various databases [159]. Rice PIPELINE can be searched by clone sequence, clone name, GenBank accession number, or keyword. Another web-based database system, called "PlantQTL-GE," was developed to facilitate quantitative traits locus (QTL)-based candidate gene identification and gene function analysis [160]. This database integrated marker data and gene expression data generated from microarray experiments and ESTs from rice and *Arabidopsis thaliana*. Specific QTL marker intervals or genomic regions can be targeted for candidate gene analysis, which could be useful for identifying new candidate genes. Both databases are publicly available.

3.2. Current applied genomics research highlights

3.2.1. Association of candidate defense genes with quantitative resistance to rice blast: a case study

The candidate gene approach has been used to integrate the molecular analysis of host-pathogen interactions, gene mapping, and disease resistance in rice. Candidate genes are similar to known genes or conserved motifs that make it possible to infer their biological functions [161]. Through their association with disease resistance, they become candidate defense response (DR) genes [122, 162, 163]. Advanced backcross lines of Vandana × Moroberekan, a japonica cultivar from Africa exhibiting durable quantitative resistance to blast in Asia, were used to demonstrate this approach for blast resistance. To accumulate different genes with quantitative resistance to blast, 15 BC_3F_5 lines of Vandana × Moroberekan showing partial resistance at IRRI and Cavinti, Philippines, and carrying DR candidate alleles were selected and crossed in all pairwise combinations. Plant selections based on blast resistance and agronomic acceptability were made in F_2 and F_3 populations, and the top 60 F_5 selections were evaluated in multilocation environments.

To identify DR candidate genes in the progenies, molecular analyses of rice genes involved in quantitative resistance were done in selected F_4 lines, using STS markers derived from rice candidate gene sequences and SSR markers located in the region of each candidate gene BAC clone showing polymorphisms between Vandana, Moroberekan, and their progenies. A total of 11 candidate genes were identified based on converging evidence (i.e., mapping, phenotyping, selection, microarray analysis) and used in this study. These candidate genes with known biological functions were oxalate oxidase/germin-like proteins, aspartyl protease (Esi-18), 14-3-3 proteins, PR-1, PBZ (PR10A), rice peroxidase (POX 22.3), heat shock protein (HSP90), putative 2-dehydro-3-deoxyphosphoheptonate aldolase, thaumatin-like pathogenesis-related protein, glyoxylase 1 (*Oryza sativa*), and S-adenosyl L-homocystein hydrolase. DR candidate genes were examined using in silico analysis of their sequences retrieved from the Rice Genome Program database. For genes occurring in gene families such as oxalate oxidase belonging to germin-like proteins, phylogenetic trees using the retrieved sequences were constructed to determine their relatedness and groups. The conserved promoter motifs were also compared and *cis*-elements in the 1000-bp upstream regions were identified. For each gene, there was variation in the copy number of *cis*-elements related to biotic stress responses, such as W box, WNPR1, and WRKY. This study suggested that these genes have potential associations with the response of rice to pathogen infection such as the blast fungus *Magnaporthe oryzae*.

3.2.2. Identification of SNP by Eco-TILLING at specific candidate genes

TILLING or "targeting induced local lesions in genomes" is a reverse genetics technique developed to identify variation in *Arabidopsis* mutant libraries obtained from chemical mutagenesis with EMS [164, 165]. The approach involves creating pools of mutant lines followed by amplification with differentially labeled, locus-specific primers on these pools. If a pool contains a mutant variant, then denaturation/renaturation of the PCR products will allow heteroduplex mismatch molecules to be formed. Treatment of the products with the single-strand-specific endonuclease CEL1 will cleave a mismatch site and generate fragments that on separation and visualization by fluorescence will indicate the position of the mutation in the amplicon. Eco-TILLING is the application of this technique to discover allelic variation in natural populations. TILLING is accomplished using pools of mutant library lines having a majority of the wild-type allele at a given locus while Eco-TILLING contrasts a reference line, such as the source of the sequence with a single diverse germplasm accession. The main requirement for both TILLING and Eco-TILLING is sufficient sequence information for the design of locus-specific primers. Hence, SNP discovery and genotyping can proceed without the need for de novo sequencing, a requirement of other SNP genotyping tools prior to assay design.

At IRRI, we have designed locus-specific primers for a range of candidate genes putatively involved in drought, general stress response, and grain quality is leveraging the high-quality sequence information for the japonica-type Nipponbare [7]. Candidate genes were identified using convergent information taking into account genome annotation, involvement of the ortholog in another species, expression data, and colocalization with QTLs. Candidate genes for drought include DREB2a, ERF3, sucrose synthase, actin depolymerizing factor, and trehalose-6-phosphate phosphatase, among others. We have conducted Eco-TILLING at these candidate genes using a diverse collection of 1536 *O. sativa* accessions from the international Genebank collection contrasted to both japonica-type Nippponbare and indica-type IR64. Depending on the contrast, from 4 to 9 haplotypes have been discovered in about 1 kb at the candidate gene locus. Representative types for the haplotype mismatch patterns have been sequenced, and association tests with phenotypic data for vegetative-stage drought characters are under way. We have also optimized a procedure that allows TILLING/Eco-TILLING products to be detected on agarose gels, thus eliminating the need for fluorescent labeling and the use of an automated genotyper, with savings in both time and costs [166]. This simplified procedure is now our method of choice and its application to breeding will be described later.

3.2.3. Genome-wide SNP discovery in diverse rice germplasm

The availability of the high-quality sequence of Nipponbare provides the unprecedented opportunity for genome-wide SNP discovery and improving our knowledge about allelic diversity in rice. IRRI along with partners in the International Rice Functional Genomics Consortium has undertaken a project to identify genome-wide SNP in a diverse collection of 20 varieties [167] with funding from IRRI, the Generation Challenge Program, and USDA-CSREES. The diverse varieties include representatives from all variety groups—temperate and tropical japonica, aromatic, aus, deep-water, and indica types—with Nipponbare included as a control. The technology being used for SNP discovery is hybridization to very high-density oligomer arrays pioneered by Perlegen Sciences, Inc. (Mountain View, Ca 94043, USA). On these arrays, four 25-mer oligomer features are tiled for each of the strands, where the middle base is present as A, T, C, or G for the four features with a single base offset occurring before the next set of features. Hence, 8 oligomer features interrogate each base of the sequence of the target genome during hybridization. Application of Perlegen's technology has led to the identification of large sets of SNPs for human [168], mouse [169], and *Arabidopsis* [170].

Funding was available for SNP discovery in 100 Mb of the rice genomes. Consequently, only the nonrepetitive regions of the Nipponbare genome were selected for tiling onto high-density oligomer arrays. However, the nonrepetitive regions span the entire genome with the majority of 100 kb windows containing several or more tiled regions. Following hybridization of the query genomes to arrays, about 260000 nonredundant SNPs were identified by Perlegen's model-based algorithms. Efforts are ongoing to extend this collection by applying the machine-learning-based techniques developed for the analysis of the Arabidopsis project [170].

The set of Perlegen model-based SNPs provides about 93% genome coverage by the criterion that at least 1 SNP occurs per 100 kb of the genome. Since existing estimates of linkage disequilibrium (LD) in rice indicate that LD extends to 100 kb or longer [171, 172], then the SNP dataset should be sufficient for identifying a collection of tag SNPs that define haplotype blocks across the rice genome. This set of tag SNPs can then be used to undertake whole-genome scans in a wider collection of rice varieties, with the resulting genotypic data applied to association studies with detailed phenotypes for traits of interest.

3.2.4. Exploiting wild species

Landraces and wild species of rice (genus *Oryza*) possess an underused source of novel alleles that have great potential for crop improvement of cultivated rice species (*O. sativa* and *O. glaberrima*), since they possess new genes that could be exploited for yield increases and for developing resistance to biotic stresses and tolerance of abiotic stresses [173, 174]. Consequently, many experiments have attempted to use wild sources to develop new breeding material and also characterize genes and QTLs from these sources. The advanced backcross QTL analysis (AB-QTL) approach—which is a method for integrating QTL mapping with simultaneous line development—has been widely used to introgress wild genes and QTLs into adapted varieties with great success for agronomic traits and yield (reviewed in [174]).

Introgression lines (ILs) are derived by generating backcross lines using MAS with relatively large, different donor chromosomal segments from wild or exotic genotypes [119, 175]. ILs are useful for many applications in genetic analysis (e.g., high-resolution mapping of QTL regions), since phenotypic evaluation can be performed over multiple years and environments. In a study analyzing ILs developed from *Oryza rufipogon* in an *indica* background (Teqing), many putative QTLs for yield and yield components were detected [176].

Genome sequence research using wild species is well under way. The *Oryza* Map Alignment Project (OMAP) was initiated to construct physical maps (derived from BAC clones) of 11 wild and 1 cultivated species (*O. glaberrima*) and align them to the Nipponbare reference genome sequence [177, 178]. Advanced backcross populations (BC_4F_2) of 3 OMAP wild accessions are also being generated for mapping important traits. Apart from providing insights into evolution of the *Oryza* genus, other expected outcomes are the identification of new genes and QTLs that could be subsequently incorporated into adapted rice varieties.

3.2.5. Association mapping

Despite the widespread use and success of QTL mapping for identifying QTLs that control traits, the method has inherent limitations [179, 180]. In practice, mapping populations are derived from bi-parental crosses that represent only a small fraction of the total allelic variation, and QTL mapping experiments may require a large investment in resources. Association mapping—based on linkage disequilibrium—may bypass these limitations of QTL mapping because a greater number of alleles are analyzed and historic phenotypic data for multiple traits can be readily used without the need for a specific evaluation of populations generated solely for the purposes of QTL mapping [181, 182]. Furthermore, association mapping can offer improvements in resolution because analysis is based on the accumulation of all meioses events throughout the breeding history.

Linkage disequilibrium has been estimated in rice to be approximately 100 to 250 kbp based on the characterization of two genes, *xa5* (chromosome 5) and *Waxy* (chromosome 6) [171, 172]. A more recent study indicated that the extent of LD was much larger: 20–30 cM [183]. The former estimate suggests that high-density whole-genome scans are required for efficient association mapping in rice. An alternative approach would be to focus on regions previously delimited by QTL analysis or regions in combination with candidate gene analysis.

Several recent studies have investigated "population structure" in rice, which is important for controlling the false discovery rate [83, 184, 185]. Various methods of data analysis have been evaluated. An example was the use of discriminant analysis involving markers associated with previous QTLs [186]. Discriminant analysis results were consistent with previous QTL results, although additional markers, not identified by QTL mapping methods, were detected which may indicate new loci associated with specific traits.

The "foundation" of previously identified QTLs for numerous traits, the availability of candidate genes from genomics research, and further improvements in statistical methodology [184] are likely to ensure that more rice researchers use association mapping approaches in the future.

3.3. Recent and new marker genotyping methods

3.3.1. Optimizing and refining current protocols

One very important point we would like to emphasize before reviewing new technology is that there are great opportunities for further optimization of currently used protocols, especially in terms of cost and throughput. Furthermore, many innovations on standard methods are possible (see, e.g., [187]). This is important because many labs have already made a considerable investment in lab equipment and have the technical expertise to use specific protocols using specific markers.

As discussed earlier, multiplexing has considerable potential for increasing the efficiency of marker genotyping although this has not been extensively explored in rice. Multiplex PCR could be complicated since numerous variables (primer combination, annealing time and temperature, extension time and temperature, and concentrations of primers and magnesium chloride) are involved [188, 189]. However, in many cases, the investment in time and resources may be justified. Coburn et al. [57] reported 80% successful PCR amplification for duplex PCR. Multiplex loading is simpler and in our opinion could be applied on a much wider scale. At IRRI, loading of two or even three markers (A. Das, pers. comm.) is frequently possible, which saves time and resources. Of course, information regarding marker allele sizes is a prerequisite for multiplex loading.

3.3.2. *Considering the adoption of new genotyping methods*

Many new promising genotyping methods could improve efficiency in terms of time and potential cost [190]. Most of these methods are targeted toward SNPs but most of them could be adapted for other marker types. Interestingly, there are many high-throughput SNP genotyping platforms (that have often been developed for medical applications), yet there has been no universally adopted system [191, 192].

In the context of plant breeding, there are several important considerations. Cost is critical due to the large number of samples breeders evaluate. Furthermore, 3 to 6 target traits usually segregate in a single population so the frequency of lines with all the desirable gene combinations is very low. This could undermine the suitability of some high-throughput whole-genome profiling programs, although there could be numerous applications in basic research.

Obviously, some genotyping methods will be more suitable for specific labs than others. For this reason, we have classified these methods into two groups: regional hub labs and remote breeding stations. A regional hub lab is defined as a research institute with a critical mass of scientists who receive sufficient funding for long-term, broad objective breeding research that includes genomics research (e.g., CGIAR centers and national breeding institutes). We refer to a remote breeding station as a "smaller" lab that has more limited capacity for marker genotyping in terms of funding and resources.

3.3.3. *Remote breeding station lab 1: gel-based methods*

PCR-based SNP methods

PCR-based SNP detection methods that use standard agarose or acrylamide electrophoresis are obviously attractive because they are technically simple and no further investment in equipment is required. The simplest form of PCR-based SNP marker is based on designing PCR primers such that a forward or reverse primer has a specific dNTP at the 3′ end; PCR amplification is successful for the appropriate primer-template combination and fails when the specific 3′ base in the primer is not complementary to the template [43]. Reliability has been an important issue with designing PCR-based SNP markers; hence, several studies, exploring methods to improve reliability including the use of additional primers, have been conducted [193–195]. Hayashi et al. [43] introduced an artificial mismatch at the 3rd base from the 3′ end—in addition to the last 3′ base—which was found to increase specificity; a 67% success rate was found for 49 target SNPs. This method can be used to develop codominant allele-specific markers. Overall, these methods are useful to complement the arsenal of CAPS markers for whichtarget SNP-containing sites are not available.

Heteroduplex cleavage SNP detection methods

TILLING and Eco-TILLING methods (discussed previously) are reverse genetics methods used to identify SNPs in target genes in mutants and germplasm collections, respectively. However, simplified TILLING/Eco-TILLING methods, using standard polyacrylamide or agarose gel electrophoresis detection methods, could be applied for MAS and would be especially useful in situations, where it is difficult to find other types of polymorphic markers [166, 196]. This method relies on the principle that CEL I cleaves heteroduplexes at the position of SNPs.

In brief, the method involves the following steps.

(i) PCR amplification of the region of interest in parental lines (A and B) (homozygous).

(ii) The PCR products are combined in equal concentration and subjected to CEL I digestion (TILLING/Eco-TILLING) in an agarose procedure to test for polymorphism.

(iii) DNA is extracted from each member of the breeding population (RIL-homozygous) and quantified.

(iv) DNA extracted from either of the parental lines (e.g., parent A) is combined with DNA from each of the RILs in a 1 : 1 ratio.

(v) The mix is subjected to CEL I digestion. If an SNP is detected, this indicates that the allele carried by the RIL is unlike that of the parent used to create the mix (in this case, parent A).

One possible limitation of this procedure is that it would be ideally done on homozygous lines. If there is doubt, the assay should be conducted with just the DNA from each of the RILs; no SNPs should be detected.

PCR-RF-SSCP

polymerase chain reaction- (PCR-) restriction fragment- (RF-) single-strand conformation polymorphism (SSCP)—abbreviated to PRS—is essentially based on a combination of the CAPS technique (i.e., restriction digestion of gene-specific PCR products) with SSCP, which on its own can be used for SNP detection of small PCR amplicons (100–400 bp) using polyacrylamide gel electrophoresis (PAGE) [197–199]. This method has been successfully used to detect SNPs in rice and other crops. One of the advantages of this method is that much longer PCR amplicons (>2000 bp) can be scanned for SNPs, and it may be well suited for labs with technical expertise in polyacrylamide gel electrophoresis and/or silver staining.

3.3.4. Remote breeding station lab 2: non-gel-based methods

Dot blots

Dot blots have been used for genotyping of rice breeding material [200]. The main advantages of this methodare that gel electrophoresis and even PCR in some cases are not required. This method used cultivar-specific sequences that were previously identified by AFLP, STS, or PRS. Genomic DNA from rice samples was spotted on membranes and short oligonucelotide (28–45 bp) or digoxigenin (DIG)-labeled PCR products (102–466 bp) were used as probes. DIG labeling methods avoid the use of radioisotopes, which is preferable in most labs and very important for remote breeding stations due to delivery, storage, and disposal. Relatively high DNA yields were required for this method (3.5–5 μg).

The dot blot genotyping method was later extended to a robust SNP detection [201]. In this method, two nucleotide probes (17 nt) were used: one allele-specific probe was DIG-labeled (at the 5′ end) and the other allele probe was unlabeled, following the principles of competitive allele-specific short oligonucleotide hybridization, which improves specificity. The probe targets were PCR products that contained the SNP regions. This method has potential for high-throughput capacity since 864 samples were blotted on a single membrane. Dot blot genotyping has been used for high-throughput, large-scale MAS in commercial companies [202]. Dot-blot assay was used in advanced Basmati-derived lines that have reached the replicated yield trial at IRRI's breeding program (Reveche et al., unpublished data). This method, however, is not yet in routine use but offers great potential for MAS in breeding program.

3.3.5. Regional hub lab

Capillary electrophoresis platforms for SSR genotyping

To maximize the efficiency of multiplexing using capillary electrophoresis platforms, marker "panels" can be assembled, which consist of markers with no overlapping allele size ranges or the same fluorescent dyes [56, 203]. In general, panels of any size and for any traits can be designed based on available primer resources and previously determined allele sizes. Coburn et al. [57] reported assembling panels consisting of 6 to 11 SSRs that were evenly spaced along all 12 chromosomes; most panels were designed such that they are chromosome-specific. A greater flexibility of panel design was demonstrated in maize, in which primers were redesigned for specific SSR loci from sequence data [114]. This permitted a tenplex level of multiplexing (i.e., scoring of 10 individual SSR marker alleles in a single gel lane). Although these panels from these two examples were designed for whole-genome scans, they have wider potential in routine MAS. Furthermore, generic fluorescently labeling primer methods, which greatly reduce costs, are other innovative methods by which the cost efficiency of capillary electrophoresis methods can be improved [204, 205]. In our opinion, it would also be feasible to adopt capillary electrophoresis systems in some remote breeding stations.

SNuPE

Many SNP detection methods are based on the commonly used principle of single nucleotide primer extension (also called single base extension, SBE). Briefly, this method works by using a genotyping primer that immediately precedes an SNP at the 3′ end in the template. This genotyping primer is extended with a specific fluorescently labeled dideoxy nucleotide (ddNTP) that is detected, which permits genotyping at a target locus. SNuPE can be performed using capillary electrophoresis systems, which could be very convenient if these platforms have been set up in labs for SSR genotyping. Capillary electrophoresis platforms have a very high throughput capacity: a pilot study in maize indicated that 1200 genotypes could be analyzed per day [206].

FRET-based genotyping

SNPs have become prominent in rice functional genomics research because of their advantage of being prevalent in the genome. For example, a recent study has reported an average occurrence of one SNP for every 40 kb in target regions in chromosomes 6 and 11 (S. McCouch, pers. comm.). If these SNPs are informative and exist in alternate alleles of a gene for resistance and susceptibility, for example, the *Xa21* gene for bacterial blight resistance, they would become useful candidates for marker development. At IRRI, we have adopted a method for SNP detection that uses the system known as fluorescence resonance energy transfer (FRET).

FRET is a radiation-less transmission of energy from a donor molecule to an acceptor molecule when they are in close proximity to one another (typically 10–100 Å). It has been mostly used in biomedical research and drug discovery to detect SNPs in the human genome [207, 208] and in homogeneous DNA diagnostics [209] as well as for other applications in protein interaction analysis [210]. In the conventional FRET reported by Takatsu et al. [211], the detection method requires special fluorescence-labeled probes, which are expensive and difficult to optimize. Later in the same year, Takatsu et al. [212] developed a method based on single base extension and applied SYBR Green I (bound to double-stranded DNA) as an energy donor and fluorescence-labeled ddNTP as an energy acceptor. This method avoids difficult probe design and allows a significant reduction in detection cost.

We have adapted the method for large-scale MAS in rice and further reduced the cost by optimization of expensive reagents (e.g., enzymes) during purification steps of single-stranded DNA prior to SBE. We employed the method as an SNP genotyping technique with the advantage of being high-throughput and non-gel-based. Here, the amplified genomic DNA containing the polymorphic site is incubated with a primer (designed to anneal immediately next to the polymorphic site) in the presence of DNA polymerase, SYBR Green I, and ddNTP labeled with a fluorophore (ROX or Cy5). The primer binds to the complementary site and is extended with a single ddNTP. When SYBR Green I is excited at its excitation wavelength of 495 nm, it will transfer the energy to the ddNTP at the polymorphic site next to it. High fluorescence intensity will be measured at each emission wavelength for SYBR Green I and the respective fluorophores for a resistant and susceptible allele, so SNP can be discriminated after the SBE reaction.

MICROARRAY-BASED GENOTYPING (MBG)

(A) SNP genotyping of alternate alleles

DNA microarray technology provides a snapshot of gene expression levels of all genes in an organism in a single experiment. Depending on the objective of the experiment, it allows the identification of genes that are expressed in different cell types to learn how their expression levels change in different developmental stages or disease states and to identify the cellular processes in which they participate. This technology platform has also been used in genotyping studies, such as the tagged microarray marker (TAM) approach and the high-throughput system that makes genotyping efficient and low cost [213]. An alternative and simpler microarray technique was described by Ji et al. [214]. MBG is based on simple hybridization with fluorescence-labeled probes, which anneal with specific alleles in PCR products. MBG for MAS of specific genes needs printing of PCR products derived from breeding materials on glass. The alternate probes of the gene (e.g., *xa5* gene for bacterial blight resistance) are labeled with fluorophores, such as Alexa-Fluor 546 (or Cy3) for the R allele and Alexa-Fluor 647 (or Cy5) for the S allele. MBG is useful when the number of samples increases, thus decreasing the cost per data point. In designing an experiment for marker-assisted breeding, we can save time, space, and labor by establishing computer-aided data acquisition. MBG is one of the most advanced techniques for automated data processing.

Although the use of some expensive equipment, including the arrayer and scanner, may make users think twice, the cost per sample will be remarkably lower by using less expensive supplies and reagents that are commercially available.

(B) Single-feature polymorphism (SFP)

Microarray-based genotyping that used indel polymorphisms or SFP provides the means to simultaneously screen hundreds to thousands of markers per individual. This technology is particularly suited to applications requiring whole-genome coverage, and the relatively low cost of this assay allows a genotyping strategy using large populations. Along with foreground selection for the target traits, high-resolution whole-genome selection will provide a greater capacity for background selection to retain the positive attributes of popular varieties in backcrossing programs. Obtaining graphical genotypes of individuals will facilitate the pyramiding of desirable alleles at multiple loci and will shorten the time needed for developing new varieties.

SFP assays are done by labeling genomic DNA (target) and hybridizing it to arrayed oligonucleotide probes that are complementary to indel loci. The SFPs can be discovered through sequence alignments or by hybridization of genomic DNA with whole-genome microarrays. Each SFP is scored by the presence or absence of a hybridization signal with its corresponding oligonucleotide probe on the array. Both spotted oligonucleotides [215] and Affymetrix-type arrays [216] have been used in these assays. For genotyping large populations, the cost per individual is more critical than the cost per data point. Spotted oligonucleotide microarrays have the potential to provide low-cost genotyping platforms [217]. The availability of genomic sequences from multiple accessions presents opportunities for the design of spotted long oligonucleotide microarrays for low-cost/high-density genotyping of rice.

The SFP genotyping slide for rice has been developed in the laboratory of D. Galbraith, University of Arizona, Ariz, USA [218]. Using the publicly available genomic sequences of rice cultivars Nipponbare and 93-11 representing the *japonica* and *indica* subspecies, respectively, they made alignment of these sequences and identified 1264 SFPs suitable for probe design. With a median distance between markers of 128 kb, the SFPs are evenly distributed over the whole genome. An early result using these probes showed conservatively 30–50% polymorphism between a pair of rice lines (the lowest between *japonica* types). Thus, a single contrast produces around 400 well-spaced, polymorphic gene-based markers for any pair of unrelated parental lines. One advantage of the DNA hybridization-based genotyping procedure is that it can be used for quantitative genotyping of pooled samples.

Both of these microarray-based genotyping platforms can be combined for foreground (e.g., SNP genotyping of alternate alleles) and background selection (e.g., SFPs) in breeding programs.

MALDI TOF MS

Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI TOF MS) has been used for SNP genotyping in other crops such as barley and oilseed rape [219, 220]. The principle of mass spectroscopy is based on mass-to-charge ratio rather than electrophoretic mobility. SNP genotypes can be discriminated after SNuPE and then determining the molecular weight differences for the incorporated ddNTPs. This system has potential for high-throughput genotyping in regional hub labs because of the capacity to screen large numbers, speed of genotyping (seconds compared with hours for gel-based systems), amenability to automation, and low-cost potential.

4. CRITICAL ASSESSMENT OF THE IMPACT OF GENOMICS RESEARCH

4.1. *Benefits to breeding*

To date, the outcomes from genomics research have had three main benefits to breeders: increased knowledge regarding important traits, the generation of new breeding lines, and a vast array of DNA marker tools. Genomics research outcomes will provide considerably more information on the biology of traits, especially for complex quantitative traits for which information can be very limited [14]. Improved knowledge regarding complex traits can be extremely useful for breeders. Currently, there is an enormous amount of QTL and candidate gene data for these traits that will be continually refined and validated until specific genes are identified.

In applied terms, one important tangible benefit has been the generation of new breeding lines arising from QTL mapping experiments. These lines may include the "best" lines segregating for the traits under study. Numerous introgression lines or chromosome segment substitution lines (CSSL) and NILs developed for specific traits have considerable potential for breeding programs [221, 222]. As discussed earlier, many new breeding lines with wild donor introgression are the output from AB-QTL analysis experiments [174]. For breeding programs, ILs or AB-QTL analysis lines can be rapidly converted into NILs via an MABC approach using only a small number of backcrosses.

From the molecular breeding perspective, the most tangible benefit from genomics research is the wealth of DNA markers associated with traits from previous research and the potential for generating thousands of new markers from the two rice genome sequences [7, 130]. This has already had a pronounced impact on plant breeding and thisimpact will undoubtedly continue in the future. In theory, the lack of polymorphism for target markers in breeding material should no longer be a problem as more and more "allele-specific marker kits" will be available or be custom-made, where required for an increasing number of traits. Marker kits will enable the precise selection of parental lines for the generation of new breeding populations and reliable selection of segregating progeny. As more and more genes are identified, the development of "functional markers" or "perfect" markers will be more common [4]. Since functional markers are the site that determines phenotype, they are thus the ultimate marker in a marker kit. Such markers have been used for *Xa21* with great success [108–110]. Rice functional markers were recently developed for betaine aldehyde dehydrogenase (*BAD2;* controlling fragrance) and *xa5* was developed for bacterial blight resistance [223, 224].

4.2. *Obstacles that genomics research will not solve*

4.2.1. *Cost of using DNA markers*

Despite the enormous potential for developing and using markers in rice, the cost of genotyping is still a prohibitive barrier to the wider application of MAS. Even with the global importance of rice, many developing countries have limited research and development capability. Therefore, cost optimization of current genotyping protocols and the development of new cost-effective protocols should be a major priority for breeding research and especially the rice molecular breeding lab. These improvements might involve simple optimizations of current laboratory practices, adopting new more efficient methods, or developing new MAS strategies and schemes.

Collard and Mackill [8] stated that preliminary cost analysis of MAS at IRRI indicated great potential for reduction. They stated a cost of US \$1.00 per marker data point achieved by a post-doctoral research fellow or US \$0.30 for a research technician, which we have since revised to US \$0.37. At first glance, this amount may not sound like much, but when one considers that this indicates a cost of US \$96 per plate, and that literally thousands (or even tens of thousands) of breeding lines are screened per annum in a typical rice breeding program, the importance of cost becomes obvious.

A detailed breakdown of cost components for the marker genotyping of a single SSR marker using standard methods indicated some interesting findings (Table 2).

(i) PCR costs the most in terms of consumables.

(ii) The DNA extraction step costs the most in terms of labor.

(iii) Overall, the DNA extraction step is the most expensive.

This analysis also provided a simple framework to investigate opportunities for some cost reduction (Table 2). In summary, scenarios 1 and 2 highlight that the optimization of technical procedures could decrease costs, scenario 3 highlights that the MAS scheme used will also vary costs, and scenario 4 shows that MAS lab planning and appropriate delegation of duties can also reduce costs.

Detailed cost-benefit analyses of using markers for specific traits could be critical information to determine the most appropriate and advantageous situations for using markers. For example, in maize, an extremely detailed cost-benefit analysis indicated that using markers for selection for *opaque2* (the gene associated with quality protein maize) was more economical than conventional screening methods [225]. In such cases, there is a clear-cut advantage of using markers in breeding.

4.2.2. *QTL application research: bridging the "application gap"*

Many research steps are required from QTL discovery to the practical application of markers in a breeding program [69]. The three main research areas can be described as "QTL confirmation," "broad-range QTL testing," and "marker validation," which we collectively refer to as "QTL application research." These research areas have been loosely defined as QTL or marker validation activities—especially in wheat and barley. See references cited in [8]. However, in this paper, we have specifically defined the overall research area as QTL application research and have defined three components. QTL confirmation is desirable because factors such as small population sizes and insufficient replication of trait data, and experimental errors can cause inaccuracies in determining QTL positions and effects. Broad-range QTL testing refers to verification of QTLs in different populations by using previously reported markers in order to evaluate

Table 2: Cost breakdown of standard marker genotyping and exploration of marker genotyping cost reduction opportunities.

Situation	Step	Consumables (US $)	Labor (US $)	Cost per marker (US $)
Standard cost	DNA	0.051	0.437	**1.001**
	PCR	0.211	0.076	
	Gel	0.052	0.174	
Scenario 1—multiplex loading	DNA	0.051	0.437	**0.910**
	PCR	0.211	0.076	
	Gel	0.000	0.043	
Scenario 2—multiplex PCR	DNA	0.051	0.437	**0.500**
	PCR	0.211	0.076	
	Gel	0.052	0.174	
Scenario 3—MAS pyramiding	DNA	0.051	0.437	**0.676**
	PCR 1	0.211	0.076	
	Gel 1	0.052	0.174	
	PCR 2	0.211	0.076	
	Gel 2	0.052	0.174	
	PCR 3	0.211	0.076	
	Gel 3	0.052	0.174	
Scenario 4—DNA extraction performed by research technician	DNA	0.051	0.040	**0.604**
	PCR	0.211	0.076	
	Gel	0.052	0.174	

(i) Standard cost calculated based on the genotyping of 96 samples using a single SSR marker at IRRI from Collard and Mackill [8].
(ii) Data in this section of the table are reported for the second marker; hence, the gel cost for consumables is zero. The calculation was performed using the data for a standard marker plus the second marker (gel consumable cost = 0) and dividing by two.
(iii) Multiplex loading by sequential loading of PCR samples, assuming different DNA samples are run in all lanes. Labor would require an extra 20 minutes for sequential loading, but gel preparation and assembly are no longer required in this scenario.
(iv) If direct pooling of PCR products is possible, only a single loading is required for 96 samples (extra 5 minutes of labor). Gel labor costs are reduced to US $0.011 and the total cost per marker is $0.893.
(v) Multiplex PCR (i.e., duplex PCR) in which two markers can be genotyped in the time and effort required for a single marker.
(vi) For MAS pyramiding, the genotyping of three loci was considered. For simplicity, it was assumed markers could not be multiloaded, but obviously, if this was possible, it would indicate a further cost reduction per marker screened.
(vii) The DNA extraction step is the most costly from our data analysis. There are considerable savings in expense, if genotyping efforts of a postdoctoral researcher are coordinated with those of a research technician at IRRI.

the effectiveness of the markers in predicting phenotype. This is required because of the effects of genetic background, possible epistatic interactions, and environmental effects that could ultimately reveal that QTLs may not be relevant in a specific breeding program. Marker validation activities are also required to evaluate the reliability of the markers and to identify polymorphism in relevant breeding lines. The latter two steps are also highly desirable for confirming marker-trait linkages identified by association mapping.

In practice, these research steps are often not performed and they represent an important obstacle for MAS to have an impact on crop improvement; this was referred to as the "application gap" by Collard and Mackill [8]. Although there are encouraging examples of marker validation research, there are relatively few published reports of QTL confirmation or broad-range application research in rice. A notable exception was the confirmation of QTLs for sheath blight resistance [226].

4.2.3. "Phenotype gap"

This term was used to refer to the increasing ratio of genomic sequence data to known gene phenotype [227]—the term "phenotype gap" was originally coined by mammalian researchers. As mentioned earlier, this limits the ability to functionally annotate the constantly growing amount of rice genome sequence data. In the next few decades, the lack of knowledge of gene function will exist for the vast majority of rice genes. For mutant studies, only a few selected genotypes have been used, including only a single *indica* variety. Phenotypic analysis of mutant lines represents a considerable workload [142]. Precision phenotyping is also critical to the success of QTL or association mapping experiments, but, unfortunately, the importance of refining and developing new methods for precise phenotypic measurement is also often neglected in the genomics era. Overcomingthe phenotype gap represents the next great challenge for scientists involved in rice genomics research.

5. FUTURE CONSIDERATIONS FOR INTEGRATING THE RICE MOLECULAR BREEDING LABORATORY IN THE 21ST CENTURY

5.1. *Molecular breeding lab activities*

QTL application research activities represent an extensive amount of time, effort, and resources. In practice, it seems that molecular breeders will ultimately have to perform this research in situations, in which important data for the application of MAS are not available. From experience, it is clear that breeding programs that do not undertake these activities risk wasting considerable time and resources. However, in practice, QTL application research activities may be constrained by funding, time, and resources; in some cases, these activities may be beyond the capacity of many rice molecular breeding labs. Furthermore, a breeder may decide that, based on the importance of the target trait, such QTL application research steps do not worth the investment in time, resources, and money, since, at the end, the markers may not turn out to be useful for selection in their own breeding program.

This poses a practical barrier to the application of MAS in breeding programs for which there may not be any simple solutions. One possible solution that might assist plant breeders and molecular breeders could be the formation of molecular breeding networks in which practical information and experiences are readily shared between labs regarding specific gene/QTL targets and marker information. A web-based medium such as a “wiki” or electronic *Rice Molecular Breeding Newsletter* could be extremely useful. Greater integration with research objectives among the research institutes involved in QTL mapping might also result in more relevant data being generated for breeding programs.

Many activities will occur in the future rice molecular breeding lab. Obviously, the primary objectives will be to support and assist the breeding program in the evaluation and selection of breeding material. To fulfill this duty, organizational and maintenance activities such as organizing protocols, marker data, supplies of consumables, equipment maintenance, and LIMS will be critical. In-house data records for marker optimization and parental screening will be critical; generally, the more detailed the records, the better. This must include field and glasshouse leaf tissue collection protocols, which cannot be neglected.

It also seems certain that the development of custom-made markers will become more commonplace, and so molecular breeders will need to be proficient in skills such as PCR primer design, DNA sequence analysis, and using bioinformatics databases and tools. Considerable in silico applied genomics research will occur prior to wet-lab experiments or before breeding populations are initiated. SNPs will be the inevitable polymorphism target of choice arising from current and future genomics research, so rice molecular breeders should consider this ahead of time. Molecular breeders will also need to keep in touch with current bioinformatics tools and future genomics advances.

5.2. *Integration within rice breeding programs*

The advancements in the field of molecular breeding and genomics are proceeding at such a rapid rate that it makes it difficult for molecular breeders, let alone conventional plant breeders and other agricultural scientists, to keep abreast of these new developments. Thus, when possible, plant breeding stations that intend to adopt molecular breeding approaches should establish a molecular breeding lab with designated molecular breeders and technical staff, in order to maximize the likelihood of gaining benefits from molecular breeding. There will be a critical need for molecular breeders—like conventional plant breeders—to be Jacks (or Jills) of all trades in order to integrate the disciplines. In addition to a background in applied genomics, the ideal molecular breeder should have a strong background in classical and quantitative genetics and plant breeding. Molecular breeders will need to work extremely closely with senior plant breeders for trait prioritization and devising effective MAS strategies.

Of course, establishing molecular breeding labs will not be possible in many plant breeding stations, especially in developing countries, because of limited funding and resources. However, collaboration with national or international research institutions or universities could still provide opportunities for such breeding programs to gain benefits from genomics research.

For genomics to be fully integrated into the overall breeding program, we propose that molecular breeders be actively engaged in “genomics extension activities” (analogous to “agricultural extension”) to explain and disseminate information regarding markers and advances in genomics. Appropriate activities may include training workshops and developing practical manuals, booklets, and other educational material and would address the knowledge gap between molecular biologists, plant breeders, and other disciplines [8, 69]. Such activities might also encourage a greater integration in situations, in which university research labs conducting basic research are closely connected with actual breeding stations.

6. CONCLUDING REMARKS

Breeding research in rice is poised to gain many direct and indirect benefits from genomics research. However, there are many challenges for rice scientists to fully exploit and apply knowledge, resources, and tools in actual rice breeding programs. There are great opportunities for more efficient rice breeding and the faster development of new rice varieties in the future. We hope that some of the ideas proposed in this article will encourage the rice scientific community to collectively work toward converting rice from a model crop species into a model species for marker-assisted breeding.

ABBREVIATIONS

AFLP: amplified fragment length polymorphism;
BAC: bacterial artifical chromosome;
BC: backcross;
CAPS: cleaved amplified polymorphic site;
CG: candidate gene;
EST: expressed sequence tag;
FRET: fluorescence resonance energy transfer;
IL: introgression line;
ILP: intron length polymorphism;
ISSR: Inter-simple sequence repeats;
LIMS: laboratory information management system;
MABC: marker-assisted backcrossing;
MAS: marker-assisted selection;
MES: marker-evaluated selection;
NIL: near-isogenic lines;
PAC: P1 phage artificial chromosome;
PCR: polymerase chain reaction;
QTL: quantitative trait loci;
RAPD: random amplified polymorphic DNA;
RFLP: restriction fragment length polymorphism;
SBE: single base extension;
SCAR: sequence characterized amplified region;
SFP: single-feature polymorphism;
SNP: single nucleotide polymorphism;
SNuPE: single nucleotide primer extension;
SSCP: single-strand conformation polymorphism;
SSR: simple sequence repeats (microsatellites);
STS: sequence tagged site.

ACKNOWLEDGMENTS

The authors thank Dr. C. Raghavan for valuable discussion regarding the manuscript, Dr. Jan E. Leach for critical reading of the manuscript, and J. Crouch for the conceptual input in initiating some of the new technology platforms presented in this review. They also thank E. Mercado, M. Y. V. Reveche, D. Skinner, J. Chen, G. Carrillo, M. Bernardo, and J. H. Chin for their intellectual and technical input for developing some of the technology platforms, and B. Hardy, science editor, IRRI, for editing the manuscript. They gratefully acknowledge the financial support from the Generation Challenge Program.

REFERENCES

[1] G. S. Khush, "Green revolution: preparing for the 21st century," *Genome*, vol. 42, no. 4, pp. 646–655, 1999.

[2] T. W. Mew, H. Leung, S. Savary, C. M. Vera Cruz, and J. E. Leach, "Looking ahead in rice disease research and management," *Critical Reviews in Plant Sciences*, vol. 23, no. 2, pp. 103–127, 2004.

[3] S. Peng, J. Huang, J. E. Sheehy, et al., "Rice yields decline with higher night temperature from global warming," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 27, pp. 9971–9975, 2004.

[4] P. Pinstrup-Andersen, R. Pandya-Lorch, and M. W. Rosegrant, "World food prospects: critical issues for the early twenty-first century," 1999, International Food Policy Research Institute, Washington, DC, USA.

[5] S. A. Goff, D. Ricke, T.-H. Lan, et al., "A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*)," *Science*, vol. 296, no. 5565, pp. 92–100, 2002.

[6] J. Yu, S. Hu, J. Wang, et al., "A draft sequence of the rice (*Oryza sativa ssp. indica*) genome," *Chinese Science Bulletin*, vol. 46, no. 23, pp. 1937–1942, 2001.

[7] T. Matsumoto, J. Wu, H. Kanamori, et al., "The map-based sequence of the rice genome," *Nature*, vol. 436, no. 7052, pp. 793–800, 2005.

[8] B. C. Y. Collard and D. J. Mackill, "Marker-assisted selection: an approach for precision plant breeding in the twenty-first century," *Philosophical Transactions of the Royal Society B*, vol. 363, no. 1491, pp. 557–572, 2008.

[9] Y. B. Xu and J. H. Crouch, "Marker-assisted selection in plant breeding: from publications to practice," *Crop Science*, vol. 48, no. 2, pp. 391–407, 2008.

[10] S. L. Dwivedi, J. H. Crouch, D. J. Mackill, et al., "The molecularization of public sector crop breeding: progress, problems, and prospects," *Advances in Agronomy*, vol. 95, pp. 163–318, 2007.

[11] E. Francia, G. Tacconi, C. Crosatti, et al., "Marker assisted selection in crop plants," *Plant Cell, Tissue and Organ Culture*, vol. 82, no. 3, pp. 317–342, 2005.

[12] M. Mohan, S. Nair, A. Bhagwat, et al., "Genome mapping, molecular markers and marker-assisted selection in crop plants," *Molecular Breeding*, vol. 3, no. 2, pp. 87–103, 1997.

[13] J.-M. Ribaut and D. Hoisington, "Marker-assisted selection: new tools and strategies," *Trends in Plant Science*, vol. 3, no. 6, pp. 236–238, 1998.

[14] R. K. Varshney, A. Graner, and M. E. Sorrells, "Genomics-assisted breeding for crop improvement," *Trends in Plant Science*, vol. 10, no. 12, pp. 621–630, 2005.

[15] H. Leung, G. P. Hettel, and R. P. Cantrell, "International rice research institute: roles and challenges as we enter the genomics era," *Trends in Plant Science*, vol. 7, no. 3, pp. 139–142, 2002.

[16] M. Morgante and F. Salamini, "From plant genomics to breeding practice," *Current Opinion in Biotechnology*, vol. 14, no. 2, pp. 214–219, 2003.

[17] N. Huang, A. Parco, T. Mew, et al., "RFLP mapping of isozymes, RAPD and QTLs for grain shape, brown planthopper resistance in a doubled haploid rice population," *Molecular Breeding*, vol. 3, no. 2, pp. 105–113, 1997.

[18] D. J. Mackill, M. A. Salam, Z. Y. Wang, and S. D. Tanksley, "A major photoperiod-sensitivity gene tagged with RFLP and isozyme markers in rice," *Theoretical and Applied Genetics*, vol. 85, no. 5, pp. 536–540, 1993.

[19] M. Mohan, S. Nair, J. S. Bentur, U. P. Rao, and J. Bennett, "RFLP and RAPD mapping of the rice *gm2* gene that confers resistance to biotype-1 of gall midge (*Orseolia oryzae*)," *Theoretical and Applied Genetics*, vol. 87, no. 7, pp. 782–788, 1994.

[20] H. Tsunematsu, A. Yoshimura, Y. Harushima, et al., "RFLP framework map using recombinant inbred lines in rice," *Breeding Science*, vol. 46, no. 3, pp. 279–284, 1996.

[21] G. Zhang, E. R. Angeles, M. L. P. Abenes, G. S. Khush, and N. Huang, "RAPD and RFLP mapping of the bacterial blight resistance gene *xa-13* in rice," *Theoretical and Applied Genetics*, vol. 93, no. 1-2, pp. 65–70, 1996.

[22] Y. Harushima, M. Yano, A. Shomura, et al., "A high-density rice genetic linkage map with 2275 markers using a single F_2 population," *Genetics*, vol. 148, no. 1, pp. 479–494, 1998.

[23] M. W. Blair and S. R. McCouch, "Microsatellite and sequence-tagged site markers diagnostic for the rice bacterial

leaf blight resistance gene *xa-5*," *Theoretical and Applied Genetics*, vol. 95, no. 1-2, pp. 174–184, 1997.

[24] T. Inoue, H. S. Zhong, A. Miyao, et al., "Sequence-tagged sites (STSs) as standard landmarkers in the rice genome," *Theoretical and Applied Genetics*, vol. 89, no. 6, pp. 728–734, 1994.

[25] N. T. Lang, P. K. Subudhi, S. S. Virmani, et al., "Development of PCR-based markers for thermosensitive genetic male sterility gene *tms3*(t) in rice (*Oryza sativa* L.)," *Hereditas*, vol. 131, no. 2, pp. 121–127, 1999.

[26] P. K. Gupta and R. K. Varshney, "The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat," *Euphytica*, vol. 113, no. 3, pp. 163–185, 2000.

[27] P. K. Gupta, R. K. Varshney, P. C. Sharma, and B. Ramesh, "Molecular markers and their applications in wheat breeding," *Plant Breeding*, vol. 118, no. 5, pp. 369–390, 1999.

[28] S. R. McCouch, X. Chen, O. Panaud, et al., "Microsatellite marker development, mapping and applications in rice genetics and breeding," *Plant Molecular Biology*, vol. 35, no. 1-2, pp. 89–99, 1997.

[29] X. Chen, S. Temnykh, Y. Xu, Y. G. Cho, and S. R. McCouch, "Development of a microsatellite framework map providing genome-wide coverage in rice (*Oryza sativa* L.)," *Theoretical and Applied Genetics*, vol. 95, no. 4, pp. 553–567, 1997.

[30] O. Panaud, X. Chen, and S. R. McCouch, "Development of microsatellite markers and characterization of simple sequence length polymorphism (SSLP) in rice (*Oryza sativa* L.)," *Molecular and General Genetics*, vol. 252, no. 5, pp. 597–607, 1996.

[31] S. Temnykh, G. DeClerck, A. Lukashova, L. Lipovich, S. Cartinhour, and S. McCouch, "Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential," *Genome Research*, vol. 11, no. 8, pp. 1441–1452, 2001.

[32] S. R. McCouch, L. Teytelman, Y. Xu, et al., "Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.)," *DNA Research*, vol. 9, no. 6, pp. 199–207, 2002.

[33] A. Rafalski, "Applications of single nucleotide polymorphisms in crop genetics," *Current Opinion in Plant Biology*, vol. 5, no. 2, pp. 94–100, 2002.

[34] F. A. Feltus, J. Wan, S. R. Schulze, J. C. Estill, N. Jiang, and A. H. Paterson, "An SNP resource for rice genetics and breeding based on subspecies *indica* and *japonica* genome alignments," *Genome Research*, vol. 14, no. 9, pp. 1812–1819, 2004.

[35] S. Nasu, J. Suzuki, R. Ohta, et al., "Search for and analysis of single nucleotide polymorphisms (SNPS) in rice (*Oryza sativa*, *Oryza rufipogon*) and establishment of SNP markers," *DNA Research*, vol. 9, no. 5, pp. 163–171, 2002.

[36] Y.-J. Shen, H. Jiang, J.-P. Jin, et al., "Development of genome-wide DNA polymorphism database for map-based cloning of rice genes," *Plant Physiology*, vol. 135, no. 3, pp. 1198–1205, 2004.

[37] L. Monna, R. Ohta, H. Masuda, A. Koike, and Y. Minobe, "Genome-wide searching of single-nucleotide polymorphisms among eight distantly and closely related rice cultivars (*Oryza sativa* L.) and a wild accession (*Oryza rufipogon* Griff.)," *DNA Research*, vol. 13, no. 2, pp. 43–51, 2006.

[38] A.-L. Hour, Y.-C. Lin, P.-F. Li, et al., "Detection of SNPs between Tainung 67 and Nipponbare rice cultivars," *Botanical Studies*, vol. 48, no. 3, pp. 243–253, 2007.

[39] K. Shirasawa, H. Maeda, L. Monna, S. Kishitani, and T. Nishio, "The number of genes having different alleles between rice cultivars estimated by SNP analysis," *Theoretical and Applied Genetics*, vol. 115, no. 8, pp. 1067–1074, 2007.

[40] S. Katagiri, J. Wu, Y. Ito, et al., "End sequencing and chromosomal *in silico* mapping of BAC clones derived from an *indica* rice cultivar, Kasalath," *Breeding Science*, vol. 54, no. 3, pp. 273–279, 2004.

[41] X. Wang, X. Zhao, J. Zhu, and W. Wu, "Genome-wide investigation of intron length polymorphisms and their potential as molecular markers in rice (*Oryza sativa* L.)," *DNA Research*, vol. 12, no. 6, pp. 417–427, 2005.

[42] R. Fjellstrom, A. M. McClung, and A. R. Shank, "SSR markers closely linked to the *Pi-z* locus are useful for selection of blast resistance in a broad array of rice germplasm," *Molecular Breeding*, vol. 17, no. 2, pp. 149–157, 2006.

[43] K. Hayashi, N. Hashimoto, M. Daigen, and I. Ashikawa, "Development of PCR-based SNP markers for rice blast resistance genes at the *Piz* locus," *Theoretical and Applied Genetics*, vol. 108, no. 7, pp. 1212–1220, 2004.

[44] K. Hayashi, H. Yoshida, and I. Ashikawa, "Development of PCR-based allele-specific and InDel marker sets for nine rice blast resistance genes," *Theoretical and Applied Genetics*, vol. 113, no. 2, pp. 251–260, 2006.

[45] T. Komori and N. Nitta, "Utilization of the CAPS/dCAPS method to convert rice SNPs into PCR-based markers," *Breeding Science*, vol. 55, no. 1, pp. 93–98, 2005.

[46] C. A. Conaway-Bormans, M. A. Marchetti, C. W. Johnson, A. M. McClung, and W. D. Park, "Molecular markers linked to the blast resistance gene *Pi-z* in rice for use in marker-assisted selection," *Theoretical and Applied Genetics*, vol. 107, no. 6, pp. 1014–1020, 2003.

[47] F. Lin, Y. Liu, L. Wang, X. Liu, and Q. Pan, "A high-resolution map of the rice blast resistance gene *Pi15* constructed by sequence-ready markers," *Plant Breeding*, vol. 126, no. 3, pp. 287–290, 2007.

[48] D.-H. Chen and P. C. Ronald, "A rapid DNA minipreparation method suitable for AFLP and other PCR applications," *Plant Molecular Biology Reporter*, vol. 17, no. 1, pp. 53–57, 1999.

[49] S. L. Dellaporta, J. Wood, and J. B. Hicks, "A plant DNA minipreparation: version II," *Plant Molecular Biology Reporter*, vol. 1, no. 4, pp. 19–21, 1983.

[50] T. M. Fulton, J. Chunwongse, and S. D. Tanksley, "Microprep protocol for extraction of DNA from tomato and other herbaceous plants," *Plant Molecular Biology Reporter*, vol. 13, no. 3, pp. 207–209, 1995.

[51] K. Zheng, P. K. Subudhi, J. Domingo, G. Magpantay, and N. Huang, "Rapid DNA isolation for marker assisted selection in rice breeding," *Rice Genetics Newsletter*, vol. 12, no. 48, 1995.

[52] X. Xu, S. Kawasaki, T. Fujimura, and C. Wang, "A protocol for high-throughput extraction of DNA from rice leaves," *Plant Molecular Biology Reporter*, vol. 23, no. 3, pp. 291–295, 2005.

[53] B. C. Y. Collard, A. Das, P. S. Virk, and D. J. Mackill, "Evaluation of 'quick and dirty' DNA extraction methods for marker-assisted selection in rice (*Oryza sativa* L.)," *Plant Breeding*, vol. 126, no. 1, pp. 47–50, 2007.

[54] H. Wang, M. Qi, and A. J. Cutler, "A simple method of preparing plant samples for PCR," *Nucleic Acids Research*, vol. 21, no. 17, pp. 4153–4154, 1993.

[55] J.-M. Ribaut, X. Hu, D. Hoisington, and D. González-de-León, "Use of STSs and SSRs as rapid and reliable preselection tools in a marker-assisted selection backcross scheme," *Plant Molecular Biology Reporter*, vol. 15, no. 2, pp. 156–163, 1997.

[56] M. W. Blair, V. Hedetale, and S. R. McCouch, "Fluorescent-labeled microsatellite panels useful for detecting allelic

diversity in cultivated rice (*Oryza sativa* L.)," *Theoretical and Applied Genetics*, vol. 105, no. 2-3, pp. 449–457, 2002.

[57] J. R. Coburn, S. V. Temnykh, E. M. Paul, and S. R. McCouch, "Design and application of microsatellite marker panels for semiautomated genotyping of rice (*Oryza sativa* L.)," *Crop Science*, vol. 42, no. 6, pp. 2092–2099, 2002.

[58] M. M. Neff, J. D. Neff, J. Chory, and A. E. Pepper, "dCAPS, a simple technique for the genetic analysis of single nucleotide polymorphisms: experimental applications in *Arabidopsis thaliana* genetics," *The Plant Journal*, vol. 14, no. 3, pp. 387–392, 1998.

[59] G. H. Jiang, C. G. Xu, J. M. Tu, X. H. Li, Y. Q. He, and Q. Zhang, "Pyramiding of insect- and disease-resistance genes into an elite *indica*, cytoplasm male sterile restorer line of rice, 'Minghui 63'," *Plant Breeding*, vol. 123, no. 2, pp. 112–116, 2004.

[60] B. Jayashree, P. T. Reddy, Y. Leeladevi, et al., "Laboratory information management software for genotyping workflows: applications in high throughput crop genotyping," *BMC Bioinformatics*, vol. 7, article 383, 2006.

[61] H. Sanchez-Villeda, S. Schroeder, M. Polacco, et al., "Development of an integrated laboratory information management system for the maize mapping project," *Bioinformatics*, vol. 19, no. 16, pp. 2022–2030, 2003.

[62] P. Jaiswal, D. Ware, J. Ni, et al., "Gramene: development and integration of trait and gene ontologies for rice," *Comparative and Functional Genomics*, vol. 3, no. 2, pp. 132–136, 2002.

[63] D. Ware, P. Jaiswal, J. Ni, et al., "Gramene: a resource for comparative grass genomics," *Nucleic Acids Research*, vol. 30, no. 1, pp. 103–105, 2002.

[64] Z. Fang, M. Polacco, S. Chen, et al., "cMap: the comparative genetic map viewer," *Bioinformatics*, vol. 19, no. 3, pp. 416–417, 2003.

[65] K. Sakata, B. A. Antonio, Y. Mukai, et al., "INE: a rice genome database with an integrated map view," *Nucleic Acids Research*, vol. 28, no. 1, pp. 97–101, 2000.

[66] Q. Yuan, S. Ouyang, J. Liu, et al., "The TIGR rice genome annotation resource: annotating the rice genome and creating resources for plant biologists," *Nucleic Acids Research*, vol. 31, no. 1, pp. 229–233, 2003.

[67] Q. Yuan, S. Ouyang, A. Wang, et al., "The institute for genomic research Osa1 rice genome annotation database," *Plant Physiology*, vol. 138, no. 1, pp. 18–26, 2005.

[68] S. Ouyang, W. Zhu, J. Hamilton, et al., "The TIGR rice genome annotation resource: improvements and new features," *Nucleic Acids Research*, vol. 35, database issue, pp. D883–D887, 2007.

[69] B. C. Y. Collard, M. Z. Z. Jahufer, J. B. Brouwer, and E. C. K. Pang, "An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts," *Euphytica*, vol. 142, no. 1-2, pp. 169–196, 2005.

[70] Y. B. Xu, "Developing marker-assisted selection strategies for breeding hybrid rice," in *Plant Breeding Reviews*, J. Janick, Ed., pp. 73–174, John Wiley & Sons, New York, NY, USA, 2003.

[71] J. O. Olufowote, Y. Xu, X. Chen, et al., "Comparative evaluation of within-cultivar variation of rice (*Oryza sativa* L.) using microsatellite and RFLP markers," *Genome*, vol. 40, no. 3, pp. 370–378, 1997.

[72] J. Yashitolaa, T. Thirumuruganb, R. M. Sundaramb, et al., "Assessment of purity of rice hybrids using microsatellite and STS markers," *Crop Science*, vol. 42, no. 4, pp. 1369–1373, 2002.

[73] P. Rajendrakumar, A. K. Biswal, S. M. Balachandran, K. Srinivasarao, and R. M. Sundaram, "Simple sequence repeats in organellar genomes of rice: frequency and distribution in genic and intergenic regions," *Bioinformatics*, vol. 23, no. 1, pp. 1–4, 2007.

[74] S. Jain, R. K. Jain, and S. R. McCouch, "Genetic analysis of Indian aromatic and quality rice (*Oryza sativa* L.) germplasm using panels of fluorescently-labeled microsatellite markers," *Theoretical and Applied Genetics*, vol. 109, no. 5, pp. 965–977, 2004.

[75] O.-Y. Jeong, M.-T. Song, J.-H. Hong, and K.-S. Lee, "Comparison of PCR-based DNA fingerprinting methods using random, microsatellite, and sequence tagged site (STS) primers for the classification of Korean rice (*Oryza sativa* L.) cultivars," *Korean Journal of Genetics*, vol. 21, no. 3, pp. 157–169, 1999.

[76] Y. Xu, H. Beachell, and S. R. McCouch, "A marker-based approach to broadening the genetic base of rice in the USA," *Crop Science*, vol. 44, no. 6, pp. 1947–1959, 2004.

[77] Y.-I. Cho, C.-W. Park, S.-W. Kwon, et al., "Key DNA markers for predicting heterosis in F_1 hybrids of *japonica* rice," *Breeding Science*, vol. 54, no. 4, pp. 389–397, 2004.

[78] J. N. Reddy, M. R. Baraoidan, M. A. Bernardo, M. L. C. George, and R. Sridhar, "Application of marker-assisted selection in rice for bacterial blight resistance gene, *Xa21*," *Current Science*, vol. 73, no. 10, pp. 873–875, 1997.

[79] G. Yi, S.-K. Lee, Y.-K. Hong, et al., "Use of *Pi5*(t) markers in marker-assisted selection to screen for cultivars with resistance to *Magnaporthe grisea*," *Theoretical and Applied Genetics*, vol. 109, no. 5, pp. 978–985, 2004.

[80] Z. Wang, Y. Jia, J. N. Rutger, and Y. Xia, "Rapid survey for presence of a blast resistance gene *Pi-ta* in rice cultivars using the dominant DNA markers derived from portions of the *Pi-ta* gene," *Plant Breeding*, vol. 126, no. 1, pp. 36–42, 2007.

[81] J. Bajracharya, K. A. Steele, D. I. Jarvis, B. R. Sthapit, and J. R. Witcombe, "Rice landrace diversity in Nepal: variability of agro-morphological traits and SSR markers in landraces from a high-altitude site," *Field Crops Research*, vol. 95, no. 2-3, pp. 327–335, 2006.

[82] L.-Z. Gao, C.-H. Zhang, D.-Y. Li, D.-J. Pan, J.-Z. Jia, and Y.-S. Dong, "Genetic diversity within *Oryza rufipogon* germplasms preserved in Chinese field gene banks of wild rice as revealed by microsatellite markers," *Biodiversity and Conservation*, vol. 15, no. 13, pp. 4059–4077, 2006.

[83] M. J. Thomson, E. M. Septiningsih, F. Suwardjo, T. J. Santoso, T. S. Silitonga, and S. R. McCouch, "Genetic diversity analysis of traditional and improved Indonesian rice (*Oryza sativa* L.) germplasm using microsatellite markers," *Theoretical and Applied Genetics*, vol. 114, no. 3, pp. 559–568, 2007.

[84] S. Chen, X. H. Lin, C. G. Xu, and Q. Zhang, "Improvement of bacterial blight resistance of 'Minghui 63', an elite restorer line of hybrid rice, by molecular marker-assisted selection," *Crop Science*, vol. 40, no. 1, pp. 239–244, 2000.

[85] S. Chen, C. G. Xu, X. H. Lin, and Q. Zhang, "Improving bacterial blight resistance of '6078', an elite restorer line of hybrid rice, by molecular marker-assisted selection," *Plant Breeding*, vol. 120, no. 2, pp. 133–137, 2001.

[86] G. H. Toenniessen, J. C. O'Toole, and J. DeVries, "Advances in plant biotechnology and its adoption in developing countries," *Current Opinion in Plant Biology*, vol. 6, no. 2, pp. 191–198, 2003.

[87] L. Shen, B. Courtois, K. L. McNally, S. Robin, and Z. Li, "Evaluation of near-isogenic lines of rice introgressed with QTLs for root depth through marker-aided selection,"

Theoretical and Applied Genetics, vol. 103, no. 1, pp. 75–83, 2001.

[88] S. Singh, J. S. Sidhu, N. Huang, et al., "Pyramiding three bacterial blight resistance genes (*xa5*, *xa13* and *Xa21*) using marker-assisted selection into *indica* rice cultivar PR106," *Theoretical and Applied Genetics*, vol. 102, no. 6-7, pp. 1011–1015, 2001.

[89] R. M. Sundaram, M. R. Vishnupriya, S. K. Biradar, et al., "Marker assisted introgression of bacterial blight resistance in Samba Mahsuri, an elite *indica* rice variety," *Euphytica*, vol. 160, no. 3, pp. 411–422, 2008.

[90] S.-P. Liu, X. Li, C.-Y. Wang, X.-H. Li, and Y.-Q. He, "Improvement of resistance to rice blast in Zhenshan 97 by molecular marker-aided selection," *Acta Botanica Sinica*, vol. 45, no. 11, pp. 1346–1350, 2003.

[91] P. H. Zhou, Y. F. Tan, Y. Q. He, C. G. Xu, and Q. Zhang, "Simultaneous improvement for four quality traits of Zhenshan 97, an elite parent of hybrid rice, by molecular marker-assisted selection," *Theoretical and Applied Genetics*, vol. 106, no. 2, pp. 326–331, 2003.

[92] M. Joseph, S. Gopalakrishnan, R. K. Sharma, et al., "Combining bacterial blight resistance and Basmati quality characteristics by phenotypic and molecular marker-assisted selection in rice," *Molecular Breeding*, vol. 13, no. 4, pp. 377–387, 2004.

[93] T. Toojinda, S. Tragoonrung, A. Vanavichit, et al., "Molecular breeding for rainfed lowland rice in the Mekong region," *Plant Production Science*, vol. 8, no. 3, pp. 330–333, 2005.

[94] E. Arnao, O. Borges, C. Ramis, A. Díaz, and I. Galindo, "Recurrent parent genome recovery in a marker assisted backcross breeding program in rice," *Interciencia*, vol. 31, no. 6, pp. 431–436, 2006.

[95] K. A. Steele, A. H. Price, H. E. Shashidhar, and J. R. Witcombe, "Marker-assisted selection to introgress rice QTLs controlling root traits into an Indian upland rice variety," *Theoretical and Applied Genetics*, vol. 112, no. 2, pp. 208–221, 2006.

[96] K. A. Steele, D. S. Virk, R. Kumar, S. C. Prasad, and J. R. Witcombe, "Field evaluation of upland rice lines selected for QTLs controlling root traits," *Field Crops Research*, vol. 101, no. 2, pp. 180–186, 2007.

[97] Y. Takeuchi, T. Ebitani, T. Yamamoto, et al., "Development of isogenic lines of rice cultivar Koshihikari with early and late heading by marker-assisted selection," *Breeding Science*, vol. 56, no. 4, pp. 405–413, 2006.

[98] C. N. Neeraja, R. Maghirang-Rodriguez, A. Pamplona, et al., "A marker-assisted backcross approach for developing submergence-tolerant rice cultivars," *Theoretical and Applied Genetics*, vol. 115, no. 6, pp. 767–776, 2007.

[99] K. A. Steele, G. Edwards, J. Zhu, and J. R. Witcombe, "Marker-evaluated selection in rice: shifts in allele frequency among bulks selected in contrasting agricultural environments identify genomic regions of importance to rice adaptation and breeding," *Theoretical and Applied Genetics*, vol. 109, no. 6, pp. 1247–1260, 2004.

[100] S. Yoshimura, A. Yoshimura, N. Iwata, et al., "Tagging and combining bacterial-blight resistance genes in rice using RAPD and RFLP markers," *Molecular Breeding*, vol. 1, no. 4, pp. 375–387, 1995.

[101] S. Hittalmani, A. Parco, T. V. Mew, R. S. Zeigler, and N. Huang, "Fine mapping and DNA marker-assisted pyramiding of the three major genes for blast resistance in rice," *Theoretical and Applied Genetics*, vol. 100, no. 7, pp. 1121–1128, 2000.

[102] A. C. Sanchez, D. S. Brar, N. Huang, Z. Li, and G. S. Khush, "Sequence tagged site marker-assisted selection for three bacterial blight resistance genes in rice," *Crop Science*, vol. 40, no. 3, pp. 792–797, 2000.

[103] A. P. Davierwala, A. P. K. Reddy, M. D. Lagu, P. K. Ranjekar, and V. S. Gupta, "Marker assisted selection of bacterial blight resistance genes in rice," *Biochemical Genetics*, vol. 39, no. 7-8, pp. 261–278, 2001.

[104] J. Ramalingam, H. S. Basharat, and G. Zhang, "STS and microsatellite marker-assisted selection for bacterial blight resistance and *waxy* genes in rice, *Oryza sativa* L," *Euphytica*, vol. 127, no. 2, pp. 255–260, 2002.

[105] P. N. Sharma, A. Torii, S. Takumi, N. Mori, and C. Nakamura, "Marker-assisted pyramiding of brown planthopper (*Nilaparvata lugens Stål*) resistance genes *Bph1* and *Bph2* on rice chromosome 12," *Hereditas*, vol. 140, no. 1, pp. 61–69, 2004.

[106] T. M. S. Nas, D. L. Sanchez, G. Q. Diaz, M. S. Mendioro, and S. S. Virmani, "Pyramiding of thermosensitive genetic male sterility (TGMS) genes and identification of a candidate *tms5* gene in rice," *Euphytica*, vol. 145, no. 1-2, pp. 67–75, 2005.

[107] J. Zhang, X. Li, G. Jiang, Y. Xu, and Y. He, "Pyramiding of *Xa7* and *Xa21* for the improvement of disease resistance to bacterial blight in hybrid rice," *Plant Breeding*, vol. 125, no. 6, pp. 600–605, 2006.

[108] H. Leung, J. Wu, B. Liu, et al., "Sustainable disease resistance in rice: current and future strategies," in *New Directions for a Diverse Planet: Proceedings of the 4th International Crop Science Congress*, Brisbane, Australia, September-October 2004.

[109] L. M. Perez, E. D. Redoña, M. S. Mendioro, C. M. Vera Cruz, and H. Leung, "Introgression of *Xa4*, *Xa7* and *Xa21* for resistance to bacterial blight in thermosensitive genetic male sterile rice (*Oryza sativa* L.) for the development of two-line hybrids," *Euphytica*. In press.

[110] J. M. S. Agarcio, L. M. Borines, D. A. Tabanao, et al., "Improving resistance of hybrid rice parental lines to bacterial blight," in *Proceedings of the International Seminar on Hybrid Rice and Agro-ecosystem*, University of Agriculture, Hanoi, Vietnam, November 2007, In press.

[111] N. N. Narayanan, N. Baisakh, C. M. Vera Cruz, S. S. Gnanamanickam, K. Datta, and S. K. Datta, "Molecular breeding for the development of blast and bacterial blight resistance in rice cv. IR50," *Crop Science*, vol. 42, no. 6, pp. 2072–2079, 2002.

[112] P. Swamy, A. N. Panchbhai, P. Dodiya, et al., "Evaluation of bacterial blight resistance in rice lines carrying multiple resistance genes and *Xa21* transgenic lines," *Current Science*, vol. 90, no. 6, pp. 818–824, 2006.

[113] K. Datta, N. Baisakh, M. K. Thet, J. Tu, and S. K. Datta, "Pyramiding transgenes for multiple resistance in rice against bacterial blight, yellow stem borer and sheath blight," *Theoretical and Applied Genetics*, vol. 106, no. 1, pp. 1–8, 2002.

[114] F. Wang, J. Zhao, J. Dai, et al., "Selection and development of representative simple sequence repeat primers and multiplex SSR sets for high throughput automated genotyping in maize," *Chinese Science Bulletin*, vol. 52, no. 2, pp. 215–223, 2007.

[115] F. Hospital, "Size of donor chromosome segments around introgressed loci and reduction of linkage drag in marker-assisted backcross programs," *Genetics*, vol. 158, no. 3, pp. 1363–1379, 2001.

[116] F. Hospital and A. Charcosset, "Marker-assisted introgression of quantitative trait loci," *Genetics*, vol. 147, no. 3, pp. 1469–1485, 1997.

[117] M. Frisch, M. Bohn, and A. E. Melchinger, "Comparison of selection strategies for marker-assisted backcrossing of a gene," *Crop Science*, vol. 39, no. 5, pp. 1295–1301, 1999.

[118] D. J. Mackill, B. C. Y. Collard, S. Heuer, and A. M. Ismail, "QTLs in rice breeding: examples for abiotic stresses," in *Proceedings of the 5th International Rice Genetics Symposium*, pp. 155–167, International Rice Research Institute, Manila, Philippines, November 2005.

[119] M. Ashikari and M. Matsuoka, "Identification, isolation and pyramiding of quantitative trait loci for rice breeding," *Trends in Plant Science*, vol. 11, no. 7, pp. 344–350, 2006.

[120] A. Yoshimura, J. X. Lei, T. Matsumoto, et al., "Analysis and pyramiding of bacterial blight resistance genes in rice by using DNA markers," in *Rice Genetics III*, pp. 577–581, International Rice Research Institute, Manila, Philippines, 1996.

[121] M. L. Shanti, M. L. C. George, C. M. Vera Cruz, et al., "Identification of resistance genes effective against rice bacterial blight pathogen in eastern India," *Plant Disease*, vol. 85, no. 5, pp. 506–512, 2001.

[122] B. Liu, S. Zhang, X. Zhu, et al., "Candidate defense genes as predictors of quantitative blast resistance in rice," *Molecular Plant-Microbe Interactions*, vol. 17, no. 10, pp. 1146–1152, 2004.

[123] J. U. Jeung, S. G. Heu, M. S. Shin, C. M. Vera Cruz, and K. K. Jena, "Dynamics of *Xanthomonas oryzae* pv. *oryzae* populations in Korea and their relationship to known bacterial blight resistance genes," *Phytopathology*, vol. 96, no. 8, pp. 867–875, 2006.

[124] Y. Wang, Y. Xue, and J. Li, "Towards molecular breeding and improvement of rice in China," *Trends in Plant Science*, vol. 10, no. 12, pp. 610–614, 2005.

[125] S. K. Datta, "Rice biotechnology: a need for developing countries," *AgBioForum*, vol. 7, no. 1-2, pp. 31–35, 2004.

[126] D. S. Brar and G. S. Khush, "Cytogenetic manipulation and germplasm enhancement of rice (*Oryza sativa* L.)," in *Genetic Resources, Chromosome Engineering, and Crop Improvement*, R. J. Singh and P. P. Jauhar, Eds., vol. 2, pp. 115–158, CRC Press, Boca Raton, Fla, USA, 2006.

[127] G. Barry, P. S. Virk, A. Das, J. Tan, and P. Herve, "International status and achievement for development of nutrient rice," in *Proceedings of the International Symposium on Rice Biofortification-"Improving Human Health through Biofortified Rice"*, pp. 9–32, Suwon, Korea, September 2006.

[128] P. S. Virk, G. Barry, and A. Das, "Transferring beta-carotene loci from improved transgenic rice into popular *indica* varieties using marker aided backcrossing," in *Proceedings of the 13th Australasian Plant Breeding Conference*, Christchurch, New Zealand, April 2006.

[129] P. S. Virk, G. Barry, A. Das, J. H. Lee, and J. Tan, "Research status of micronutrient rice development in Asia," in *Proceedings of the International Symposium on Rice Biofortification-'Improving Human Health through Biofortified Rice'*, pp. 123–148, Suwon, South Korea, September 2006.

[130] M. Gowda, R. C. Venu, K. Roopalakshmi, M. V. Sreerekha, and R. S. Kulkarni, "Advances in rice breeding, genetics and genomics," *Molecular Breeding*, vol. 11, no. 4, pp. 337–352, 2003.

[131] H. Jiang, L.-B. Guo, and Q. Qian, "Recent progress on rice genetics in China," *Journal of Integrative Plant Biology*, vol. 49, no. 6, pp. 776–790, 2007.

[132] A. K. Tyagi, J. P. Khurana, P. Khurana, et al., "Structural and functional analysis of rice genome," *Journal of Genetics*, vol. 83, no. 1, pp. 79–99, 2004.

[133] R. K. Varshney, D. A. Hoisington, and A. K. Tyagi, "Advances in cereal genomics and applications in crop breeding," *Trends in Biotechnology*, vol. 24, no. 11, pp. 490–499, 2006.

[134] Y. Xu, S. R. McCouch, and Q. Zhang, "How can we use genomics to improve cereals with rice as a reference genome?" *Plant Molecular Biology*, vol. 59, no. 1, pp. 7–26, 2005.

[135] W. A. Rensink and C. R. Buell, "Microarray expression profiling resources for plant genomics," *Trends in Plant Science*, vol. 10, no. 12, pp. 603–609, 2005.

[136] S. Kikuchi, K. Satoh, T. Nagata, et al., "Collection, mapping, and annotation of over 28,000 cDNA clones from *japonica* rice," *Science*, vol. 301, no. 5631, pp. 376–379, 2003.

[137] K. Xie, J. Zhang, Y. Xiang, et al., "Isolation and annotation of 10828 putative full length cDNAs from *indica* rice," *Science in China Series C*, vol. 48, no. 5, pp. 445–451, 2005.

[138] K. Sakata, Y. Nagamura, H. Numa, et al., "RiceGAAS: an automated annotation system and database for rice genome sequence," *Nucleic Acids Research*, vol. 30, no. 1, pp. 98–102, 2002.

[139] H. Ohyanagi, T. Tanaka, H. Sakai, et al., "The Rice Annotation Project Database (RAP-DB): hub for *Oryza sativa* ssp. *japonica* genome information," *Nucleic Acids Research*, vol. 34, database issue, pp. D741–D744, 2006.

[140] H. Leung and G. L. An, "Rice functional genomics: large-scale gene discovery and applications to crop improvement," *Advances in Agronomy*, vol. 82, pp. 55–111, 2004.

[141] G. An, S. Lee, S.-H. Kim, and S.-R. Kim, "Molecular genetics using T-DNA in rice," *Plant and Cell Physiology*, vol. 46, no. 1, pp. 14–22, 2005.

[142] H. Hirochika, E. Guiderdoni, G. An, et al., "Rice mutant resources for gene discovery," *Plant Molecular Biology*, vol. 54, no. 3, pp. 325–334, 2004.

[143] D.-H. Jeong, S. An, S. Park, et al., "Generation of a flanking sequence-tag database for activation-tagging lines in *japonica* rice," *The Plant Journal*, vol. 45, no. 1, pp. 123–132, 2006.

[144] D. Miki, S. Moritoh, and K. Shimamoto, "Gene-specific RNAi using 3′ UTR-dsRNA in rice," *Plant and Cell Physiology*, vol. 44, pp. S12–S12, 2003.

[145] Z. Wang, C. Chen, Y. Xu, et al., "A practical vector for efficient knockdown of gene expression in rice (*Oryza sativa* L.)," *Plant Molecular Biology Reporter*, vol. 22, no. 4, pp. 409–417, 2004.

[146] J.-L. Wu, C. Wu, C. Lei, et al., "Chemical- and irradiation-induced mutants of *indica* rice IR64 for forward and reverse genetics," *Plant Molecular Biology*, vol. 59, no. 1, pp. 85–97, 2005.

[147] C. G. McLaren, R. M. Bruskiewich, A. M. Portugal, and A. B. Cosico, "The International Rice Information System. A platform for meta-analysis of rice crop data," *Plant Physiology*, vol. 139, no. 2, pp. 637–642, 2005.

[148] G. Droc, M. Ruiz, P. Larmande, et al., "OryGenesDB: a database for rice reverse genetics," *Nucleic Acids Research*, vol. 34, database issue, pp. D736–D740, 2006.

[149] L. J. G. van Enckevort, G. Droc, P. Piffanelli, et al., "EU-OSTID: a collection of transposon insertional mutants for functional genomics in rice," *Plant Molecular Biology*, vol. 59, no. 1, pp. 99–110, 2005.

[150] D. Y. Chao, Y. H. Luo, M. Shi, D. Luo, and H. X. Lin, "Salt-responsive genes in rice revealed by cDNA microarray analysis," *Cell Research*, vol. 15, no. 10, pp. 796–810, 2005.

[151] Q. Li, F. Chen, L. Sun, Z. Zhang, Y. Yang, and Z. He, "Expression profiling of rice genes in early defense responses to blast and bacterial blight pathogens using cDNA microarray," *Physiological and Molecular Plant Pathology*, vol. 68, no. 1–3, pp. 51–60, 2006.

[152] M. A. Rabbani, K. Maruyama, H. Abe, et al., "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 Physiology*, vol. 133, no. 4, pp. 1755–1767, 2003.

[153] H. Yamakawa, T. Hirose, M. Kuroda, and T. Yamaguchi, "Comprehensive expression profiling of rice grain filling-related genes under high temperature using DNA microarray," *Plant Physiology*, vol. 144, no. 1, pp. 258–277, 2007.

[154] H. Yamakawa, M. Kuroda, and T. Yamaguchi, "Microarray analysis of rice grain filling under high temperature," *Plant and Cell Physiology*, vol. 47, p. 562, 2006.

[155] L. Li, X. Wang, V. Stolc, et al., "Genome-wide transcription analyses in rice using tiling microarrays," *Nature Genetics*, vol. 38, no. 1, pp. 124–129, 2006.

[156] V. Stolc, L. Li, X. Wang, et al., "A pilot study of transcription unit analysis in rice using oligonucleotide tiling-path microarray," *Plant Molecular Biology*, vol. 59, no. 1, pp. 137–149, 2005.

[157] S. Komatsu and N. Tanaka, "Rice proteome analysis: a step toward functional analysis of the rice genome," *Proteomics*, vol. 5, no. 4, pp. 938–949, 2005.

[158] R. Rakwal and G. K. Agrawal, "Rice proteomics: current status and future perspective," *Electrophoresis*, vol. 24, no. 19-20, pp. 3378–3389, 2003.

[159] J. Yazaki, K. Kojima, K. Suzuki, N. Kishimoto, and S. Kikuchi, "The rice PIPELINE: a unification tool for plant functional genomics," *Nucleic Acids Research*, vol. 32, database issue, pp. D383–D387, 2004.

[160] H. Zeng, L. Luo, W. Zhang, et al., "PlantQTL-GE: a database system for identifying candidate genes in rice and *Arabidopsis* by gene expression and QTL information," *Nucleic Acids Research*, vol. 35, database issue, pp. D879–D882, 2007.

[161] S. Pflieger, V. Lefebvre, and M. Causse, "The candidate gene approach in plant genetics: a review," *Molecular Breeding*, vol. 7, no. 4, pp. 275–291, 2001.

[162] J. Ramalingam, C. M. Vera Cruz, K. Kukreja, et al., "Candidate defense genes from rice, barley, and maize and their association with qualitative and quantitative resistance in rice," *Molecular Plant-Microbe Interactions*, vol. 16, no. 1, pp. 14–24, 2003.

[163] J. D. Faris, W. L. Li, D. J. Liu, P. D. Chen, and B. S. Gill, "Candidate gene analysis of quantitative disease resistance in wheat," *Theoretical and Applied Genetics*, vol. 98, no. 2, pp. 219–225, 1999.

[164] C. M. McCallum, L. Comai, E. A. Greene, and S. Henikoff, "Targeting induced local lesions IN genomes (TILLING) for plant functional genomics," *Plant Physiology*, vol. 123, no. 2, pp. 439–442, 2000.

[165] L. Comai, K. Young, B. J. Till, et al., "Efficient discovery of DNA polymorphisms in natural populations by Ecotilling," *The Plant Journal*, vol. 37, no. 5, pp. 778–786, 2004.

[166] C. Raghavan, M. E. B. Naredo, H. Wang, et al., "Rapid method for detecting SNPs on agarose gels and its application in candidate gene mapping," *Molecular Breeding*, vol. 19, no. 2, pp. 87–101, 2007.

[167] K. L. McNally, R. Bruskiewich, D. J. Mackill, C. R. Buell, J. E. Leach, and H. Leung, "Sequencing multiple and diverse rice varieties: connecting whole-genome variation with phenotypes," *Plant Physiology*, vol. 141, no. 1, pp. 26–31, 2006.

[168] D. A. Hinds, L. L. Stuve, G. B. Nilsen, et al., "Whole-genome patterns of common DNA variation in three human populations," *Science*, vol. 307, no. 5712, pp. 1072–1079, 2005.

[169] K. A. Frazer, E. Eskin, H. M. Kang, et al., "A sequence-based variation map of 8.27 million SNPs in inbred mouse strains," *Nature*, vol. 448, no. 7157, pp. 1050–1053, 2007.

[170] R. M. Clark, G. Schweikert, C. Toomajian, et al., "Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*," *Science*, vol. 317, no. 5836, pp. 338–342, 2007.

[171] A. J. Garris, S. R. McCouch, and S. Kresovich, "Population structure and its effect on haplotype diversity and linkage disequilibrium surrounding the *xa5* locus of rice (*Oryza sativa* L.)," *Genetics*, vol. 165, no. 2, pp. 759–769, 2003.

[172] K. M. Olsen, A. L. Caicedo, N. Polato, A. McClung, S. McCouch, and M. D. Purugganan, "Selection under domestication: evidence for a sweep in the rice *waxy* genomic region," *Genetics*, vol. 173, no. 2, pp. 975–983, 2006.

[173] S. McCouch, "Diversifying selection in plant breeding," *PLoS Biology*, vol. 2, no. 10, p. e347, 2004.

[174] S. R. McCouch, M. Sweeney, J. Li, et al., "Through the genetic bottleneck: *O. rufipogon* as a source of trait-enhancing alleles for *O. sativa*," *Euphytica*, vol. 154, no. 3, pp. 317–339, 2007.

[175] D. Zamir, "Improving plant breeding with exotic genetic libraries," *Nature Reviews Genetics*, vol. 2, no. 12, pp. 983–989, 2001.

[176] L. Tan, F. Liu, W. Xue, et al., "Development of *Oryza* rufipogon and *O. sativa* introgression lines and assessment for yield-related quantitative trait loci," *Journal of Integrative Plant Biology*, vol. 49, no. 6, pp. 871–884, 2007.

[177] J. S. S. Ammiraju, M. Luo, J. L. Goicoechea, et al., "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 Research*, vol. 16, no. 1, pp. 140–147, 2006.

[178] R. A. Wing, J. S. S. Ammiraju, M. Luo, et al., "The *Oryza* map alignment project: the golden path to unlocking the genetic potential of wild rice species," *Plant Molecular Biology*, vol. 59, no. 1, pp. 53–62, 2005.

[179] M. J. Asíns, "Present and future of quantitative trait locus analysis in plant breeding," *Plant Breeding*, vol. 121, no. 4, pp. 281–291, 2002.

[180] M. J. Kearsey and A. G. L. Farquhar, "QTL analysis in plants; where are we now?" *Heredity*, vol. 80, no. 2, pp. 137–142, 1998.

[181] S. A. Flint-Garcia, J. M. Thornsberry, and E. S. Buckler IV, "Structure of linkage disequilibrium in plants," *Annual Review of Plant Biology*, vol. 54, pp. 357–374, 2003.

[182] P. K. Gupta, S. Rustgi, and P. L. Kulwal, "Linkage disequilibrium and association studies in higher plants: present status and future prospects," *Plant Molecular Biology*, vol. 57, no. 4, pp. 461–485, 2005.

[183] H. A. Agrama, G. C. Eizenga, and W. Yan, "Association mapping of yield and its components in rice cultivars," *Molecular Breeding*, vol. 19, no. 4, pp. 341–356, 2007.

[184] H. Iwata, Y. Uga, Y. Yoshioka, K. Ebana, and T. Hayashi, "Bayesian association mapping of multiple quantitative trait loci and its application to the analysis of genetic variation among *Oryza sativa* L. germplasms," *Theoretical and Applied Genetics*, vol. 114, no. 8, pp. 1437–1449, 2007.

[185] H. Lu, M. A. Redus, J. R. Coburn, J. N. Rutger, S. R. McCouch, and T. H. Tai, "Population structure and breeding patterns of 145 U.S. rice cultivars based on SSR marker analysis," *Crop Science*, vol. 45, no. 1, pp. 66–76, 2005.

[186] N. Zhang, Y. Xu, M. Akash, S. McCouch, and J. H. Oard, "Identification of candidate markers associated with agronomic traits in rice using discriminant analysis," *Theoretical and Applied Genetics*, vol. 110, no. 4, pp. 721–729, 2005.

[187] P. Tanhuanpäauml and J. Vilkki, "Marker-assisted selection for oleic acid content in spring turnip rape," *Plant Breeding*, vol. 118, no. 6, pp. 568–570, 1999.

[188] P. Donini, P. Stephenson, G. J. Bryan, and R. M. D. Koebner, "The potential of microsatellites for high throughput genetic diversity assessment in wheat and barley," *Genetic Resources and Crop Evolution*, vol. 45, no. 5, pp. 415–421, 1998.

[189] O. Henegariu, N. A. Heerema, S. R. Dlouhy, G. H. Vance, and P. H. Vogt, "Multiplex PCR: critical parameters and step-by-step protocol," *BioTechniques*, vol. 23, no. 3, pp. 504–511, 1997.

[190] X. Chen and P. F. Sullivan, "Single nucleotide polymorphism genotyping: biochemistry, protocol, cost and throughput," *Pharmacogenomics Journal*, vol. 3, no. 2, pp. 77–96, 2003.

[191] A.-C. Syvänen, "Accessing genetic variation: genotyping single nucleotide polymorphisms," *Nature Reviews Genetics*, vol. 2, no. 12, pp. 930–942, 2001.

[192] A.-C. Syvänen, "Toward genome-wide SNP genotyping," *Nature Genetics*, vol. 37, no. 6s, pp. S5–S10, 2005.

[193] J. S. Bao, H. Corke, and M. Sun, "Nucleotide diversity in *starch synthase IIa* and validation of single nucleotide polymorphisms in relation to starch gelatinization temperature and other physicochemical properties in rice (*Oryza sativa* L.)," *Theoretical and Applied Genetics*, vol. 113, no. 7, pp. 1171–1183, 2006.

[194] P. C. Bundock, M. J. Cross, F. M. Shapter, and R. J. Henry, "Robust allele-specific polymerase chain reaction markers developed for single nucleotide polymorphisms in expressed barley sequences," *Theoretical and Applied Genetics*, vol. 112, no. 2, pp. 358–365, 2006.

[195] E. Domon, T. Yanagisawa, A. Saito, and K. Takeda, "Single nucleotide polymorphism genotyping of the barley *waxy* gene by polymerase chain reaction with confronting two-pair primers," *Plant Breeding*, vol. 123, no. 3, pp. 225–228, 2004.

[196] S. B. Kadaru, A. S. Yadav, R. G. Fjellstrom, and J. H. Oard, "Alternative Ecotilling protocol for rapid, cost-effective single-nucleotide polymorphism discovery and genotyping in rice (*Oryza sativa* L.)," *Plant Molecular Biology Reporter*, vol. 24, no. 1, pp. 3–22, 2006.

[197] Y. Sato and T. Nishio, "Efficient detection of DNA polymorphism in cabbage and rice cultivars by PCR-RF-SSCP (PRS)," *Plant Cell Reports*, vol. 21, no. 3, pp. 276–281, 2002.

[198] Y. Sato and T. Nishio, "Mutation detection in rice *waxy* mutants by PCR-RF-SSCP," *Theoretical and Applied Genetics*, vol. 107, no. 3, pp. 560–567, 2003.

[199] K. Shirasawa, L. Monna, S. Kishitani, and T. Nishio, "Single nucleotide polymorphisms in randomly selected genes among *japonica* rice (*Oryza sativa* L.) varieties identified by PCR-RF-SSCP," *DNA Research*, vol. 11, no. 4, pp. 275–283, 2004.

[200] K. Shirasawa, S. Kishitani, and T. Nishio, "Dot-blot analysis for identification of *japonica* rice cultivars and genotyping of recombinant inbred lines," *Breeding Science*, vol. 55, no. 2, pp. 187–192, 2005.

[201] K. Shirasawa, S. Shiokai, M. Yamaguchi, S. Kishitani, and T. Nishio, "Dot-blot-SNP analysis for practical plant breeding and cultivar identification in rice," *Theoretical and Applied Genetics*, vol. 113, no. 1, pp. 147–155, 2006.

[202] D. J. Cahill and D. H. Schmidt, "Use of marker assisted selection in a product development breeding program," in *New Directions for a Diverse Planet: Proceedings of the 4th International Crop Science Congress*, T. Fischer, N. Turner, J. Angus, et al., Eds., Brisbane, Australia, September-October 2004.

[203] M. Pessoa-Filho, A. Beló, A. A. N. Alcochete, P. H. N. Rangel, and M. E. Ferreira, "A set of multiplex panels of microsatellite markers for rapid molecular characterization of rice accessions," *BMC Plant Biology*, vol. 7, article 23, 2007.

[204] L. R. Rampling, N. Harker, M. R. Shariflou, and M. K. Morell, "Detection and analysis systems for microsatellite markers in wheat," *Australian Journal of Agricultural Research*, vol. 52, no. 12, pp. 1131–1141, 2001.

[205] M. Schuelke, "An economic method for the fluorescent labeling of PCR fragments," *Nature Biotechnology*, vol. 18, no. 2, pp. 233–234, 2000.

[206] J. Batley, G. Barker, H. O'Sullivan, K. J. Edwards, and D. Edwards, "Mining for single nucleotide polymorphisms and insertions/deletions in maize expressed sequence tag data," *Plant Physiology*, vol. 132, no. 1, pp. 84–91, 2003.

[207] S. Grösch, E. Niederberger, J. Lötsch, C. Skarke, and G. Geisslinger, "A rapid screening method for a single nucleotide polymorphism (SNP) in the human MOR gene," *British Journal of Clinical Pharmacology*, vol. 52, no. 6, pp. 711–714, 2001.

[208] J.-B. Fan, X. Chen, M. K. Halushka, et al., "Parallel genotyping of human SNPs using generic high-density oligonucleotide tag arrays," *Genome Research*, vol. 10, no. 6, pp. 853–860, 2000.

[209] X. Chen, B. Zehnbauer, A. Gnirke, and P.-Y. Kwok, "Fluorescence energy transfer detection as a homogeneous DNA diagnostic method," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 94, no. 20, pp. 10756–10761, 1997.

[210] L. He, D. P. Olson, X. Wu, T. S. Karpova, J. G. McNally, and P. E. Lipsky, "A flow cytometric method to detect protein-protein interaction in living cells by directly visualizing donor fluorophore quenching during CFP→YFP fluorescence resonance energy transfer (FRET)," *Cytometry Part A*, vol. 55A, no. 2, pp. 71–85, 2003.

[211] K. Takatsu, T. Yokomaku, S. Kurata, and T. Kanagawa, "A new approach to SNP genotyping with fluorescently labeled mononucleotides," *Nucleic Acids Research*, vol. 32, no. 7, p. e60, 2004.

[212] K. Takatsu, T. Yokomaku, S. Kurata, and T. Kanagawa, "A FRET-based analysis of SNPs without fluorescent probes," *Nucleic Acids Research*, vol. 32, no. 19, p. e156, 2004.

[213] A. J. Flavell, V. N. Bolshakov, A. Booth, et al., "A microarray-based high throughput molecular marker genotyping method: the tagged microarray marker (TAM) approach," *Nucleic Acids Research*, vol. 31, no. 19, p. e115, 2003.

[214] M. Ji, P. Hou, S. Li, N. He, and Z. Lu, "Microarray-based method for genotyping of functional single nucleotide polymorphisms using dual-color fluorescence hybridization," *Mutation Research*, vol. 548, no. 1-2, pp. 97–105, 2004.

[215] M. T. Barrett, A. Scheffer, A. Ben-Dor, et al., "Comparative genomic hybridization using oligonucleotide microarrays

and total genomic DNA," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 51, pp. 17765–17770, 2004.

[216] J. O. Borevitz, D. Liang, D. Plouffe, et al., "Large-scale identification of single-feature polymorphisms in complex genomes," *Genome Research*, vol. 13, no. 3, pp. 513–523, 2003.

[217] H. L. Stickney, J. Schmutz, I. G. Woods, et al., "Rapid mapping of zebrafish mutations with SNPs and oligonucleotide microarrays," *Genome Research*, vol. 12, no. 12, pp. 1929–1934, 2002.

[218] D. W. Galbraith, "DNA microarray analyses in higher plants," *OMICS: A Journal of Integrative Biology*, vol. 10, no. 4, pp. 455–473, 2006.

[219] M. Paris, R. H. Potter, R. C. M. Lance, C. D. Li, and M. G. K. Jones, "Typing *Mlo* alleles for powdery mildew resistance in barley by single nucleotide polymorphism analysis using MALDI-ToF mass spectrometry," *Australian Journal of Agricultural Research*, vol. 54, no. 11-12, pp. 1343–1349, 2003.

[220] R. J. Snowdon and W. Friedt, "Molecular markers in *Brassica* oilseed breeding: current status and future possibilities," *Plant Breeding*, vol. 123, no. 1, pp. 1–8, 2004.

[221] T. Ebitani, Y. Takeuchi, Y. Nonoue, T. Yamamoto, K. Takeuchi, and M. Yano, "Construction and evaluation of chromosome segment substitution lines carrying overlapping chromosome segments of *indica* rice cultivar 'Kasalath' in a genetic background of *japonica* elite cultivar 'Koshihikari'," *Breeding Science*, vol. 55, no. 1, pp. 65–73, 2005.

[222] Z.-Y. Xi, F.-H. He, R.-Z. Zeng, et al., "Development of a wide population of chromosome single-segment substitution lines in the genetic background of an elite cultivar of rice (*Oryza sativa* L.)," *Genome*, vol. 49, no. 5, pp. 476–484, 2006.

[223] L. M. T. Bradbury, R. J. Henry, Q. Jin, R. F. Reinke, and D. L. E. Waters, "A perfect marker for fragrance genotyping in rice," *Molecular Breeding*, vol. 16, no. 4, pp. 279–283, 2005.

[224] A.S. Iyer-Pascuzzi and S. R. McCouch, "Functional markers for *xa5*-mediated resistance in rice (*Oryza sativa*, L.)," *Molecular Breeding*, vol. 19, no. 4, pp. 291–296, 2007.

[225] K. Dreher, M. Khairallah, J.-M. Ribaut, and M. Morris, "Money matters (I): costs of field and laboratory procedures associated with conventional and marker-assisted maize breeding at CIMMYT," *Molecular Breeding*, vol. 11, no. 3, pp. 221–234, 2003.

[226] S. R. M. Pinson, F. M. Capdevielle, and J. H. Oard, "Confirming QTLs and finding additional loci conditioning sheath blight resistance in rice using recombinant inbred lines," *Crop Science*, vol. 45, no. 2, pp. 503–510, 2005.

[227] B. Miflin, "Crop improvement in the 21st century," *Journal of Experimental Botany*, vol. 51, no. 342, pp. 1–8, 2000.

Hindawi Publishing Corporation
International Journal of Plant Genomics
Volume 2008, Article ID 362451, 6 pages
doi:10.1155/2008/362451

Review Article

Genomics of Sorghum

Andrew H. Paterson

Plant Genome Mapping Laboratory, Department of Crop and Soil Sciences, University of Georgia, 111 Riverbend Road, Athens, GA 30602, USA

Correspondence should be addressed to Andrew H. Paterson, paterson@uga.edu

Received 26 April 2007; Accepted 13 January 2008

Recommended by Yunbi Xu

Sorghum (*Sorghum bicolor* (L.) Moench) is a subject of plant genomics research based on its importance as one of the world's leading cereal crops, a biofuels crop of high and growing importance, a progenitor of one of the world's most noxious weeds, and a botanical model for many tropical grasses with complex genomes. A rich history of genome analysis, culminating in the recent complete sequencing of the genome of a leading inbred, provides a foundation for invigorating progress toward relating sorghum genes to their functions. Further characterization of the genomes other than Saccharinae cereals may shed light on mechanisms, levels, and patterns of evolution of genome size and structure, laying the foundation for further study of sugarcane and other economically important members of the group.

1. WHY SORGHUM?

As a food and feed crop, sorghum is an important "failsafe" in the global agroecosystem. Worldwide, sorghum is the 5th most important grain crop grown based on tonnage, after maize, wheat, rice, and barley (www.fao.org). Sorghumis unusually tolerant of low input levels, an essential trait for areas such as Northeast Africa and the US Southern Plains that receive too little rainfall for most other grains. Increased demand for limited fresh water supplies, increasing use of marginal farmland, and global climatic trends, all suggest that dryland crops such as sorghum will be of growing importance to feed the world's expanding populations.

Currently the 2nd source of grain-based ethanol in the US (after maize), sorghum is a biofuel crop of growing importance. The generally lower water demands and market price for sorghum than maize, versus their equal per-bushel ethanol yields, suggests that sorghum will be of growing importance in meeting grain-based biofuels needs. Cellulosic biofuel production offers compelling advantages over seed-based production [1], but will require greater utilization of marginal lands to make the low per-unit value of biomass production economical, and will be heavily dependent upon the use of perennials to be sustainable [2, 3]. A relatively advanced state of knowledge of the genetic control of perenniality in sorghum [4, 5] and early progress in functional genomics of perenniality [6] add to its promise as a cellulosic biofuels crop. "Sweet sorghums" with high sugar content in stems, already grown for forage and silage, may be especially promising.

The *Sorghum* genus also offers the opportunity to gain new insights into biology of weeds and invasives. Vegetative dispersal by rhizomes (underground stems) and seed dispersal by disarticulation of the mature inflorescence (shattering) cause "Johnsongrass" [*Sorghum halepense* (L.) Pers, $2n = 2x = 40$] to rank among the world's most noxious weeds [7]. Johnsongrass is an interspecific hybrid of *Sorghum bicolor* and *S. propinquum*, the latter contributing rhizomatousness. *Sorghum bicolor* and *S. propinquum* are readily crossed, and their progeny provide a system in which to dissect the genetic basis of rhizomatousness [4]. The same features that make Johnsongrass such a troublesome weed are actually desirable in many forage, turf, and biomass crops which are genetically complex. Therefore, sorghum offers novel learning opportunities relevant to weed biology as well as to improvement of a wide range of other forage, turf, and biomass crops.

The small genome of sorghum has long been an attractive model for advancing understanding of the structure, function, and evolution of cereal genomes. Sorghum is representative of tropical grasses in that it has "C4" photosynthesis, using complex biochemical and morphological specializations to improve carbon assimilation at high temperatures. By contrast, rice is more representative of temperate grasses, using "C3" photosynthesis. Its lower level of gene duplication than many other tropical cereals makes sorghum, like rice, an attractive model for functional genomics. However, sorghum is much more closely related than rice to many major cereal crops with complex genomes and high levels of gene duplication. *Sorghum* and *Zea* (maize, the leading US crop with a farm-gate value of $15–20 billion/y) diverged from a common ancestor ∼12 mya [8, 9] versus ∼42 mya for rice and the maize/sorghum lineage [10]. *Saccharum* (sugarcane), arguably the most important biofuels crop worldwide, valued at ∼$30 billion including $1 billion/y in the US*, may have shared ancestry with sorghum as little as 5 million years ago [11], retains similar gene order [12], and even produces viable progeny in some intergeneric crosses [13]. *Zea* has undergone one whole-genome duplication since its divergence from *Sorghum* [14], and *Saccharum* has undergone at least two [12].

2. PROGRESS IN SORGHUM GENOME CHARACTERIZATION

2.1. Genetic mapping

Linkage mapping in sorghum takes advantage of its straightforward diploid genetics, amenability to inbreeding, high levels of DNA polymorphism between *Sorghum* species, and manageable levels of DNA polymorphism within *S. bicolor*. High-density reference maps of one intraspecific *S. bicolor* [15–18] and one interspecific *S. bicolor* x *S. propinquum* [19, 20] cross provide about 2600 sequence-tagged-sites (based on low-copy probes that have been sequenced), 2454 AFLP, and ∼1375 sequence-scanned (based on sequences of genetically anchored BAC clones) loci. These two maps share one common parent (*S. bicolor* "BTx623") and are essentially colinear [21]. Cytological characterization of the individual sorghum chromosomes has provided a generally adopted numbering system [22].

More than 800 markers mapped in sorghum are derived from other taxa (hence serve as comparative anchors) and additional sorghum markers have been mapped directly in other taxa, or can be plotted based on sequence similarity. Anchoring of the sorghum maps to those of rice [10, 23], maize [20, 24], sugarcane [12, 25], millet [26], switchgrass [27], bermuda grass [28], and others provides for the cross-utilization of results to simultaneously advance knowledge of many important crops.

2.2. Physical mapping

Sorghum was the first angiosperm for which a BAC library was published [29]. Estimates of the physical size of the sorghum genome range from 700 Mbp based on Cot analysis [30] to 772 Mbp based on flow cytometry [31]. This makes the sorghum genome about 60% larger than that of rice, but only about 1/4 the size of the genomes of maize or human. DNA renaturation kinetic analysis [30] shows the sorghum genome to be comprised of about 16% foldback DNA, 15% highly repetitive DNA (with individual families occurring at an average of 5200 copies per genome), 41% middle-repetitive DNA (average 72 copies) and 24% low-copy DNA. About 4% of the DNA remained single-stranded at very high Cot values and is assumed to have been damaged (thus the other percentages are slight underestimates).

High-coverage BAC libraries are available for BTx623 (about 12X coverage from H*ind*III and 8X from B*am*HI), *S. propinquum* (13-14X coverage from E*co*RI (∼7X) and H*ind*III (∼7X) and IS3620C (∼9X coverage from H*ind*III). A total of 69 545 agarose-based fingerprints from BTx623 BACs are also anchored with 211,558 hybridization loci from 7292 probes (about 2000 of which are genetically mapped). In parallel, 40 957 agarose-based fingerprints from *S. propinquum* are anchored with 189 735 hybridization loci from 7481 probes (2000 genetically mapped). Targeted HICF of additional contig-terminal BACs has been used to fill gaps. Each of these has been assembled into WebFPC-accessible physical maps (http://www.stardaddy.uga.edu/fpc/WebAGCoL/bicolor/WebFPC and http://www.stardaddy.uga.edu/fpc/WebAGCoL/propinquum/WebFPC), for which earlier versions have been described in detail [32]. About 456 *S. propinquum* and 303 *S. bicolor* BAC contigs (41% of BACs, 80% of single-copy loci) appear to be well-anchored to euchromatic regions, with the percentage of the genome attributable to euchromatin likely to rise with additional anchoring. The finding that 41% of BACs are anchored to euchromatin while only 24% of the sorghum genomic DNA is single- or low-copy [with an overall kinetic complexity of 1.64×10^8 [30]], suggests that sorghum euchromatin includes a mixture of low-copy and repetitive DNA.

2.3. Genome sequence

The shotgun sequencing of a leading US sorghum inbred, BTx623, is now complete, with ∼10.5 million reads (∼8X coverage) deposited in the NCBI Trace Archive. Early analysis confirms that the sorghum genome sequence will be a suitable substrate for a complete and high-quality annotation. In a preliminary assembly (that is expected to further improve with ongoing analysis), more than 97% of sorghum protein-coding genes (ESTs) were captured in the ∼250 longest scaffolds. The vast majority of these can be linked, ordered, and oriented using the genetic and physical map to reconstruct complete chromosomes. Alignments of the preliminary assembly to sorghum methyl-filtered sequence; sorghum, maize, and sugarcane transcript assemblies; and the *Arabidopsis* and rice proteomes confirms the base-level accuracy of the assembly and correct local structure of protein-coding loci.

Additional resources from reduced-representation sequencing will contribute to the identification of expressed portions of the genome sequence. The sorghum gene space is presently represented by approximately 204 000 expressed sequence tags, many of which have been clustered into ∼22 000 unigenes representing more than 20 diverse libraries from several genotypes [33]. About 500 000 methyl-filtered (MF) reads that provide an estimated 1x coverage of the

MF-estimated gene space [34] have been assembled into contigs (SAMIs, http://magi.plantgenomics.iastate.edu).

3. POSTGENOMICS OF SORGHUM

With the genome sequence available, one can anticipate renewed interest and accelerated progress in relating sorghum genes to their functions. Prior efforts will benefit from the sequence as a means of integrating diverse data types, providing for the formulation and testing of new hypotheses about roles of specific genes in particular traits. Existing data from QTL mapping, expression profiling, and early association genetics studies are likely to figure prominently in this merger. To fully realize the fruits of the sorghum sequence, additional functional genomics resources will be needed that provide for identification and study of crippling mutations in specific sorghum genes, in a manner that can be targeted to the subset of genes for which sorghum is a preferred system over rice, maize, or other cereal models.

3.1. QTL mapping

Motivated by interest in a range of basic and applied questions, the linkage maps of sorghum have been employed in the "tagging" (mapping) of genes for a large number of traits. The interspecific population has been especially useful for characterization of genes related to domestication, such as seed size, shattering [23], tillering, and rhizomatousness [4]. Plant height and flowering time [35, 36] have been a high priority. Similarly, the importance of hybrid sorghum motivated much research into the genetic control of fertility restoration [37–39]. Resistance genes have been tagged for numerous diseases [40–47], key insect pests [48–51], and also the parasitic weed, striga [40, 52]. Genes and QTLs have been identified that are related to abiotic stresses including postreproductive stage drought tolerance (stay-green) [53–56]; preharvest sprouting [57, 58], and aluminum tolerance [59]. Additional morphological characteristics have also been mapped in interspecific and/or intraspecific populations [21].

3.2. Expression profiling

Progress in characterization of the transcriptome has been paralleled by identification of differential gene expression in response to biotic and abiotic factors, including greenbug feeding [60], dehydration, high salinity and ABA [61], and methyl jasmonate, salicylic acid, and aminocyclopropane carboxylic acid treatments [62].

3.3. Association genetics

Much of the value of the sorghum sequence may be realized through better understanding of the levels and patterns of diversity in extant germ plasm, which can contribute both to functional analysis of specific sorghum genes and to deterministic improvement of sorghum for specific needs and environments. Sorghum is well suited to association mapping methods because of its medium-range patterns of linkage disequilibrium [63] and its self-pollinating mating system. Extensive *ex situ* sorghum germplasm collections exist within the U.S. National Plant Germplasm System and ICRISAT. Early characterization of complementary association genetics panels developed by a group of US scientists [6], and by Subprogram 1 of the Generation Challenge Program, is in progress. At present, more than 750 SSR alleles and 1402 SNP alleles discovered in 3.3 Mb of sequence [63–66] are freely available from the *Comparative Grass Genomics Center* relational database [67]. Extensive studies of sequence variation in sorghum show that haplotype diversity is low, even when nucleotide diversity is high: for regions of average length 671 bp surveyed in 17 accessions, the median number of haplotypes was three and the mode was two [63]. Common sequence variation can therefore be captured in a small sample of accessions.

3.4. Need for mutants and their characterization

A collection of ~400 *S. bicolor* mutants, now under the curation of C. Franks (USDA-ARS, Lubbock TX), provides a start toward testing hypotheses about the functions of individual genes, but a much broader set is needed, ideally providing for the identification of multiple loss-of-function mutants in each gene. Sorghum offers an opportunity to complement more extensive reverse genetics resources in for *Oryza* and *Zea*, providing for the study of genes/gene families that are less tractable in maize or rice (e.g., which remain duplicated in both taxa, but are single copy in sorghum), and also for targeting functional analyses to specific sorghum genes implicated in key traits by association genetics or other approaches.

To accelerate identification in a targeted manner of mutants useful to relate *Sorghum* genes to their functions, 1600 M3 annotated individually pedigreed mutagenized lines using ethyl methane sulfonate have been generated for sorghum genotype BTx623 and their preliminary characterization is in progress [68]. To date, every M3 row inspected closely has been distinguishable from the original stock, and many have multiple mutant phenotypes (Z. Xin, personal communication). More effort in this area is desirable.

Transposon tagging warrants further exploration as a means to obtain additional mutants in sorghum. *Cs1* is the first active transposable element isolated from sorghum, and offers several advantages as an insertion mutagen. *Cs1*-homologous sequences are present in low copy number in sorghum and other grasses, including sudangrass, maize, rice, teosinte, and sugarcane [69]. The low copy number and high transposition frequency of *Cs1* implies that this transposon could prove to be an efficient gene isolation tool. Preliminary studies of *Cs1* as a mutagen (S. Chopra, personal communication) indicate the feasibility of using this transposon as a tagging tool.

4. BEYOND SORGHUM-BROADER CHARACTERIZATION OF THE SACCHARINEAE

Sorghum sprung from the loins of the Saccharinae group of cereals, which also includes cultivated sugarcane and weedy/invasive Johnsongrass and *Microstegium*. This curious group shows a 6-fold variation in genome size among closely related species with the same chromosome number (*S, bicolor* and *propinquum* versus *nitidum*) [70]; an apparent reduction in chromosome number from the ancestral 20 to 10 in most parasorghums [71]; at least two chromosome doublings in *Saccharum* since its divergence from the remainder of the group [12]; and both natural (*Sorghum halepense*: [4]) and human-mediated polyploidization (*Saccharum* cultivars: [12]). Knowledge of the mechanisms, levels, and patterns of evolution of genome size and structure in this curious group will help to reveal the path by which the sorghum genome has arrived at its present state, also laying the foundation for further study of sugarcane and other economically important members of the group.

Of singular importance is the role that sorghum may play in clarifying the fates and consequences of genes duplicated in recent whole-genome duplications in *Saccharum*, and *Zea* (albeit not in the Saccharinae). *Zea* is the less complicated of these opportunities—a genomewide (or largely so) duplication in the *Zea* lineage shortly followed the *Sorghum-Zea* divergence [14, 72], making *Sorghum* an excellent outgroup for deducing the ancestral state at duplicated loci with regard to location, sequence, regulatory and other features. This opportunity is less complicated in that *Zea* is relatively advanced in restoration of the diploid state with regard to chromosome pairing, behaving for practical purposes as a diploid. *Saccharum* offers insight into an earlier stage following polyploid formation, behaving largely as an autopolyploid although with varying degrees of preferential pairing in different taxa and crosses [12, 73, 74]. *Sorghum halepense*, although far less well studied than either *Zea* or *Saccharum*, appears to be even closer to polyploid formation, in that its formation postdates the divergence of *S. bicolor* and *S. propinquum* which we roughly estimate to be 1-2 million years ago (based on $\sim$1.2% divergence of coding nucleotides). While it is very possible that these three polyploidizations differed in the degree of pairing specificity that was possible at the outset of polyploid evolution, insight into the relative degrees of duplicate gene loss, and/or silencing would be a valuable resource toward clarifying recent hypotheses about adaptation of genomes to the polyploid state [75].

ACKNOWLEDGMENT

The author thanks the USDA-CSREES, NSF Plant Genome Research Program, and International Consortium for Sugarcane Biotechnology for funding relevant aspects of his research, and the US Department of Energy Joint Genome Institute Community Sequencing Program for sequencing of the sorghum genome.

REFERENCES

[1] A. E. Farrell, R. J. Plevin, B. T. Turner, A. D. Jones, M. O'Hare, and D. M. Kammen, "Ethanol can contribute to energy and environmental goals," *Science*, vol. 311, no. 5760, pp. 506–508, 2006.

[2] P. Wagoner, "Perennial grain development: past efforts and potential for the future," *Critical Reviews in Plant Sciences*, vol. 9, no. 5, pp. 381–408, 1990.

[3] P. L. Scheinost, D. L. Lammer, X. Cai, T. D. Murray, and S. S. Jones, "Perennial wheat: the development of a sustainable cropping system for the U.S. Pacific Northwest," *American Journal of Alternative Agriculture*, vol. 16, no. 4, pp. 147–151, 2001.

[4] A. H. Paterson, K. F. Schertz, Y.-R. Lin, S.-C. Liu, and Y.-L. Chang, "The weediness of wild plants: molecular analysis of genes influencing dispersal and persistence of johnsongrass, *Sorghum halepense* (L.) Pers," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 92, no. 13, pp. 6127–6131, 1995.

[5] F. Y. Hu, D. Y. Tao, E. Sacks, et al., "Convergent evolution of perenniality in rice and sorghum," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 7, pp. 4050–4054, 2003.

[6] S. Kresovich, B. Barbazuk, J. A. Bedell, et al., "Toward sequencing the Sorghum genome. A U.S. National Science Foundation-sponsored workshop report," *Plant Physiology*, vol. 138, no. 4, pp. 1898–1902, 2005.

[7] L. G. Holm, D. L. Plucknett, J. V. Pancho, and J. P. Herberger, *The World's Worst Weeds: Distribution and Biology*, University Press of Hawaii, Honolulu, Hawaii, USA, 1977.

[8] B. S. Gaut, L. G. Clark, J. F. Wendel, and S. V. Muse, "Comparisons of the molecular evolutionary process at *rbcL* and *ndhF* in the grass family (Poaceae)," *Molecular Biology and Evolution*, vol. 14, no. 7, pp. 769–777, 1997.

[9] Z. Swigoňová, J. Lai, J. Ma, et al., "Close split of sorghum and maize genome progenitors," *Genome Research*, vol. 14, no. 10, pp. 1916–1923, 2004.

[10] A. H. Paterson, J. E. Bowers, and B. A. Chapman, "Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 26, pp. 9903–9908, 2004.

[11] B. W. S. Sobral, D. P. V. Braga, E. S. LaHood, and P. Keim, "Phylogenetic analysis of chloroplast restriction enzyme site mutations in the *Saccharinae* griseb. subtribe of the *Andropogoneae* Dumort. tribe," *Theoretical and Applied Genetics*, vol. 87, no. 7, pp. 843–853, 1994.

[12] R. Ming, S.-C. Liu, Y.-R. Lin, et al., "Detailed alignment of saccharum and sorghum chromosomes: comparative organization of closely related diploid and polyploid genomes," *Genetics*, vol. 150, no. 4, pp. 1663–1682, 1998.

[13] J. M. J. de Wet, S. C. Gupta, J. R. Harlan, and C. O. Grassl, "Cytogenetics of introgression from *Saccharum* into *Sorghum*," *Crop Science*, vol. 16, pp. 568–572, 1976.

[14] Z. Swigoňová, J. Lai, J. Ma, et al., "On the tetraploid origin of the maize genome," *Comparative and Functional Genomics*, vol. 5, no. 3, pp. 281–284, 2004.

[15] G.-W. Xu, C. W. Magill, K. F. Schertz, and G. E. Hart, "A RFLP linkage map of *Sorghum bicolor* (L.) Moench," *Theoretical and Applied Genetics*, vol. 89, no. 2-3, pp. 139–145, 1994.

[16] D. Bhattramakki, J. Dong, A. K. Chhabra, and G. E. Hart, "An integrated SSR and RFLP linkage map of *Sorghum bicolor* (L.) Moench," *Genome*, vol. 43, no. 6, pp. 988–1002, 2000.

[17] P. E. Klein, R. R. Klein, S. W. Cartinhour, et al., "A high-throughput AFLP-based method for constructing integrated genetic and physical maps: progress toward a sorghum genome map," *Genome Research*, vol. 10, no. 6, pp. 789–807, 2000.

[18] M. A. Menz, R. R. Klein, J. E. Mullet, J. A. Obert, N. C. Unruh, and P. E. Klein, "A high-density genetic map of *Sorghum bicolor* (L.) Moench based on 2926 AFLP®, RFLP and SSR markers," *Plant Molecular Biology*, vol. 48, no. 5-6, pp. 483–499, 2002.

[19] L. M. Chittenden, K. F. Schertz, Y.-R. Lin, R. A. Wing, and A. H. Paterson, "A detailed RFLP map of *Sorghum bicolor* X *S.propinquum*, suitable for high-density mapping, suggests ancestral duplication of *Sorghum* chromosomes or chromosomal segments," *Theoretical and Applied Genetics*, vol. 87, no. 8, pp. 925–933, 1994.

[20] J. E. Bowers, C. Abbey, S. Anderson, et al., "A high-density genetic recombination map of sequence-tagged sites for *Sorghum*, as a framework for comparative structural and evolutionary genomics of tropical grains and grasses," *Genetics*, vol. 165, no. 1, pp. 367–386, 2003.

[21] F. A. Feltus, G. E. Hart, K. F. Schertz, et al., "Genetic map alignment and QTL correspondence between inter- and intra-specific sorghum populations," *Theoretical and Applied Genetics*, vol. 112, no. 7, pp. 1295–1305, 2006.

[22] J.-S. Kim, P. E. Klein, R. R. Klein, H. J. Price, J. E. Mullet, and D. M. Stelly, "Chromosome identification and nomenclature of *Sorghum bicolor*," *Genetics*, vol. 169, no. 2, pp. 1169–1173, 2005.

[23] A. H. Paterson, Y.-R. Lin, Z. Li, et al., "Convergent domestication of cereal crops by independent mutations at corresponding genetic loci," *Science*, vol. 269, no. 5231, pp. 1714–1718, 1995.

[24] R. Whitkus, J. Doebley, and M. Lee, "Comparative genetic mapping of sorghum and maize," *Genetics*, vol. 132, pp. 1119–1130, 1992.

[25] P. Dufour, M. Deu, L. Grivet, et al., "Construction of a composite sorghum genome map and comparison with sugarcane, a related complex polyploid," *Theoretical and Applied Genetics*, vol. 94, no. 3-4, pp. 409–418, 1997.

[26] R. W. Jessup, B. L. Burson, G. Burow, et al., "Segmental allotetraploidy and allelic interactions in buffelgrass (*Pennisetum ciliare* (L.) Link syn. *Cenchrus ciliaris* L.) as revealed by genome mapping," *Genome*, vol. 46, no. 2, pp. 304–313, 2003.

[27] A. M. Missaoui, A. H. Paterson, and J. H. Bouton, "Investigation of genomic organization in switchgrass (*Panicum virgatum* L.) using DNA markers," *Theoretical and Applied Genetics*, vol. 110, no. 8, pp. 1372–1383, 2005.

[28] C. M. Bethel, E. B. Sciara, J. C. Estill, J. E. Bowers, W. Hanna, and A. H. Paterson, "A framework linkage map of bermudagrass (*Cynodon dactylon* x *transvaalensis*) based on single-dose restriction fragments," *Theoretical and Applied Genetics*, vol. 112, no. 4, pp. 727–737, 2006.

[29] S.-S. Woo, J. Jiang, B. S. Gill, A. H. Paterson, and R. A. Wing, "Construction and characterization of a bacterial artificial chromosome library of *Sorghum bicolor*," *Nucleic Acids Research*, vol. 22, no. 23, pp. 4922–4931, 1994.

[30] D. G. Peterson, S. R. Schulze, E. B. Sciara, et al., "Integration of cot analysis, DNA cloning, and high-throughput sequencing facilitates genome characterization and gene discovery," *Genome Research*, vol. 12, no. 5, pp. 795–807, 2002.

[31] K. Arumuganathan and E. Earle, "Estimation of nuclear DNA content of plants by flow cytometry," *Plant Molecular Biology Reporter*, vol. 9, no. 3, pp. 208–218, 1991.

[32] J. E. Bowers, M. A. Arias, R. Asher, et al., "Comparative physical mapping links conservation of microsynteny to chromosome structure and recombination in grasses," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 37, pp. 13206–13211, 2005.

[33] L. H. Pratt, C. Liang, M. Shah, et al., "Sorghum expressed sequence tags identify signature genes for drought, pathogenesis, and skotomorphogenesis from a milestone set of 16,801 unique transcripts," *Plant Physiology*, vol. 139, no. 2, pp. 869–884, 2005.

[34] J. A. Bedell, M. A. Budiman, A. Nunberg, et al., "Sorghum genome sequencing by methylation filtration," *PLoS Biology*, vol. 3, no. 1, p. e13, 2005.

[35] Y.-R. Lin, K. F. Schertz, and A. H. Paterson, "Comparative analysis of QTLs affecting plant height and maturity across the poaceae, in reference to an interspecific sorghum population," *Genetics*, vol. 141, no. 1, pp. 391–411, 1995.

[36] P. E. Ulanch, K. L. Childs, P. W. Morgan, and J. E. Mullet, "Molecular markers linked to Ma(1) in sorghum," *Plant Physiology*, vol. 111, p. 709, 1996.

[37] R. R. Klein, P. E. Klein, A. K. Chhabra, et al., "Molecular mapping of the *rf1* gene for pollen fertility restoration in sorghum (*Sorghum bicolor* L.)," *Theoretical and Applied Genetics*, vol. 102, no. 8, pp. 1206–1212, 2001.

[38] L. Wen, H. V. Tang, W. Chen, et al., "Development and mapping of AFLP markers linked to the sorghum fertility restorer gene *rf4*," *Theoretical and Applied Genetics*, vol. 104, no. 4, pp. 577–585, 2002.

[39] R. R. Klein, P. E. Klein, J. E. Mullet, P. Minx, W. L. Rooney, and K. F. Schertz, "Fertility restorer locus *Rf1* of sorghum (*Sorghum bicolor* L.) encodes a pentatricopeptide repeat protein not present in the colinear region of rice chromosome 12," *Theoretical and Applied Genetics*, vol. 111, no. 6, pp. 994–1012, 2005.

[40] C. S. Mutengwa, P. B. Tongoona, and I. Sithole-Niang, "Genetic studies and a search for molecular markers that are linked to *Striga asiatica* resistance in sorghum," *African Journal of Biotechnology*, vol. 4, no. 12, pp. 1355–1361, 2005.

[41] C. L. McIntyre, R. E. Casu, J. Drenth, et al., "Resistance gene analogues in sugarcane and sorghum and their association with quantitative trait loci for rust resistance," *Genome*, vol. 48, no. 3, pp. 391–400, 2005.

[42] C. L. McIntyre, S. M. Hermann, R. E. Casu, et al., "Homologues of the maize rust resistance gene *Rp1-D* are genetically associated with a major rust resistance QTL in sorghum," *Theoretical and Applied Genetics*, vol. 109, no. 4, pp. 875–883, 2004.

[43] D. S. Multani, R. B. Meeley, A. H. Paterson, J. Gray, S. P. Briggs, and G. S. Johal, "Plant-pathogen microevolution: molecular basis for the origin of a fungal disease in maize," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, no. 4, pp. 1686–1691, 1998.

[44] A. S. Totad, B. Fakrudin, and M. S. Kuruvinashetti, "Isolation and characterization of resistance gene analogs (RGAs) from sorghum (*Sorghum bicolor* L. Moench)," *Euphytica*, vol. 143, no. 1-2, pp. 179–188, 2005.

[45] M. Singh, K. Chaudhary, H. R. Singal, C. W. Magill, and K. S. Boora, "Identification and characterization of RAPD and SCAR markers linked to anthracnose resistance gene in sorghum [*Sorghum bicolor* (L.) Moench]," *Euphytica*, vol. 149, no. 1-2, pp. 179–187, 2006.

[46] M. L. Wang, R. Dean, J. Erpelding, and G. Pederson, "Molecular genetic evaluation of sorghum germplasm differing in response to fungal diseases: rust (*Puccinia purpurea*) and anthracnose (*Collectotrichum graminicola*)," *Euphytica*, vol. 148, no. 3, pp. 319–330, 2006.

[47] Y. Z. Tao, D. R. Jordan, R. G. Henzell, and C. L. McIntyre, "Identification of genomic regions for rust resistance in sorghum," *Euphytica*, vol. 103, no. 3, pp. 287–292, 1998.

[48] C. S. Katsar, R. H. Paterson, G. L. Teetes, and G. C. Peterson, "Molecular analysis of sorghum resistance to the greenbug (Homoptera: Aphididae)," *Journal of Economic Entomology*, vol. 95, no. 2, pp. 448–457, 2002.

[49] H. Agrama, G. Widle, J. Reese, L. Campbell, and M. Tuinstra, "Genetic mapping of QTLs associated with greenbug resistance and tolerance in *Sorghum bicolor*," *Theoretical and Applied Genetics*, vol. 104, no. 8, pp. 1373–1378, 2002.

[50] N. Nagaraj, J. C. Reese, M. R. Tuinstra, et al., "Molecular mapping of sorghum genes expressing tolerance to damage by greenbug (Homoptera: Aphididae)," *Journal of Economic Entomology*, vol. 98, no. 2, pp. 595–602, 2005.

[51] Y. Z. Tao, A. Hardy, J. Drenth, et al., "Identifications of two different mechanisms for sorghum midge resistance through QTL mapping," *Theoretical and Applied Genetics*, vol. 107, no. 1, pp. 116–122, 2003.

[52] B. I. G. Haussmann, D. E. Hess, G. O. Omanya, et al., "Genomic regions influencing resistance to the parasitic weed *Striga hermonthica* in two recombinant inbred populations of sorghum," *Theoretical and Applied Genetics*, vol. 109, no. 5, pp. 1005–1016, 2004.

[53] W. Xu, P. K. Subudhi, O. R. Crasta, D. T. Rosenow, J. E. Mullet, and H. T. Nguyen, "Molecular mapping of QTLs conferring stay-green in grain sorghum (*Sorghum bicolor* L. Moench)," *Genome*, vol. 43, no. 3, pp. 461–469, 2000.

[54] P. K. Subudhi, D. T. Rosenow, and H. T. Nguyen, "Quantitative trait loci for the stay green trait in sorghum (*Sorghum bicolor* L. Moench): consistency across genetic backgrounds and environments," *Theoretical and Applied Genetics*, vol. 101, no. 5-6, pp. 733–741, 2000.

[55] O. R. Crasta, W. Xu, D. T. Rosenow, J. Mullet, and H. T. Nguyen, "Mapping of post flowering drought resistance traits in grain sorghum: association between QTLs influencing premature senescence and maturity," *Molecular and General Genetics*, vol. 262, no. 3, pp. 579–588, 1999.

[56] B. I. G. Haussmann, V. Mahalakshmi, B. V. S. Reddy, N. Seetharama, C. T. Hash, and H. H. Geiger, "QTL mapping of stay-green in two sorghum recombinant inbred populations," *Theoretical and Applied Genetics*, vol. 106, no. 1, pp. 133–142, 2002.

[57] F. Carrari, R. Benech-Arnold, R. Osuna-Fernandez, et al., "Genetic mapping of the *Sorghum bicolor vp1* gene and its relationship with preharvest sprouting resistance," *Genome*, vol. 46, no. 2, pp. 253–258, 2003.

[58] D. Lijavetzky, M. C. Martínez, F. Carrari, and H. E. Hopp, "QTL analysis and mapping of pre-harvest sprouting resistance in sorghum," *Euphytica*, vol. 112, no. 2, pp. 125–135, 2000.

[59] J. V. Magalhaes, D. F. Garvin, Y. Wang, et al., "Comparative mapping of a major aluminum tolerance gene in sorghum and other species in the Poaceae," *Genetics*, vol. 167, no. 4, pp. 1905–1914, 2004.

[60] S.-J. Park, Y. Huang, and P. Ayoubi, "Identification of expression profiles of sorghum genes in response to greenbug phloem-feeding using cDNA subtraction and microarray analysis," *Planta*, vol. 223, no. 5, pp. 932–947, 2006.

[61] C. D. Buchanan, S. Lim, R. A. Salzman, et al., "*Sorghum bicolor*'s transcriptome response to dehydration, high salinity and ABA," *Plant Molecular Biology*, vol. 58, no. 5, pp. 699–720, 2005.

[62] R. A. Salzman, J. A. Brady, S. A. Finlayson, et al., "Transcriptional profiling of sorghum induced by methyl jasmonate, salicylic acid, and aminocyclopropane carboxylic acid reveals cooperative regulation and novel gene responses," *Plant Physiology*, vol. 138, no. 1, pp. 352–368, 2005.

[63] M. T. Hamblin, M. G. Salas Fernandez, A. M. Casa, S. E. Mitchell, A. H. Paterson, and S. Kresovich, "Equilibrium processes cannot explain high levels of short- and medium-range linkage disequilibrium in the domesticated grass *Sorghum bicolor*," *Genetics*, vol. 171, no. 3, pp. 1247–1256, 2005.

[64] S. J. Schloss, S. E. Mitchell, G. M. White, et al., "Characterization of RFLP probe sequences for gene discovery and SSR development in *Sorghum bicolor* (L.) Moench," *Theoretical and Applied Genetics*, vol. 105, no. 6-7, pp. 912–920, 2002.

[65] M. T. Hamblin, S. E. Mitchell, G. M. White, et al., "Comparative population genetics of the panicoid grasses: sequence polymorphism, linkage disequilibrium and selection in a diverse sample of *Sorghum bicolor*," *Genetics*, vol. 167, no. 1, pp. 471–483, 2004.

[66] A. M. Casa, S. E. Mitchell, M. T. Hamblin, et al., "Diversity and selection in sorghum: simultaneous analyses using simple sequence repeats," *Theoretical and Applied Genetics*, vol. 111, no. 1, pp. 23–30, 2005.

[67] A. R. Gingle, Y. Huang, H. Yang, J. Bowers, S. Kresovich, and A. H. Paterson, "CGGC: an integrated web resource for sorghum," submitted to *Plant Physiology*.

[68] Z. Xin, M. Wang, N. Barkley Jr, et al., "Development of a tilling population for sorghum functional genomics," in *Proceedings of the 15th International Plant & Animal Genome Conference*, San Diego, Calif, USA, January 2007.

[69] S. Chopra, V. Brendel, J. Zhang, J. D. Axtell, and T. Peterson, "Molecular characterization of a mutable pigmentation phenotype and isolation of the first active transposable element from *Sorghum bicolor*," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96, no. 26, pp. 15330–15335, 1999.

[70] H. J. Price, S. L. Dillon, G. Hodnett, W. L. Rooney, L. Ross, and J. S. Johnston, "Genome evolution in the genus *Sorghum* (Poaceae)," *Annals of Botany*, vol. 95, no. 1, pp. 219–227, 2005.

[71] R. Spangler, B. Zaitchik, E. Russo, and E. Kellogg, "Andropogoneae evolution and generic limits in *Sorghum* (Poaceae) using *ndhF* sequences," *Systematic Botany*, vol. 24, no. 2, pp. 267–281, 1999.

[72] Z. Swigoňová, J. Lai, J. Ma, W. Ramakrishna, V. Llaca, J. L. Bennetzen, and J. Messing, "Close split of sorghum and maize genome progenitors," *Genome Research*, vol. 14, no. 10, pp. 1916–1923, 2004.

[73] S. M. Al-Janabi, R. J. Honeycutt, M. McClelland, and B. W. S. Sobral, "A genetic linkage map of *Saccharum spontaneum* L. 'SES 208'," *Genetics*, vol. 134, no. 4, pp. 1249–1260, 1993.

[74] L. Grivet, A. D'Hont, D. Roques, P. Feldmann, C. Lanaud, and J. C. Glaszmann, "RFLP mapping in cultivated sugarcane (*Saccharum* spp.): genome organization in a highly polyploid and aneuploid interspecific hybrid," *Genetics*, vol. 142, no. 3, pp. 987–1000, 1996.

[75] A. H. Paterson, B. A. Chapman, J. C. Kissinger, J. E. Bowers, F. A. Feltus, and J. C. Estill, "Convergent retention or loss of gene/domain families following independent whole-genome duplication events in *Arabidopsis, Oryza, Saccharomyces*, and *Tetraodon*," *Trends in Genetics*, vol. 22, no. 11, pp. 597–602, 2006.

Hindawi Publishing Corporation
International Journal of Plant Genomics
Volume 2008, Article ID 536104, 7 pages
doi:10.1155/2008/536104

Review Article

Brachypodium Genomics

Bahar Sogutmaz Ozdemir,[1] Pilar Hernandez,[2] Ertugrul Filiz,[1] and Hikmet Budak[1]

[1] *Biological Science and Bioengineering Program, Faculty of Engineering and Natural Sciences, Sabanci University Orhanli, 34956 Tuzla-Istanbul, Turkey*
[2] *Institute for Sustainable Agriculture (IAS), Spanish National Research Council (CSIC), Alameda del Obispo s/n, Apartado 4084, 14080 Cordoba, Spain*

Correspondence should be addressed to Hikmet Budak, budak@sabanciuniv.edu

Received 19 July 2007; Accepted 25 November 2007

Recommended by P. K. Gupta

Brachypodium distachyon (L.) Beauv. is a temperate wild grass species; its morphological and genomic characteristics make it a model system when compared to many other grass species. It has a small genome, short growth cycle, self-fertility, many diploid accessions, and simple growth requirements. In addition, it is phylogenetically close to economically important crops, like wheat and barley, and several potential biofuel grasses. It exhibits agricultural traits similar to those of these target crops. For cereal genomes, it is a better model than *Arabidopsis thaliana* and *Oryza sativa* (rice), the former used as a model for all flowering plants and the latter hitherto used as model for genomes of all temperate grass species including major cereals like barley and wheat. Increasing interest in this species has resulted in the development of a series of genomics resources, including nuclear sequences and BAC/EST libraries, together with the collection and characterization of other genetic resources. It is expected that the use of this model will allow rapid advances in generation of genomics information for the improvement of all temperate crops, particularly the cereals.

1. INTRODUCTION

Brachypodium P. Beauv (from the Greek *brachys* “short” and *podion* “a little foot,” referring to its subsessile spikelets, [1]) is a genus representing some temperate wild grass species. In particular, *Brachypodium distachyon* (L.) Beauv., also described as “purple false broom,” has recently emerged as a new model plant for the diverse and economically important group of temperate grasses and herbaceous energy crops [2]. Temperate crops such as wheat, barley, and forage grasses are the basis for the food and feed supply. However, the size and complexity of their genomes are major barriers to genomics research and molecular breeding. Similarly, although the herbaceous energy crops (especially grasses) are becoming novel target sources of renewable energy, very little is known about the biological basis underlying their bioenergy traits. Therefore, there is a growing need for a temperate grass model to address questions directly relevant both for improving grain crops and forage grasses that are indispensable to our food/feed production systems, and for developing grasses into superior energy crops. The present status of genomics research conducted in this model grass species is briefly summarized in this review.

2. *BRACHYPODIUM* GENOMICS AS A MODEL SYSTEM

2.1. *Desirable attributes*

B. distachyon has many attributes that make it a suitable model for conducting functional genomics research not only among cereal crops like wheat and barley, but also for biofuel crops like Switchgrass [2]. Due to its small haploid genome (~355 Mbp) and availability of a polyploid series with a basic chromosome number of $x = 5$, ($2n = 2x = 10$), the diploid race of *B. distachyon* can be used as a model for the much larger polyploid genomes of crops such as bread wheat (16979 Mbp, $2n = 6x = 42$), durum wheat (12030 Mbp, $2n = 4x = 28$), and barley (5439 Mbp, $2n = 2x = 14$) (all C-values from [3]). Besides its small genome size, other desirable attributes include a small physical stature (approximately 20 cm), self-fertility, lack of seed-shattering, a short lifecycle that is normally completed within 11–18 weeks depending on the vernalization requirement [2] (might be as fast as 8 weeks under optimized conditions, [4]), and simple growth requirements with large planting density and easy genetic transformation [4, 5]. This combination of desirable attributes, together with the biological similarities with its target crops, is responsible for the recent research interest in this species. A few ecotypes of this taxon collected from diverse geographic regions of Turkey are shown in Figure 1, indicating a high level of variation among different accessions.

Brachypodium species range from annuals to strongly rhizomatous perennials that exhibit breeding systems ranging from strictly inbreeding to highly self-incompatible [6]. Some of the characteristic features of the genus *Brachypodium* include the following [17]: (i) hairy terminal ovary appendage, (ii) the single starch grains, (iii) the outermost thick layer of the nucellus, (iv) the long narrow caryopsis, (v) spicate or racemose inflorescences, and (vi) hairy nodules [7].

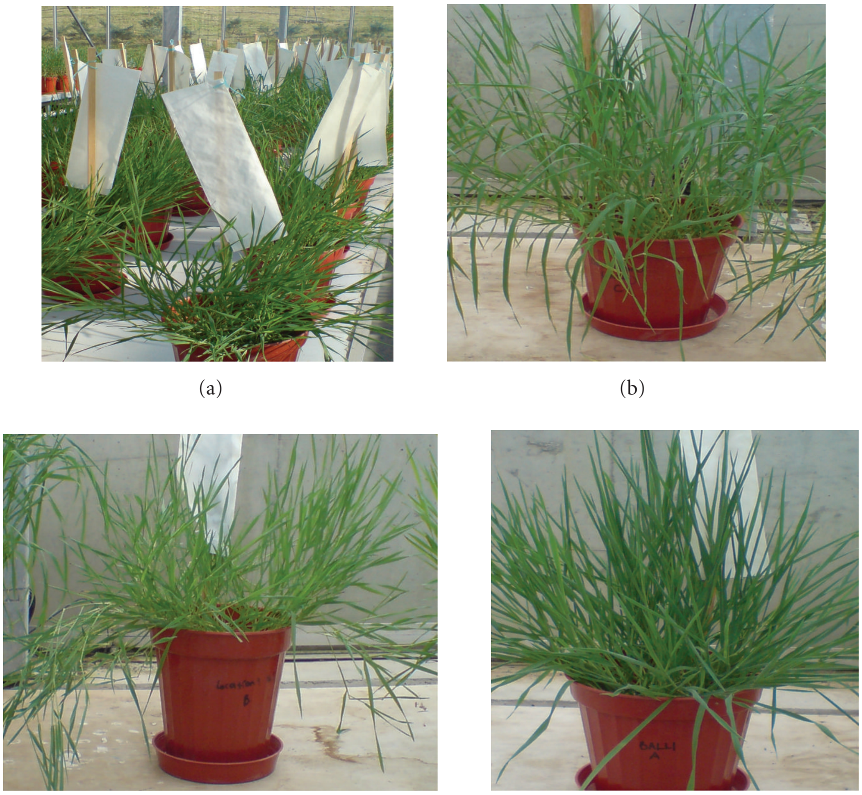
2.2. Brachypodium as a model system: a comparison with Arabidopsis and Oryza

The available genome sequences of the model plants *Arabidopsis* [9] and rice [10] are considered to be the major resources for plant genomics research. Nevertheless, these model species are not suitable for the functional genomics studies of temperate grasses. *Arabidopsis* has all the desirable attributes for a model plant: it is small in size, grows easily and quickly (reaching maturity in 6 weeks), has a small diploid genome, is self-compatible and easily transformable. Its utility as a model system has been proven by the wealth of genomic discoveries, useful for a broad range of crops (including cereals) it has generated. However, as a dicot species, it does not share with grass crops most of the biological features related to agricultural traits and in this sense, rice would provide a better alternative. The rice plant, however, does not fulfill the requirements of short size, rapid life cycle, inbreeding reproductive strategy, simple growth requirements, or easy transformation, thus imposing practical limitations. As a tropical species, it does not display all agronomic traits that are relevant to temperate grasses, especially to forage grasses; these agronomic traits include resistance to specific pathogens, freezing tolerance, vernalization, perenniality, injury tolerance, meristem dormancy mechanisms, mycorrhizae, sward ecology, or postharvest biochemistry of silage [2]. Moreover, rice is phylogenetically distant from the *Pooidae* subfamily that includes wheat, barley, and temperate grasses [11], whereas *Brachypodium* diverged from the ancestral *Pooidae* clade immediately prior to the radiation of the modern “core pooids” (*Triticeae*, *Bromeae*, and *Avenae*), which include majority of important temperate cereals and forage grasses [8]. Based upon cytological, anatomical, and physiological studies, *Brachypodium* is placed into its own tribe *Brachypodieae* of the Poaceae family [12]. In fact, the perennial outbreeding species, *B. sylvaticum* ($2n = 18$) was considered suitable for study of archetypal grass centromere sequences, which allowed detection of repetitive DNA sequences that are conserved among wheat, maize, rice, and *Brachypodium* [13]. Several species of this genus were studied using combined sequences of chloroplast *ndh*F gene and nuclear ITS to reconstruct phylogeny among these selected species within the genus [8]. Similarly, RFLPs and RAPDs were used for nuclear genome analysis to establish the evolutionary position of the genus. The genus *Brachypodium* constitutes morphologically more or less closely resembling species that are native to different ecological regions such that *B. distachyon* is the Mediterranean annual, nonrhizomatous *B. mexicanum* is from the New World, *B. pinnatum* and *B. sylvaticum* are Eurasian, and *B. rupestre* is a European taxon [7]. In view of the above, *B. distachyon* has been proposed as an alternative model for functional genomics of temperate grasses [2].

2.3. Brachypodium and the tribe Triticeae

The tribe *Triticeae* Dumort belongs to the grass family, Poaceae, and constitutes one of the economically most important plant groups. It includes three major cereal crops, wheat, barley, and rye (belonging to the genera *Triticum*, *Hordeum*, and *Secale*, resp.), which are traditionally cultivated in the temperate zone. The tribe has basic chromosome number of seven, and contains taxa ranging from diploids ($2n = 2x = 14$) to duodecaploids ($2n = 12x = 84$), including all intermediate ploidy levels [14]. Polyploidy (the most important cytogenetic process in higher plants, [15]) and more specifically, allopolyploidy, has played and plays a main role in the tribe's evolution. With around 350 species, crossability barriers are poorly understood, and it is remarkable how its species, even species in different genera, can be made to hybridize even if they do not hybridize naturally. Therefore, the tribe also includes man-made crops such as ×*Triticosecale* (triticale) and ×*Tritordeum* (an amphiploid of *Triticum aestivum* ×*Hordeum chilense* [16, 17] and *Triticum aestivum* × *Leymus arenarius* [18]). It has also been possible to apply interspecific and intergeneric hybridization to increase the genetic variability of crops belonging to this tribe (mainly wheat, [19, 20]).

As mentioned earlier, the chromosome numbers within *B. distachyon* accessions range from 10 to 30 [21], and the haploid genome size in diploid *Brachypodium* ($2n = 2x = 10$) varies from 172 Mbp to 355 Mbp [2, 22], although the former value may be an underestimate [22]; the genome size is thus assumed to be approximately 355 Mbp. Therefore, within Poaceae, *B. distachyon* carries one of the smallest genomes, which is intermediate between the genomes of *Arabidopsis thaliana* with 157 Mbp, and rice with 490 Mbp (All C-values from [3]). These data are consistent with previous reports that species of *Brachypodium* have the smallest 5S rDNA spacer among the grasses, and contain less than 15% highly repetitive DNA [7]. Additionally, GISH analysis of somatic chromosomes has shown preponderance of repetitive DNA in the pericentromeric regions, reflecting the compactness and economy of this genome [12]. This study also revealed structural uniformity of the diploid accessions ABR1 and ABR5, confirming their status as model genotypes, from which two BAC libraries have recently been prepared for functional genomics analysis [23]. Recent analysis of BAC end sequences (BESs) also corroborates this unusually compact genome [24, 25]. The other accessions having



(a) (b) (c) (d)

Figure 1: The *Brachypodium distachyon* lines grown under greenhouse conditions. (a) Seed heads are covered to prevent crosspollination in case it persists. (b), (c), and (d) seeds were collected from a diverse geographic region of Turkey.

chromosome numbers in multiples of 10 suggested that this species has evolved as a polyploid series based upon $2n = 2x = 10$, and that ecotypes that deviate from multiples of 10 evolved due to aneuploid or dysploid changes in chromosome number [21]. A cytotaxonomic analysis of the members of the polyploid series has revealed hybrid origin of several of the polyploid genotypes, suggesting a complex evolution of this species that is not entirely based on chromosome doubling [12]. For example, allotetraploid artificial hybrids between *B. distachyon* and *B. sylvaticum* exhibited irregular meiosis and infertility [6], although allotetraploids were fertile, one of them (ABR100) showed normal meiosis [12]. This indicates that either the constituent diploids of this allotetraploid are more compatible in hybrids, or the hybrids themselves have evolved pairing control mechanisms similar to those of wheat and other allopolyploids. It has also been shown using GISH that the genomes of the constituent diploids remain separated in the allotetraploid, and that there is no recombination between homoeologous chromosomes. These features make the natural polyploid hybrids within the genus *Brachypodium* a suitable material for the isolation and characterization of diplodizing genes.

For the reasons stated above, the whole tribe Triticeae is considered to be an enormous gene pool for crop improvement, deserving efforts not only for its morphological, physiological, genetic, and genomic characterization, but also for the establishment of phylogenetic relationships among different species of the tribe. The large and complex genomes of some members of the tribe are a main constraint for genomics research within this tribe, which would be greatly facilitated with the availability of a suitable model species like *B. distachyon*.

3. CURRENT STATUS OF *BRACHYPODIUM* GENOMICS

3.1. Development of inbred lines

Inbred lines make an important resource for genomics research. Keeping this in view, diploid inbred lines have been developed in *B. distachyon* by selfing [4] as well as through selection from segregating populations derived from crosses among diploid ecotypes [26].

3.2. Development of transformation and regeneration protocols

An efficient transformation procedure and an optimized plant regeneration protocol have been developed in *B. distachyon*. For instance, in a study reported in 1995, callus induction and plant regeneration from mature embryos, as well as *in vitro* clonal propagation of shoots were successfully achieved in *B. distachyon* [27]. In our own studies also,

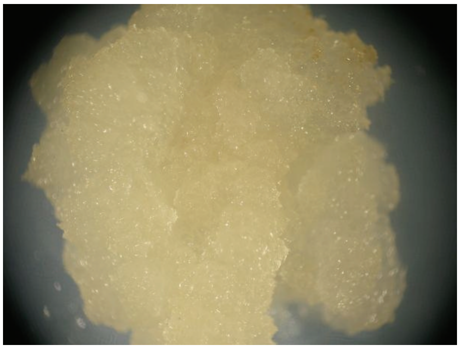


Figure 2: Callus induction of mature *Brachypodium distachyon* embryos supplemented with 5 mg/L 2,4-D (2,4-dichlorophenoxyacetic acid).

efficient callus formation from mature embryos of *B. distachyon* was successfully achieved using MS basal medium supplemented with sucrose and 2,4-dichlorophenoxyacetic (2,4-D) acid at a concentration ranging from 2.5 to 5 mg\L (see Figure 2). The results are suggesting that in *Brachypodium* species, higher rates of callus induction can be achieved through the use of (i) MS basal medium [43] rather than LS basal medium [44] (ii) sucrose rather than maltose, and (iii) higher concentrations of auxin as a plant growth regulator (unpublished data).

Agrobacterium-mediated transformation involving insertions of single genes has also been achieved in several genotypes of *B. distachyon* (including diploid and polyploid taxa), giving T_1 transgenic plants [4]. These transformation studies also involved the diploid genotype Bd21, which has also been used for construction of BAC libraries [28] and for generating 20440 ESTs [29]. *Agrobacterium*-mediated transformation was successful in 10 out of the 19 lines, with efficiencies ranging from 0.4% to 15% [4]. Embryogenic calli derived from immature embryos were also transformed through biolistic transformation leading to transgene expression in T_1 progeny [2, 5]. In later study, transformation with an average efficiency of 5.3% was achieved. In this study, testing of T_0 as well as T_1 generations and seed production in T_2 was achieved within one year due to the short life cycle of *B. distachyon* [5], confirming its importance as a model plant species.

3.3. BAC libraries and expression sequence tags (ESTs)

BAC libraries of two diploid ecotypes of *B. distachyon*, ABR1 and ABR5, have also been constructed and have been used to determine synteny among rice, *Brachypodium*, and other species of Poaceae family. For this purpose, BACs were marker-selected (BAC landing) using primers designed according to previously mapped rice and Poaceae sequences. Most BACs hybridized as single loci in known *Brachypodium* chromosomes, whereas contiguous BACs colocalized on individual chromosomes, thus confirming conservation of genome synteny [23].

3.4. Mutagenesis

Mutagenesis with sodium azide was also successful in *B. distachyon* although response to this mutagen differed among different accessions [30]. The results obtained were comparable with those earlier obtained in barley and rice under higher concentrations of mutagens. Application of ethylmethane sulphonate (EMS) is currently on the way in diploid *Brachypodium* accessions.

4. *BRACHYPODIUM* GENOMES: ADVANCES ON THE WAY

4.1. BAC-based physical maps

A BAC-based physical map of *B. distachyon* is being developed at the John Innes Centre (Norwich, UK) as an aid to the international effort to make BAC-based physical maps of the genomes of Chinese Spring bread wheat [31]. Since establishing a physical map of the genome of bread wheat, one of the most important crops worldwide, is a major challenge due to the enormous size of the genome and its hexaploid constitution, it is expected that the availability of a *Brachypodium* physical map will greatly facilitate this task. The close phylogenetic relationship of *Brachypodium* to wheat leads to high similarity in gene sequences. Unambiguous hybridization signals are also generated, when *Brachypodium* probes are used on wheat BAC filters and southern blots. Preliminary experiments have also shown that it is feasible to anchor *Brachypodium* BACs to the rice genome by BES to create an outline physical map. An outline physical map of *B. distachyon* genotype, Bd3-1 using BES and fingerprinting, is being established and will be used to start assembling contigs in wheat chromosome groups [31].

Another *B. distachyon* physical map is being developed at the University of California and US Department of Agriculture (USDA) [25] by using two BAC libraries constructed from *B. distachyon* genotype, BD-21. These BACs are being fingerprinted using snapshot-based fingerprinting. This physical map of *B. distachyon* will also be integrated with BES, again providing genome-wide *Brachypodium* resources for sequence assembly, comparative genome analysis, gene isolation, and functional genomics analysis.

4.2. B. distachyon genome and retrotransposons

The genome of *B. distachyon* is also being examined for the presence, diversity, and distribution of the major classes of plant transposable elements, particularly the retrotransposons [32], since retrotransposons comprise most of the existing DNA between genes in the large cereal genomes. The compact genome of *B. distachyon* contains relatively few retrotransposons, which include copia, gypsy, TRIM, and LARD groups of elements. The availability of retrotransposon sequences will facilitate the development of retrotransposon-based molecular markers like IRAP, REMAP, SSAP, and RBIP markers, which have a variety of applications.

4.3. Genetic linkage maps

A genetic map of *B. distachyon* genotype, Bd21 is being developed by the International *Brachypodium* Initiative [33]. Genetic maps will provide anchor points linking the genome of *Brachypodium* with those of rice, wheat, and some biofuel crops, and will establish chromosome-scale physical maps of BACs for whole genome sequencing. In order to develop these genetic maps, mapping populations are being developed, which currently comprise several hundred F_2 lines derived from the cross Bd21 × Bd3-1. These will be advanced to F6 to establish RILs that can serve as a common mapping resource for the community. Several approaches have been used to identify polymorphisms between parents of the mapping population. First, conserved orthologous sequence (COS) markers derived from wheat and millet were used to identify a set of 80 confirmed polymorphisms between these two parental lines (Bd21, Bd3-1). Another strategy was the use of ESTs derived from Bd21 in order to identify additional polymorphisms [29]. The most productive approach has been to predict introns in *Brachypodium* genes, based on a comparison of Bd21 ESTs with the annotated rice genome sequence; nearly all primers designed from predicted introns gave amplified products in PCR reactions. Most markers developed thus were polymorphic among the 5 diploid inbred lines used for testing, and thus proved to be useful markers for genetic mapping.

4.4. Whole genome sequencing

The *Brachypodium* nuclear genome is currently being sequenced within a project that was funded in early 2006 by the US Department of Energy (DOE). A draft genome sequence is expected to be completed by the end of 2007. This project is generating a whole-genome shotgun sequence of *B. distachyon* genotype, Bd21 genome, and is coupled with another project aimed at generating nearly 250.000 ESTs. Data from both projects will be made publicly available through an online database (BrachyBase at http://www.brachybase.org) and a community-dedicated portal (http://www.brachypodium.org). BrachyBase will enable efficient exploitation of genome and transcriptome sequences to identify genes underlying traits and will facilitate comparisons with other grass genomes [34].

Generation and analysis of over 60 000 BES from large-insert BAC clones has provided the first view of *Brachypodium* genome composition, structure, and organization [35]. In this study, ~10% of the BES show similarity to known repetitive DNA sequences in existing databases, whereas ~40% matched sequences in the EST database, which suggests that a considerable portion of the *Brachypodium* genome is transcribed. Gene-related BESs that were identified for the *Brachypodium* genome were also aligned *in silico* to the rice genome sequences. On the basis of gene colinearity between *Brachypodium* and rice, conserved and diverged regions were identified. BES with significant matches to wheat ESTs that have been mapped to individual chromosome and bin positions were also identified. These BACs represent regions that are colinear with mapped ESTs and will be useful in identifying additional markers for specific regions of wheat chromosomes.

A 371-kb region in *B. sylvaticum* has already been sequenced was also compared with orthologous regions from rice and wheat genomes [36]. In this region, *Brachypodium* and wheat showed perfect macrocolinearity, whereas rice contains an approximately 220-kb inversion. Using conserved genomic and EST sequences, divergence between *Brachypodium* and wheat was estimated to be 35–40 million years, which is significantly more recent than the divergence of rice and wheat, which is estimated to have occurred approximately 50 million years [37].

Chosen target loci from *Brachypodium* genome are also being sequenced and compared with genomic sequences from a variety of plant species including the following: (i) wheat species (*Triticum* and *Aegilops*) with different ploidy levels, (ii) rice, and (iii) *B. sylvaticum*, for which a BAC library is available [38, 39]. This comparison revealed that there is a better conservation of microcolinearity between wheat and *Brachypodium* orthologous regions than between wheat and rice, as was also shown in an earlier study [36]. For instance, sequence comparison at the grain hardness locus shows that genes responsible for grain hardness/softness, which is seed quality trait in wheat, are absent from the rice orthologous region, but present in the *B. sylvaticum* orthologous region. The gene density found in *B. sylvaticum* genome is comparable to that of rice (one gene per 8 kb). These results illustrate that *Brachypodium* species may represent an intermediate model for wheat genome analysis.

To test the potential of *Brachypodium* as a model for the functional analysis of ryegrass (*Lolium perenne*) flowering genes, expression of two *Terminal Flower* 1 orthologs, namely, *LpTFL* 1 (from *L. perenne*) and *TFL* 1 (from *Arabidopsis*), was examined in two different *B. distachyon* accessions [40]. Both these repressors significantly delayed heading date. The short life cycle of *Brachypodium* and the rapid transformation system allowed heading date scoring of T1s within the first year after transformation, thus demonstrating the potential of *Brachypodium* as a model for ryegrass (*L. perenne*) also.

Brachypodium is also being explored as a model for the genomics research involving study of cereals-pathogen interactions. For instance, varying degrees of susceptibility and resistance to *Magnaporthe grisea* (economically destructive pathogen and casual agent of Rice Blast disease that can also infect temperate cereals and forage grasses) have been found in several *Brachypodium* accessions. Aetiology of fungal development and disease progression in *Brachypodium* closely resembled those of rice infections; an overexpression of genes that were homologous with barley genomic probes was also observed [41]. Recent advances in *Brachypodium* genomics also involved use of metabolic profiling using Fourier-transform infrared spectroscopy (FT-IR) for high-throughput metabolic fingerprinting and electrospray ionization mass spectrometry (ESI-MS). These metabolomic approaches have shown considerable differential phospholipids processing of membrane lipids during *M. grisea-B. distachyon* accessions ABR1 (susceptible) and ABR5 (resistant) interactions [42]. *Brachypodium distachyon,* being a host for

M. grisea and other disease-causing pathogens of Pooid cereals [42], is a suitable model for conducting functional genomics research involving study of *M. grisea* pathology and plant responses [41].

5. CONCLUSIONS

With the small genome size and simple growth requirements, *Brachypodium* provides us with a genome, which is a model for in-depth understanding of functional genomics of temperate grass genome. As a model, it overcomes some of the drawbacks that are inherent in the genomes of *Arabidopsis* and rice that have already been sequenced and have been hitherto considered models for the improvement of crop species like wheat and barley. Therefore, elucidation and an improved understanding of *Brachypodium* genomics has enormous potential to benefit all phases of society. It provides improved, efficient, and effective genetics and genomics program. The knowledge on *Brachypodium* genome is also useful for an in-depth understanding of evolutionary relationships among different plant genomes. This will play a pivotal role in comparative studies in diverse fields such as ecology, molecular evolution, and comparative genetics.

ACKNOWLEDGMENT

We greatly acknowledge Professor Z. Sayers for critical review of the manuscript.

REFERENCES

[1] L. Watson and M. Dallwitz, "The grass genera of the world: descriptions, illustrations, identification, and information retrieval; including synonyms, morphology, anatomy, physiology, phytochemistry, cytology, classification, pathogens, world and local distribution, and references," 1992, http://delta-intkey.com.

[2] J. Draper, L. A. J. Mur, G. Jenkins, et al., "*Brachypodium distachyon*. A new model system for functional genomics in grasses," *Plant Physiology*, vol. 127, no. 4, pp. 1539–1555, 2001.

[3] M. D. Bennett and I. J. Leitch, "Plant DNA C-values database (release 3.0, Dec. 2004)," 2004, http://www.rbgkew.org.uk/cval/homepage.html.

[4] J. P. Vogel, D. F. Garvin, O. M. Leong, and D. M. Hayden, "*Agrobacterium*-mediated transformation and inbred line development in the model grass *Brachypodium distachyon*," *Plant Cell, Tissue and Organ Culture*, vol. 84, no. 2, pp. 199–211, 2006.

[5] P. Christiansen, C. H. Andersen, T. Didion, M. Folling, and K. K. Nielsen, "A rapid and efficient transformation protocol for the grass *Brachypodium distachyon*," *Plant Cell Reports*, vol. 23, no. 10-11, pp. 751–758, 2005.

[6] M. Khan and C. Stace, "Breeding relationships in the genus *Brachypodium* (Poaceae: Pooideae)," *Nordic Journal of Botany*, vol. 19, no. 3, pp. 257–269, 1999.

[7] P. Catalan, Y. Shi, L. Armstrong, J. Draper, and C. A. Stace, "Molecular phylogeny of the grass genus *Brachypodium* P-Beauv based on RFLP and RAPD analysis," *Botanical Journal of the Linnean Society*, vol. 117, no. 4, pp. 263–280, 1995.

[8] P. Catalán and R. G. Olmstead, "Phylogenetic reconstruction of the genus *Brachypodium* P. Beauv. (Poaceae) from combined sequences of chloroplast*ndh*F gene and nuclear ITS," *Plant Systematics and Evolution*, vol. 220, no. 1-2, pp. 1–19, 2000.

[9] The Arabidopsis Genome Initiative, "Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*," *Nature*, vol. 408, no. 6814, pp. 796–815, 2000.

[10] International Rice Genome Sequencing Project, "The map-based sequence of the rice genome," *Nature*, vol. 436, no. 7052, pp. 793–800, 2005.

[11] E. A. Kellogg, "Evolutionary history of the grasses," *Plant Physiology*, vol. 125, no. 3, pp. 1198–1205, 2001.

[12] R. Hasterok, J. Draper, and G. Jenkins, "Laying the cytotaxonomic foundations of a new model grass, *Brachypodium distachyon* (L.) Beauv," *Chromosome Research*, vol. 12, no. 4, pp. 397–403, 2004.

[13] L. Aragón-Alcaide, T. Miller, T. Schwarzacher, S. Reader, and G. Moore, "A cereal centromeric sequence," *Chromosoma*, vol. 105, no. 5, pp. 261–268, 1996.

[14] D. R. Dewey, "The genomic system of classification as a guide to intergeneric hybridization with the perennial *Triticeae*," in *Gene Manipulation in Plant Improvement*, J. P. Gustafson, Ed., pp. 209–279, Plenum, New York, NY, USA, 1984.

[15] G. L. Stebbins, "The morphological, physiological, and cytogenetic significance of polyploidy," in *Chromosomal Evolution in Higher Plants*, F. R. S. Barrington and A. J. Willis, Eds., Addison-Wesley, Reading, Mass, USA, 1971.

[16] A. Martín and V. Chapman, "A hybrid between *Hordeum chilense* and *Triticum aestivum*," *Cereal Research Communication*, vol. 4, pp. 365–368, 1977.

[17] A. Martín and E. Sánchez-Monge-Laguna, "A hybrid between *Hordeum chilense* and *Triticum turgidum*," *Cereal Research Communication*, vol. 8, pp. 349–353, 1980.

[18] K. Anamthawat-Jónsson, S. K. Bödvarsdóttir, B. Th. Bragason, J. Gudmundsson, P. K. Martin, and R. M. D. Koebner, "Wide hybridization between wheat (*Triticum* L.) and lymegrass (*Leymus* Hochst.)," *Euphytica*, vol. 93, no. 3, pp. 293–300, 1997.

[19] Y. Cauderon, "Use of agropyron species for wheat improvement," in *Broadening the Genetic Base of Crops*, A. C. Zeven and A. M. van Harten, Eds., pp. 175–186, Pudoc, Wageningen, The Netherlands, 1978.

[20] H. C. Sharma and B. S. Gill, "Current status of wide hybridization in wheat," *Euphytica*, vol. 32, no. 1, pp. 17–31, 1983.

[21] I. H. Robertson, "Chromosome numbers in *Brachypodium* Beauv. (Gramineae)," *Genetica*, vol. 56, no. 1, pp. 55–60, 1981.

[22] M. D. Bennett and I. J. Leitch, "Nuclear DNA amounts in angiosperms: progress, problems and prospects," *Annals of Botany*, vol. 95, no. 1, pp. 45–90, 2005.

[23] R. Hasterok, A. Marasek, I. S. Donnison, et al., "Alignment of the genomes of *Brachypodium distachyon* and temperate cereals and grasses using bacterial artificial chromosome landing with fluorescence *in situ* hybridization," *Genetics*, vol. 173, no. 1, pp. 349–362, 2006.

[24] W. Gu, C. M. Post, G. D. Aguirre, and K. Ray, "Individual DNA bands obtained by RAPD analysis of canine genomic DNA often contain multiple DNA sequences," *Journal of Heredity*, vol. 90, no. 1, pp. 96–98, 1999.

[25] M. Luo, Y. Ma, and N. Huo, "Construction Of physical map for *Brachypodium distachyon*," in *Plant & Animal Genomes XV Conference*, San Diego, Calif, USA, January 2007.

[26] D. Garvin, "*Brachypodium distachyon*. A new model plant for structural and functional analysis of grass genomes," in *Model Plants and Crop Improvement*, R. Koebner and R. Varshney, Eds., pp. 109–123, CRC Press, Boca Raton, Fla, USA, 2006.

[27] P. Bablak, J. Draper, M. R. Davey, and P. T. Lynch, "Plant regeneration and micropropagation of *Brachypodium distachyon*," *Plant Cell, Tissue and Organ Culture*, vol. 42, no. 1, pp. 97–107, 1995.

[28] N. Huo, Y. Q. Gu, G. R. Lazo, et al., "Construction and characterization of two BAC libraries from *Brachypodium distachyon*, a new model for grass genomics," *Genome*, vol. 49, no. 9, pp. 1099–1108, 2006.

[29] J. P. Vogel, Y. Q. Gu, P. Twigg, et al., "EST sequencing and phylogenetic analysis of the model grass *Brachypodium distachyon*," *Theoretical and Applied Genetics*, vol. 113, no. 2, pp. 186–195, 2006.

[30] K. C. Engvild, "Mutagenesis of the model grass *Brachypodium distachyon* with sodium azide," Risoe-R-1510 (EN) Report, Risoe National Laboratory, 2005.

[31] M. W. Bevan, "Establishing a BAC-based physical map of *Brachypodium distachyon* as an aid to physical mapping in bread wheat," in *Plant & Animal Genomes XIV Conference*, San Diego, Calif, USA, January 2006.

[32] R. Kalendar and A. H. Schulman, "Retrotransposons and their use as molecular markers in *Brachypodium*," in *Plant & Animal Genomes XIV Conference*, San Diego, Calif, USA, January 2006.

[33] M. Bevan, N. McKenzie, M. Trick, et al., "Developing a genetic map of *Brachypodium distachyon* Bd21," in *Plant & Animal Genomes XV Conference*, San Diego, Calif, USA, January 2007.

[34] T. C. Mockler, S. Givan, C. Sullivan, and R. Shen, "Bioinformatics and genomics resources for *Brachypodium distachyon*," in *Plant & Animal Genomes XV Conference*, San Diego, Calif, USA, January 2007.

[35] Y. Q. Gu, N. Huo, G. R. Lazo, et al., "Towards *Brachypodium* genomics: analysis of 60,000 BAC end sequences and sequence comparison with cereal crops," in *Plant & Animal Genomes XV Conference*, San Diego, Calif, USA, January 2007.

[36] E. Bossolini, T. Wicker, P. A. Knobel, and B. Keller, "Comparison of orthologous loci from small grass genomes *Brachypodium* and rice: implications for wheat genomics and grass genome annotation," *The Plant Journal*, vol. 49, no. 4, pp. 704–717, 2007.

[37] A. H. Paterson, J. E. Bowers, and B. A. Chapman, "Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 26, pp. 9903–9908, 2004.

[38] M. Charles, N. Choisne, S. Samain, N. Boudet, and B. Chalhoub, "*Brachypodium* species as intermediate models for wheat genomics," in *Plant & Animal Genomes XIV Conference*, San Diego, Calif, USA, January 2006.

[39] T. N. Foote, S. Griffiths, S. Allouis, and G. Moore, "Construction and analysis of a BAC library in the grass *Brachypodium sylvaticum*: its use as a tool to bridge the gap between rice and wheat in elucidating gene content," *Functional & Integrative Genomics*, vol. 4, no. 1, pp. 26–33, 2004.

[40] P. Olsen, I. Lenk, C. S. Jensen, et al., "Analysis of two heterologous flowering genes in *Brachypodium distachyon* demonstrates its potential as a grass model plant," *Plant Science*, vol. 170, no. 5, pp. 1020–1025, 2006.

[41] A. P. M. Routledge, G. Shelley, J. V. Smith, N. J. Talbot, J. Draper, and L. A. J. Mur, "*Magnaporthe grisea* interactions with the model grass *Brachypodium distachyon* closely resemble those with rice (*Oryza sativa*)," *Molecular Plant Pathology*, vol. 5, no. 4, pp. 253–265, 2004.

[42] J. W. Allwood, D. I. Ellis, J. K. Heald, R. Goodacre, and L. A. J. Mur, "Metabolomic approaches reveal that phosphatidic and phosphatidyl glycerol phospholipids are major discriminatory non-polar metabolites in responses by *Brachypodium distachyon* to challenge by *Magnaporthe grisea*," *The Plant Journal*, vol. 46, no. 3, pp. 351–368, 2006.

[43] T. Murashige and F. Skoog, "A revised medium for rapid growth bioassays with tobacco tissue cultures," *Physiologia Plantarum*, vol. 15, no. 3, pp. 473–497, 1962.

[44] E. M. Linsmaier and F. Skoog, "Organic growth factor requirements of tobacco tissue cultures," *Physiologia Plantarum*, vol. 18, no. 1, pp. 100–127, 1965.

Hindawi Publishing Corporation
International Journal of Plant Genomics
Volume 2008, Article ID 458732, 11 pages
doi:10.1155/2008/458732

Review Article

Sugarcane Functional Genomics: Gene Discovery for Agronomic Trait Development

M. Menossi,[1] M. C. Silva-Filho,[2] M. Vincentz,[1] M.-A. Van-Sluys,[3] and G. M. Souza[4]

[1] *Departmento de Genetica e Evolução IB-Unicamp, Centro de Biologia Molecular e Engenharia Genética, Universidade Estadual de Campinas, C.P. 6010, CEP 13083-970 Campinas, SP, Brazil*
[2] *Departamento de Genética, Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Av. Pádua Dias, 11, C.P. 83, 13400-970 Piracicaba, SP, Brazil*
[3] *Departamento de Botânica, Instituto de Biociências, Universidade de São Paulo, Rua do Matão 277, 05508-090 São Paulo, SP, Brazil*
[4] *Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, Av. Prof. Lineu Prestes 748, 05508-900 São Paulo, SP, Brazil*

Correspondence should be addressed to G. M. Souza, glmsouza@iq.usp.br

Received 29 March 2007; Accepted 21 November 2007

Recommended by Yunbi Xu

Sugarcane is a highly productive crop used for centuries as the main source of sugar and recently to produce ethanol, a renewable bio-fuel energy source. There is increased interest in this crop due to the impending need to decrease fossil fuel usage. Sugarcane has a highly polyploid genome. Expressed sequence tag (EST) sequencing has significantly contributed to gene discovery and expression studies used to associate function with sugarcane genes. A significant amount of data exists on regulatory events controlling responses to herbivory, drought, and phosphate deficiency, which cause important constraints on yield and on endophytic bacteria, which are highly beneficial. The means to reduce drought, phosphate deficiency, and herbivory by the sugarcane borer have a negative impact on the environment. Improved tolerance for these constraints is being sought. Sugarcane's ability to accumulate sucrose up to 16% of its culm dry weight is a challenge for genetic manipulation. Genome-based technology such as cDNA microarray data indicates genes associated with sugar content that may be used to develop new varieties improved for sucrose content or for traits that restrict the expansion of the cultivated land. The genes can also be used as molecular markers of agronomic traits in traditional breeding programs.

1. SUGARCANE: A HIGHLY SUCCESSFUL CROP WITH A CHALLENGING GENOME

Sugarcane is an important tropical crop and has served as a source of sugar for hundreds of years. With an originally soft, watery culm sugarcane acquired through human selection a distinctive feature of partitioning carbon into sucrose in the stem. The striking ability of accumulating levels of sucrose that can reach around 0.7 M in mature internodes [1] is an almost unique feature in cultivated plants.

Sugarcane is cultivated in more than 20 million hectares in tropical and subtropical regions of the world, producing up to 1.3 billion metric tons of crushable stems. It is generally used to produce sugar, accounting for almost two thirds of the world's production and has recently gained increased attention because ethanol derived from cane sugar represents an important renewable biofuel source, which could turn it into a global commodity and important energy source. Sugarcane bagasse (the major waste product generated by sugar mills after extraction of the sucrose from cane juice) is largely used for energy cogeneration at the mill or for the production of animal feed increasing the overall efficiency of the crop system. Recently, there has been increased interest in using bagasse for processes such as paper production, as a dietary fiber in bread, as a wood substitute in the production of wood composite, and in the synthesis of carbon fibres [2–6]. It is expected that enzymatic and hydrolytic processes that allow the bagasse carbon units from cellulose and hemicellulose to be fermented, will soon be scaled up for ethanol production, turning sugarcane into an efficient crop for energy production.

Commercial sugarcane relies on vegetative propagation through stem cuttings to generate a new clonal plant, resulting from lateral bud growth, and subsequently stools, with a large number of tillers. In 12 months the plant will reach 4-5 meters, with extractable culms measuring 2-3 meters and a sugar content of 13–16%. After harvest, underground buds will sprout starting a new crop season. In most situations 4–6 harvests are possible before the field is renewed. After each harvest, leaves and plant toppings removed from the stems are left in the fields allowing for nutrient recycling, soil protection and growth without crop rotation.

Sugarcane belongs to the genus *Saccharum* L composed of hybrids [7, 8] derived from *Saccharum officinarum* (Noble clones), *S. sinense* (Chinese clones), *S. barberi* (North Indian clones), and *S. spontaneum* [9]. The hybrids are highly polyploid and aneuploid and on average contain 100–120 chromosomes with an estimated somatic cell size of 10,000 Mbp [10]. The number of chromosomes can vary in commercial cultivars. The basic genome size ranges from 760 to 926 Mbp, which is twice the size of the rice genome (389 Mbp) and similar to sorghum's (760 Mbp) [11]. Even in the face of the economic importance, it represents to many countries, the complexity of the sugarcane genome inhibited large efforts and investments in the development of biotechnology and genetic tools for this crop. Cultivar improvement has been achieved over the years using traditional breeding, which can take up to 15 years of selections. Nevertheless sugarcane transgenics are still lagging behind. Herbicide-, herbivory-, and viral-resistant transgenic plants have been reported but so far there has been no commercial release. This is probably due to intellectual property and regulatory issues, but may also be related to the fact that for complex traits, such as sucrose content, the genes to be used have not yet been proved ideal for improving agronomic performance. Gene discovery and identification is essential for breeding programs, either for transgenic plant development or for marker-assisted breeding.

The complete genome sequence of a sugarcane cultivar is not yet available. Significant progress has been noted recently with the development of tools such as expressed sequence tags (ESTs). Large collections have become available to explore the large polyploid sugarcane genome and consequently renewed the interest in sugarcane genetics [12–14]. This review will focus on describing EST development and subsequent progress that led to the identification of genes associated with agronomic traits of interest in sugarcane. It will also highlight some of the possible functions of genes associated with sucrose content, including biotic and abiotic stress and the role that phytohormones may play in the adaptive responses of this plant.

2. EST PROFILING FOR GENE DISCOVERY

ESTs represent tags of the expressed portion of a genome and therefore potentially identify genes encoding proteins, natural antisense transcripts [15–18], miRNA, transacting siRNA precursors [19, 20], and more generally noncoding RNA [21]. The information carried by an EST collection is a significant starting point to determine an organism's genome content but more pragmatically, when considering important crops, it can directly point to genes which may contribute to agronomical trait development (e.g., tolerance to abiotic and biotic stresses, mineral nutrition, and sugar content amongst others).

Several sugarcane ESTs collections have been developed [22–28]. The publicly available sugarcane ESTs were assembled into tentative consensus sequences (virtual transcripts), singletons, and mature transcripts, referred to as the Sugarcane Gene Index (SGI; http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?gudb=s_officinarum). The Brazilian sugarcane EST project collection (SUCEST, http://sucest-fun.org, [26]) generated 237,954 ESTs, which were organized into 43,141 putative unique sugarcane transcripts (26,803 contigs and 16,338 singletons) referred to as sugarcane assembled sequences (SASs). An internal redundancy analysis suggested that this collection of SASs represented 33,000 sugarcane genes [13, 26, 29] but this estimation was likely to have been an overestimation, since a two-fold redundancy among SASs that presented significant similarity with *Arabidopsis* proteins (60% of the SASs) was detected (M. Vincentz, unpublished data). A detailed organization of sugarcane genes into functional categories (i.e., signal transduction components, regulation of gene expression, development, biotic and abiotic stresses, transposable elements, metabolism, etc., [26]) was completed and represents the basis to develop functional genomic approaches.

The contribution of this large set of SASs to our understanding of the processes underlying angiosperm evolution was also of significance. A comparison of the SASs with the DNA and protein sequences from other angiosperms confirmed that lineage-specific gene loss, high evolutionary rate of specific sequences, and exon shuffling were important processes involved in the divergence among angiosperms [30]. Of particular interest are the monocot-specific sequences that evolve at high rates and are found in members of conserved angiosperm gene families, because they may lead to functional diversification and may therefore be related to the differentiation of specific lineages. Interestingly, two SASs (SCEZSD2038A10.g and SCSFRT2070F09.g), only detected in sugarcane, sorghum and maize, point to the existence of recent innovations in the Andropogoneae tribe (M. Vincentz, unpublished) and raise the question of what kind of adaptive traits are associated with these sequences.

Finally, it is important to note the contribution that the EST collections have made to our understanding of the sugarcane genome structure, number of alleles and the complex relationship of specific alleles and allele dosage to phenotypes. Single sequence repeats (SSRs) and single nucleotide polymorphisms (SNPs) have been annotated in a number of genotypes [31, 32]; and with the advent of pyrosequencing, their identification has been increasingly adding to our knowledge of large genomes [33]. A comprehensive functional map of the sugarcane genome has recently been described with an enhanced resolution, creating the means for developing "perfect markers" associated with key QTL [34].

3. GENE EXPRESSION BLUEPRINT OF SUGARCANE TISSUES

The availability of ESTs allows for large-scale gene expression analysis using a variety of tools. Several studies have reported an in silicoanalysis of transcript enrichment when different cDNA libraries were compared [26, 35]. Of the 43,141 SUCEST SASs, 1234 were considered to be tissue-enriched. The maximum number of ESTs in a tissue-enriched SAS (i.e., with higher transcript amounts in one or more tissues) was found to be in prolamin, which contained 360 ESTs. Developing seeds contained 1902 specific ESTs (33% of the total), with almost half of these (919) encoding prolamins, the major seed storage protein found in cereals. These ESTs included six putative new genes with a high level of expression in seeds (up to 32 ESTs/SAS). The most frequent protein domains found in tissue-enriched SASs were the protein kinase domain, followed by the trypsin-amylase inhibitor, seed storage protein, and lipid transfer protein domains. Overall, 13 transcription factor families were found to be specific for flowers, five for roots, three for *Herbaspirillum*-inoculated plantlets, and two for developing seeds and other tissues.

Following sequence identification, a functional genomics project, the SUCEST-FUN Project (http:sucest-fun.org), was implemented to associate putative roles with the sugarcane genes. cDNA microarrays containing sugarcane ESTs were used to determine temporal and spatial gene expression. Determining the distribution of gene transcripts in sugarcane tissues has helped define tissue-specific activities and ubiquitous genes, and point out genes in which promoter sequences could be searched for. This is of particular interest if one is interested in directing the expression of transgenes to particular plant tissues to avoid pleiotropic effects. Individual gene expression variation was investigated using cDNA microarrays containing 1280 distinct elements, on plants grown in the field. Transcript abundance in six plant organs (flowers, roots, leaves, lateral buds, 1st (immature), and 4th (mature) internodes) was analyzed [36], resulting in the identification of 217 genes with a tissue-enriched expression patterns, while 153 genes showed highly similar expression levels in all the tissues analyzed. A virtual profile matrix was constructed where tissue expression was compared amongst 24 tissue samples. Amongst the tissue-enriched genes,a caffeic acid 3-*O*-methyltransferase (COMT) gene expressed primarily in the mature internode was identified. This enzyme is involved in lignin biosynthesis and, in association with other enzymes like the CCOMT (caffeoyl CoA 3-*O*-methyltransferase), keeps the cell lignin content and composition in check. The identification of this culm-enriched enzyme may lead to improved sugarcane varieties with an altered lignin content: a trait highly valuable for the paper industry and for those interested in increasing hydrolysis of the sugarcane bagasse for fermentation purposes. The tissue specificity data was also evaluated against data from plants submitted to biotic and abiotic stresses, which can shed light on the putative roles of newly identified genes, such as genes for which no similarity has been found with genes in the public databases [37].

Active transcription of transposable elements (TEs) was also detected in the SUCEST database [38] and enabled the identification of a previously unknown set of genetic mobile elements in sugarcane. Further studies confirmed the expression profile of 68 individual TE clones [39]. Four actively dividing tissues were examined (callus, apical meristem, leaf roll, and flower), and callus was determined to be the tissue expressing the most diverse group of TEs. Both transposons and retrotransposons are expressed, which suggest that some of these mobile elements may have an important role in genome metabolism, as previously described for other elements in several biological systems [40, 41]. Further analysis of these transcribed TEs revealed that some of the families were constituted by both bona fide transposable elements and domesticated variants that had been captured by the plant genome to perform a yet unknown function [42]. Mutator-like elements were the most expressed transposons in sugarcane, and four groups were identified that showed similarity with the MURA transposase protein [43], of which two represented domesticated elements related to the Mustang-like genes described in rice [42, 44]. Amongst the retrotransposons, Hopscotch-like sequences, the most prevalent in the SUCEST database, showed a highly diverse expression profile. Retrotransposons carry their promoter region along the length of their transcribed mRNA and GUS-fusion expression analyses for three out of four TEs, this is being confirmed in transient assays [44], leading to the possibility of using these sequences as promoters for the expression of genes of interest in sugarcane.

4. INSIGHTS IN THE SUGARCANE RESPONSES TO BIOTIC STRESS

Plants are constantly challenged by a wide array of biotic stresses, such as herbivorous insects, nematodes, and by fungal, bacterial, and viral infestations. Phytohormones largely mediate plant responses to attacks by triggering conserved defence mechanisms, each with an intricate signalling pathway leading to plant protection. Cross-talk signalling pathways leading to plant defence have been reported, with synergistic and antagonistic outcomes [45]. Specific and general responses are mediated by distinct signals, mainly jasmonic acid, ethylene, and salicylic acid. It has been shown that both the ethylene and jasmonic acid signalling pathways act synergistically in plant defence. For example, ethylene synthesis increases the response to severaltypes of biotic challenges (e.g., bacteria, fungi [46], and insects [47]). In sugarcane, a putative ethylene receptor and two putative transcription factors, which are members of the ethylene signalling pathway, have been shown to be regulated during the association with nitrogen-fixing endophytic bacteria [48]. In addition, other signals such as green leaf volatiles (GLVs) may be involved in the orchestration of plant defences since their production is drastically enhanced when they are under biotic stress [49].

Biotic stress is responsible for significant sugarcane losses, posing a demand for the development of new stress-tolerant cultivars. In order to reduce insect and pathogen damage, plants have developed complex and varied defence mechanisms, including chemical and physical barriers.

In the last few years, an extensive amount of work has been undertaken in order to decipher the sugarcane response to biotic stress, mainly related to some insect herbivores and pathogens. Amongst the SUCEST sequences, dozens of orthologous genes involved in the sugarcane response to insect herbivores [50] and *Diazothophic* endophytes [51, 52] were identified. Although the sugarcane-endophytic bacteria interaction is an advantageous association for both organisms, it is thought that sugarcane plants activate defence responses before the establishment of such symbiosis [53]. A study based on a wide gene expression analysis of 1,545 genes in sugarcane revealed that *Gluconacetobacter diazotrophicus* and *Herbaspirillum seropedicae* endophytic bacteria activated distinct classes of defence proteins, including four plant disease-resistant genes (R-genes), salicylic acid biosynthesis genes, five transcription factors, and so on. On the other hand, *Diatraea saccharalis* herbivory specifically upregulated the expression of a pathogenesis-related protein similar to thaumatin [37]. Transcript profiling of sugarcane-resistant plants to either *Ustilago scitaminea* or *Bipolaris sacchari* (also known as *Helminthosporium sacchari* or *Drechslera sacchari*), causal agents of smut and eyespot, respectively, identified 62 differentially regulated genes, of which 10 were downregulated and 52 were induced. Nineteen out of 52 transcript-derived fragments showed homology to known plant gene sequences, most being related to defense or signaling [54].

A considerable amount of data was obtained on how the plant hormone methyl jasmonate (MeJa) could be regulating plant defence reactions [28, 37, 55]. cDNA microarrays containing 829 ESTs from roots treated with MeJA, and 4793 ESTs from immature and mature stem tissues were used to evaluate gene expression changes produced by MeJA [28]. An MeJA solution was applied to the soil containing the plants, and the roots were harvested after 1, 3, and 10 days. The highest induction was observed for genes encoding the dirigentprotein, which is involved in lignin assembly and can protect plants against fungal attack [56]. Gene categories with increased transcript levels included signal transduction, the phenylpropanoid pathway, oxidative stress, and MeJA synthesis, indicating that several processes were altered by MeJA. In agreement with several studies involving transcription profiling, most of the up- or downregulated genes had unknown functions, reinforcing the great challenge of understanding plant gene function. Responses of sugarcane leaves sprayed with MeJA for 0.5, 1, 3, 6, and 12 hours, were investigated using nylon cDNA arrays containing 1536 ESTs from several cDNA libraries [55]. A total of 15 genes were upregulated, while 11 were downregulated. As observed in sugarcane roots [28], MeJA changed the expression of genes involved in several biological processes including transcription (a zinc finger protein), signalling (a protein kinase), and abiotic stress responses (a carboxy-peptidase, a peroxidase, and a heat shock factor). The authors complemented their analysis using a digital mRNA expression profiling of the differentially expressed genes, providing an overview of their expression patterns in different sugarcane tissues. These results support the idea that different in silicostrategies can be used to enrich functional genomics analyses.

Changes in gene expression in leaves exposed to MeJA were also evaluated using cDNA microarrays containing 1545 genes [37], mostly corresponding to signal transduction components [57]. The upregulation of transcription factors (MYB, NAC, and Aux/IAA) and histone homologues (H4 and H2B) strongly suggested chromatin remodelling followed by the activation of a cascade of signalling genes. Several protein kinases were up- and downregulated, indicating a complex network of sugarcane responses to MeJA.

Several strategies have been used to improve plant defence against insects and pathogens. The activation of stress-response transcription factors was found to enhance plant tolerance to fungal and bacterial pathogens in transgenic plants [58]. However, little is known about the function of other components of the plant transcription machinery during stress. The identification and characterization of agronomically-interesting genes related to herbivores and pathogens is a major challenge for sugarcane functional genomics. Several candidates have been tested in the last few years and incorporated into elite genotypes [59–66]. The heterologous expression of defence-related proteins in sugarcane, such as the soybean proteinase inhibitors encoding genes [67] or cry proteins from *Bacillus thuringiensis* [68], has led to increased resistance against the sugarcane borer *D. saccharalis*, the major sugarcane pest in Brazil. In addition, the molecular and functional characterization of cysteine proteinase inhibitors opened up new perspectives on pathogen control, since sugarcane cystatins inhibited the growth of the filamentous fungus *Trichoderma reesei*, suggesting that it can also be employed to inhibit the growth of pathogenic sugarcane fungi [69]. The use of inducible promoters will have a significant impact on the effectiveness and management of transgenic plants. One such promoter has been cloned in sugarcane that responds to the sugarcane borer (Silva-Filho, unpublished results). Taken together, the combination of new genes with appropriate regulatory sequences will be a major outcome of the sugar cane OMICS in breeding programs.

5. ASSESSING SUGARCANE GENES RELATED TO ABIOTIC STRESS

Plants face several restrictions in their environment and have developed a wide array of strategies to either avoid or cope with the stress condition. Most of the studies using high throughput assays, such as cDNA microarrays have been conducted with model plants, such as *Arabidopsis* or species not considered as tropical crops. Recently, the first insights into the responses of sugarcane to environmental stress have been provided.

Amongst abiotic stress, water deficit plays a major role, and increasing water scarcity has been observed throughout the world. Plant irrigation currently accounts for approximately 65% of global freshwater use, indicating that the development of drought-resistant plant varieties will be a necessity in the near future [70, 71]. Agricultural irrigation is one of the most water demanding human activities. In the case of sugarcane, agricultural frontiers are expanding, in part, in areas where irrigation is needed [72, 73].

To increase the knowledge on the sugarcane responses to drought, cDNA microarrays were used to evaluate gene expression in plants submitted to 24, 72, and 120 hours of water deprivation [37]. Drought stress caused dramatic changes in the gene expression profile of sugarcane plants, with 93 genes being up- or downregulated. Among the genes differentially expressed, transcription factor orthologs of the Myb, WRKY, NAC, and DREB proteins, which are known as role players in the drought responses of other systems [74–77], were upregulated. Sugarcane plants also selectively activated proteases in response to hydric stress, since a homologue to the cysteine proteinase RD19A precursor was induced. This gene is also induced by water stress in *Arabidopsis* [78].

Although it may sound surprising, another important stress in the case of sugarcane is cold stress, caused by temperatures below 0°C (freezing) or by low temperatures above 0°C (chilling). Cold stress is unusual in tropical areas, where most of the world's sugarcane is grown, but occasionally cold can severely affect crops in these regions. This is because most plants in the tropics have not developed strategies to avoid the devastating consequences of cold to the cells [79]. There is evidence that sugarcane varieties differ in their sensitivity to cold [80], suggesting the presence of alleles that might help this tropical crop to cope with this stress. These genes would have a great potential in breeding programs and also in the engineering of sugarcane plants with higher cold tolerance, a highly valuable trait that would allow the cultivation of this plant in temperate climates.

The first report of the use of cDNA arrays to discover sugarcane genes modulated by cold stress was conducted by Nogueira et al. [81]. The exposure of sugarcane plantlets to 4°C repressed the expression of 25 genes, while a further 34 genes were upregulated. Sugarcane homologues to several genes known to be induced by cold stress were found together with genes induced by drought in other species. This is probably because the cold induces the formation of ice, dehydrating the cell. Interestingly, 20 genes that had not previously been associated with cold or drought stress were identified, suggesting that sugarcane might activate novel cold response pathways. One example is the gene encoding a putative NAD-dependent dehydrogenase that might be involved in the protection against oxidative stress due to cold exposure. One of the genes, *SsNAC23*, is a member of the NAC family of transcriptional factors that are involved in biotic and abiotic stress and development [82]. In a further characterization of SsNAC23, Nogueira et al. [81] showed that the protein is targeted to the nucleus. In addition, *SsNAC23* transcripts also increased in response to herbivory and water stress. This data further reinforces the view that different kinds of stress may have common signalling pathways. Based on this expression profiling experiment, the authors proposed a hypothetical model integrating the several components activated by sugarcane in response to low temperature. The same data analyzed using PmmA [83] revealed a new set of 30 genes as differentially expressed. Among the genes upregulated was a putative endonuclease involved in nucleic acid repair, indicating that low temperature stress might cause DNA damage. Most genes in this new set were repressed by cold stress, such as those encoding a myo-inositol 1-phosphate synthase and an MAP Kinase.

Several plant responses to environmental stress are mediated by phytohormones, with a well-known cross-talk between them [84, 85]. To assess the role of ABA in sugarcane, Rocha et al. [37] sprayed ABA on sugarcane leaves and evaluated the gene expression profile using the cDNA arrays described above. Two genes encoding orthologs to receptor Ser/Thr kinases were upregulated. A phosphatase and a small GTPase were also induced, while a protein kinase was repressed. These findings help to depict an overview of the network of ABA signal transduction in sugarcane. The cDNA array data pinpointed several aspects of the sugarcane metabolism that seem to have been changed in response to ABA. For example, changes in the fatty acid composition probably take place, since a fatty acid desaturase was induced, while transpiration would be decreased due to the action of a PP2C-like protein homologous to ABI1 and ABI2. Moreover, the work of Rocha et al. [37] also showed drought responses similar to those elicited by ABA. For example, two delta-12 oleate desaturases, an S-adenosylmethionine decarboxylase, and a PP2C-like protein phosphatase were induced by both ABA and drought.

The cross-talk between ABA and MeJA also become evident from the activation of two genes involved in salicylic acid and MeJA biosynthesis in the ABA-treated plants [37]. ABA treatment elicited an antagonistic response between the ABA and auxin pathways. A gene coding for a protein similar to the auxin responsive protein GH3 [86] was found to be repressed by ABA. Furthermore, a gene coding for a protein with a predicted auxin-repressed domain found in dormancy-associated and auxin-repressed proteins [87] was upregulated by this hormone. The cross-talk of other hormone signalling pathways during water stress was further highlighted by the differential expression of several genes encoding proteins involved in ethylene, gibberellin, salicylic acid biosynthesis, as well as other proteins involved in hormone perception and action. In the same line, several genes induced by drought stress were also observed in sugarcane plants exposed to MeJA, suggesting that this hormone might play a role in gene expression changes during water deficit in this crop. These genes are interesting tools in the engineering of plants aimed at increasing drought tolerance. In fact, transgenic tobacco and rice plants over expressing a DREB protein and an NAC protein, respectively, had improved performance in response to water scarcity [88, 89].

Last, but not least, a study on the evaluation of sugarcane responses to low P availability was also reported. Most of the world's agriculture takes place in soils with low availability of P and other nutrients [90]. Phosphorus, a key nutrient for plant growth and development, is taken up as inorganic phosphate (Pi), and most soils have very low Pi concentrations (around 2 mM) compared to the range 5–20 mM found inside the plant cells [91]. Soil supplementation with rock phosphate is widely used to increase P availability, increasing the productivity of several crops, including sugarcane [92]. However, since the P fertilizers may be exhausted within the next 60–90 years, and the P released into watercourses increases eutrophication of the water sources, there is a need to minimize P fertilization.

Phosphorus starvation experiments were used to access the changes in the gene expression profiles and gain information on the strategies used by sugarcane to overcome this nutrient deficiency stress [37]. The effect of P deficiency on the gene expression was evaluated in the roots of sugarcane plantlets. Fourteen genes were found to be repressed after 6 hours and 48 hours due to the absence of P in the nutrient solution. Surprisingly, no upregulated genes were identified. This was probably because of the highly stringent statistical test used, based on the outliers searching method [93]. When an alternative approach was used, based on the SOM algorithm [94], 146 genes were found, of which several were upregulated due to P stress [37]. This is an example of how the use of multiple statistical tests might improve the reach of large-scale gene expression profiling. Based on this larger set of genes, it was clear that P starvation triggered oxidative stress, since genes involved in the detoxification of reactive oxygen species, such as those encoding a gluthatione S-transferase and a superoxide dismutase, were induced. The role of GTPases in sugarcane responses to low P was pointed out by the differential expression of several small GTPases and their regulators, one Ran GTPase activator, one Rho GTPase activator, and one Rho GDP dissociation inhibitor. P starvation repressed one homologue of an auxin-repressed protein. The authors [37] found an interesting link between this protein and the fact that P-starvation in *Arabidopsis* caused an increase in the number of lateral roots, which is linked to increased auxin sensitivity. Interestingly, MeJA treatment also repressed the expression of this sugarcane gene, again showing a complex cross-talk between the hormones. Another indication of the hormonal regulation of the root architecture in response to low P levels was the repression of an EIL transcription factor, which was involved in root development in rice. The wide array of genes induced by P stress in sugarcane were in line with the complex responses observed in other species, such as tomato [95], *Arabidopsis* [96], white lupin [97], and rice [98].

6. THE SEARCH FOR REGULATORS OF SUGAR SYNTHESIS, TRANSPORT, AND ACCUMULATION

CO_2 fixed during photosynthesis is used to synthesize carbohydrates [1, 99, 100]. Several adaptations were developed by some grass species, such as sugarcane and maize, aiming at optimizing CO_2 fixation for carbohydrate biosynthesis. They developed a distinct carbon cycle, which defines them as the "C4 plants" [101]. The compound transported in bundle sheath cells (malate or aspartate), or the compound returned to the mesophyll cells (alanine or pyruvate), varies between the species. Also, different enzymes are involved in the decarboxylation reactions: phosphoenolpyruvate carboxykinase, NAD malic enzyme, and NADP malic enzyme (which is the case with sugarcane) [102]. In sugarcane leaves, CO_2 fixation starts in the mesophyll cells, where CO_2 is combined with phosphoenol pyruvate acid in a reaction catalysed by phosphenolpyruvate carboxylase. The resulting C4 compound, oxaloacetate, is converted to malate by NADP-malate dehydrogenase. Malate is transported to the bundle sheath cells and decarboxylated by the NADP-malic enzyme, releasing (and concentrating) CO_2 for the RuBisCo action, which catalyzes the carboxylation of CO_2 with ribulose-1,5-bisphosphate, the first step in the Benson-Calvin cycle. Glyceraldehyde 3-phosphate is formed and after several steps during which fructose-1,6-bisphosphatase (FBPase) and sucrose-phosphate synthase (SPS) play a major control role, sucrose is synthesized. Sucrose transfer to the phloem cells allows for its transport to the parenchyma cells located in the stem, the major sink tissue in sugarcane [100]. All these steps raise the possibility that the sugarcane sucrose content in the stem could be even higher, considering that it is possible, at least theoretically, to have higher rates of phloem loading transport to the stalks, to its parenchyma cells, and finally to the vacuoles of these cells, as well as the control of the use of sucrose for vegetative growth [100].

The interest in sugar transporters is obvious in sugarcane, and recent findings have indicated that sink strength is a driver for photosynthesis [103], highlighting their potential for sugarcane improvement. The SUCEST database contains nine monosaccharide and four disaccharide transporters (M. Menossi, unpublished results), and this diversity of transporters is in line with the findings that sugar transport involves either symplastic or apoplastic steps [104, 105]. An EST survey comparing transcripts from immature and mature internodes revealed transcripts encoding proteins homologous with known sugar transporters more abundant in the mature internodes [24]. The only sugarcane transporter showing high selectivity for sucrose, ShSUT1, was characterized in [106, 107]. This protein is supposed to act in the loading of sucrose from the vascular tissue into the parenchyma cells from the stem.

The large-scale analysis of gene expression in a population segregated for brix was used to identify genes associated with sucrose content [108]. The plants analyzed derived from multiple crossings among *S. officinarum* and *S. spontaneum* genotypes, and from commercial varieties selected for sugar content over 12–15 years. Sucrose accumulating internodes from field grown plants were assayed using cDNA microarrays containing 1545 elements. Transcriptome comparisons aimed at identifying differentially expressed genes were made by comparing high-sugar and low-sugar plants directly, and also by comparing high-sugar and low-sugar internodes. A total of 125 genes were found to have expression patterns correlated with sugar content. Genes encoding SNF-related kinases and involved in auxin signalling were found, providing insights into the regulatory network that might control sucrose accumulation. Intriguingly, several proteins related to stress responses, such as cytochrome P450 monoxygenases, were also found. Approximately half of the sucrose content-associated genes were found to be developmentally regulated during culm maturation, and many were related to stress responses. A comparison of this differential expression dataset with the results obtained when the plants were submitted to drought [37] revealed that approximately half of the genes identified as associated with the sucrose content were responsive to drought. They belonged to several functional categories including calcium signalling, stress responses, and protein phosphorylation. The data indicated that the sucrose accumulating tissues activate pathways during culm development, which overlap with drought and other stress responses such as cold and injury. This is corroborated by the observation that several SnRKs associated with the sucrose content and with drought belonged to the SnRK2 and SnRK3 family of kinases involved in osmotic stress responses [109].

The usefulness of evaluating progenies for gene expression studies has been reviewed by Casu et al. [110]. From their studies of a progeny contrasting for sucrose content, they showed that few of the differentially expressed genes were involved in carbohydrate metabolism. Additionally, a collection of 7409 ESTs from maturing sugarcane stems in combination with a smaller collection (1089) of ESTs from immature stems [23, 24, 111] were analyzed by bioinformatic techniques and by cDNA microarray methods, allowing for the identification of genes that are differentially regulated by stem maturity. The low level of sucrose metabolism gene expression observed indicated that when the culm matured and the sugar content increased, so sucrose synthesis and catalysis decreased. GeneChips from Affymetrix containing approximately 6,024 distinct *S. officinarum* genes were also used to study culm maturation, leading to the identification of the developmentally regulated genes involved in cellulose synthesis, cell wall metabolism, and lignification [112].

Source tissues might affect the efficiency and control of carbon fixation and allocation [113]. Gene expression in source tissues has been investigated in sugarcane using the EST analysis conducted by Ma et al. [27], and more recently using SAGE [114]. As pointed out before, the sugarcane photosynthetic carbon cycle is suspected to rely on the NADP malic enzyme pathway [115], but a high expression of a phosphoenolpyruvate carboxykinase was found in the bundle sheath cells from maize leaves, a species considered to employ the same mechanism as sugarcane [116, 117]. Recent evidence [114] indicated that the phosphoenolpyruvate carboxykinase might even be more active than the NADP malic enzyme in sugarcane leaves. This data highlighted how large-scale gene expression profiling can help in understanding complex traits such as sucrose content. In another study of sugarcane leaves, 24 genes were found to be differentially expressed in plants with high- and low-sugar content, selected from an F1 segregating population derived from a cross between two commercial sugarcane varieties [118]. Evidence that hormone signalling is related to the sucrose content was also found. One of the upregulated genes encoded an omega-3 fatty acid desaturase that might be involved in methyl jasmonate signalling. Other genes associated with high brix include a receptor-like serine/threonine kinase and a transcription factor containing an Myb domain. Surprisingly, from the 24 differentially expressed genes, 19 were more expressed in plants containing low-sugar content. Three of these genes encoded 14-3-3 like proteins, which have been found to reduce SPS activity [119, 120]. Another encoded an SNF1-related protein similar to a protein quinase that phosphorylates SPS in vitro [121] making it a target for the interaction with 14-3-3 proteins, which in turn reduces SPS activity. This data reinforced the usefulness of genomic approaches to uncover how sucrose metabolism can be regulated in sugarcane.

7. CONCLUDING REMARKS

Sugarcane cultivars tolerant to drought, cold, and poor soils are increasingly important in countries that are aiming to expand their plantations. The impending need to decrease fossil fuel usage together with the fact that ethanol is a less-pollutant renewable source of energy has renewed interest in the cultivation of sugarcane around the world, and many countries are developing an ethanol/biofuel industry. Cultivars adapted to grow in colder climates and high altitudes would be a highly attractive option. In Brazil, the largest sugarcane producer, sugarcane cultivation increased 11.20% in 2007 and the area planted increased by 7.4% [73]. This expansion occurred mostly in pastures and was possible due to the increased usage of irrigation and of new varieties adapted to the climate and to the soil of the regions. Small increases in sucrose content also contributed to the increased productivity.

Knowledge of the plant responses to drought, cold, and low levels of P help to provide a framework for improving sugarcane production using biotechnological tools. The possibility of using these genes as markers for breeding purposes or by genetic engineering of the sugarcane will certainly reduce the impact of the sugarcane crop on the environment. For example, the use of plants capable of growing in low-P soil would lead to reduced liming in the *cerrados* (savannas), which is a new agricultural frontier, and is characterized by low P availability. Additionally, as stated before, a large amount of the water used by men goes into agriculture, and improvements in sugarcane drought tolerance would reduce the impact on the water supply. A better understanding of how sugarcane plants cope with cold and drought stresses could aid in the development of cultivars better suited to particular areas.

The classical breeding of sugarcane takes 15 years of crosses and agronomical evaluation before a new cultivar is released for commercialisation. Gene discovery through the SUCEST sequencing program has been a major breakthrough for the breeding programs throughout the world, and functional studies based on cDNA arrays are uncovering pathways of plant adaptation and responses to the environment. EST-simple sequence repeats (SSRs) have been successfully used for genetic relationship analysis, extending the knowledge of the genetic diversity of sugarcane to a functional level. Development of new markers based on ESTs and their integration in genetic maps will renew breeding programs and help MAB technology speed up the breeding programs.

For over ten years now [122, 123] the directed genetic modification of sugarcane has been a reality in laboratories, and field trials have been conducted [124–127]. Genes can be silenced or overexpressed to study their function and to produce new phenotypes not possible through conventional breeding. Metabolic profiling associated with gene expression studies are certainly the future tools of the sugarcane industry. Also, the analysis of the transcriptome in transgenic plants altered for genes of interest would certainly prove to be an excellent tool to unravel sugarcane regulatory networks associated with important traits.

ACKNOWLEDGMENTS

This work is supported by Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

REFERENCES

[1] P. H. Moore, "Temporal and spatial regulation of sucrose accumulation in the sugarcane stem," *Australian Journal Of Plant Physiology*, vol. 22, no. 4, pp. 661–679, 1995.

[2] A. Pandey, C. R. Soccol, P. Nigam, and V. T. Soccol, "Biotechnological potential of agro-industrial residues—I: sugarcane bagasse," *Bioresource Technology*, vol. 74, no. 1, pp. 69–80, 2000.

[3] J. X. Sun, X. F. Sun, H. Zhao, and R. C. Sun, "Isolation and characterization of cellulose from sugarcane bagasse," *Polymer Degradation and Stability*, vol. 84, no. 2, pp. 331–339, 2004.

[4] A. Sangnark and A. Noomhorm, "Effect of dietary fiber from sugarcane bagasse and sucrose ester on dough and bread properties," *Lebensmittel-Wissenschaft und-Technologie*, vol. 37, no. 7, pp. 697–704, 2004.

[5] J. M. F. Paiva, W. G. Trindade, E. Frollini, and L. C. Pardini, "Carbon fiber reinforced carbon composites from renewable sources," *Polymer-Plastics Technology and Engineering*, vol. 43, no. 4, pp. 1187–1211, 2004.

[6] G. Han and Q. Wu, "Comparative properties of sugarcane rind and wood strands for structural composite manufacturing," *Forest Products Journal*, vol. 54, no. 12, pp. 283–288, 2004.

[7] G. Arceneaux, "Cultivated sugarcanes of the world and their botanical derivation," *Proceedings of the International Society of Sugar Cane Technology*, vol. 12, pp. 844–845, 1967.

[8] S. Price, "Interspecific hybridization in sugarcane breeding," *Proceedings of the International Society of Sugar Cane Technology*, vol. 12, pp. 1021–1026, 1965.

[9] B. T. Roach, "Nobilisation of sugarcane," *Proceedings of the International Society of Sugar Cane Technology*, vol. 14, pp. 206–216, 1972.

[10] A. D'Hont, "Unraveling the genome structure of polyploids using FISH and GISH; examples of sugarcane and banana," *Cytogenetic and Genome Research*, vol. 109, no. 1–3, pp. 27–33, 2005.

[11] A. D'Hont and J. C. Glaszman, "Sugarcane genome analysis with molecular markers, a first decade of research," *Proceedings of the International Society for Sugar Cane Technology*, vol. 24, pp. 556–559, 2001.

[12] M. K. Butterfield, A. D'Hont, and N. Berding, "The sugarcane genome: a synthesis of current understanding, and lessons for breeding and biotechnology," in *Proceedings of the South African Sugar Technologists' Association (SASTA '01)*, vol. 75, pp. 1–5, Durban, South Africa, July-August 2001.

[13] L. Grivet and P. Arruda, "Sugarcane genomics: depicting the complex genome of an important tropical crop," *Current Opinion in Plant Biology*, vol. 5, no. 2, pp. 122–127, 2002.

[14] R. Ming, P. H. Moore, K. K. Wu, et al., "Sugarcane improvement through breeding and biotechnology," *Plant Breeding Reviews*, vol. 27, pp. 15–118, 2006.

[15] X.-J. Wang, T. Gaasterland, and N.-H. Chua, "Genome-wide prediction and identification of cis-natural antisense transcripts in *Arabidopsis thaliana*," *Genome Biology*, vol. 6, no. 4, p. R30, 2005.

[16] H. Wang, N.-H. Chua, and X.-J. Wang, "Prediction of *trans*-antisense transcripts in *Arabidopsis thaliana*," *Genome Biology*, vol. 7, no. 10, p. R92, 2006.

[17] F. Vazquez, "*Arabidopsis* endogenous small RNAs: highways and byways," *Trends in Plant Science*, vol. 11, no. 9, pp. 460–468, 2006.

[18] J. Ma, D. J. Morrow, J. Fernandes, and V. Walbot, "Comparative profiling of the sense and antisense transcriptome of maize lines," *Genome Biology*, vol. 7, no. 3, p. R22, 2006.

[19] B. H. Zhang, X. P. Pan, Q. L. Wang, G. P. Cobb, and T. A. Anderson, "Identification and characterization of new plant microRNAs using EST analysis," *Cell Research*, vol. 15, no. 5, pp. 336–360, 2005.

[20] B. H. Zhang, X. P. Pan, and T. A. Anderson, "Identification of 188 conserved maize microRNAs and their targets," *FEBS Letters*, vol. 580, no. 15, pp. 3753–3762, 2006.

[21] J. S. Mattick and I. V. Makunin, "Non-coding RNA," *Human Molecular Genetics*, vol. 15, no. 1, pp. R17–R29, 2006.

[22] D. Carson and F. Botha, "Genes expressed in sugarcane maturing internodal tissue," *Plant Cell Reports*, vol. 20, no. 11, pp. 1075–1081, 2002.

[23] R. E. Casu, C. M. Dimmock, M. Thomas, N. Bower, and D. Knight, "Genetic and expression profiling in sugarcane," *Proceedings of the International Society for Sugar Cane Technology*, vol. 24, pp. 542–546, 2001.

[24] R. E. Casu, C. P. L. Grof, A. L. Rae, C. L. McIntyre, C. M. Dimmock, and J. M. Manners, "Identification of a novel sugar transporter homologue strongly expressed in maturing stem vascular tissues of sugarcane by expressed sequence tag and microarray analysis," *Plant Molecular Biology*, vol. 52, no. 2, pp. 371–386, 2003.

[25] D. L. Carson, B. I. Huckett, and F. C. Botha, "Sugarcane ESTs differentially expressed in immature and maturing internodal tissue," *Plant Science*, vol. 162, no. 2, pp. 289–300, 2002.

[26] A. L. Vettore, F. R. da Silva, E. L. Kemper, et al., "Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane," *Genome Research*, vol. 13, no. 12, pp. 2725–2735, 2003.

[27] H.-M. Ma, S. Schulze, S. Lee, et al., "An EST survey of the sugarcane transcriptome," *Theoretical and Applied Genetics*, vol. 108, no. 5, pp. 851–863, 2004.

[28] N. I. Bower, R. E. Casu, D. J. Maclean, A. Reverter, S. C. Chapman, and J. M. Manners, "Transcriptional response of sugarcane roots to methyl jasmonate," *Plant Science*, vol. 168, no. 3, pp. 761–772, 2005.

[29] L. Grivet, J.-C. Glaszmann, M. Vincentz, F. da Silva, and P. Arruda, "ESTs as a source for sequence polymorphism discovery in sugarcane: example of the Adh genes," *Theoretical and Applied Genetics*, vol. 106, no. 2, pp. 190–197, 2003.

[30] M. Vincentz, F. A. A. Cara, V. K. Okura, et al., "Evaluation of monocot and eudicot divergence using the sugarcane transcriptome," *Plant Physiology*, vol. 134, no. 3, pp. 951–959, 2004.

[31] L. R. Pinto, K. M. Oliveira, T. Marconi, A. A. F. Garcia, E. C. Ulian, and A. P. De Souza, "Characterization of novel sugarcane expressed sequence tag microsatellites and their comparison with genomic SSRs," *Plant Breeding*, vol. 125, no. 4, pp. 378–384, 2006.

[32] C. L. McIntyre, M. Jackson, G. M. Cordeiro, et al., "The identification and characterisation of alleles of sucrose phosphate synthase gene family III in sugarcane," *Molecular Breeding*, vol. 18, no. 1, pp. 39–50, 2006.

[33] G. M. Cordeiro, F. Eliott, C. L. McIntyre, R. E. Casu, and R. J. Henry, "Characterisation of single nucleotide polymorphisms in sugarcane ESTs," *Theoretical and Applied Genetics*, vol. 113, no. 2, pp. 331–343, 2006.

[34] K. M. Oliveira, T. G. Marconi, G. R. A. Margarido, et al., "Functional integrated genetic linkage map based on EST-markers for a sugarcane (*Saccharum* spp.) commercial cross," *Molecular Breeding*, vol. 20, no. 3, pp. 189–208, 2007.

[35] P. Arruda, "Sugarcane transcriptome. A landmark in plant genomics in the tropics," *Genetics and Molecular Biology*, vol. 24, no. 1–4, 2001.

[36] F. S. Papini-Terzi, F. R. Rocha, R. Z. Nicoliello Vêncio, et al., "Transcription profiling of signal transduction-related genes in sugarcane tissues," *DNA Research*, vol. 12, no. 1, pp. 27–38, 2005.

[37] F. R. Rocha, F. S. Papini-Terzi, M. Y. Nishiyama Jr., et al., "Signal transduction-related responses to phytohormones and environmental challenges in sugarcane," *BMC Genomics*, vol. 8, p. 71, 2007.

[38] M. Rossi, P. G. de Araujo, and M.-A. Van-Sluys, "Survey of transposable elements in sugarcane expressed sequence tags (ESTs)," *Genetics and Molecular Biology*, vol. 24, no. 1–4, pp. 147–154, 2001.

[39] P. G. de Araujo, M. Rossi, E. M. de Jesus, et al., "Transcriptionally active transposable elements in recent hybrid sugarcane," *The Plant Journal*, vol. 44, no. 5, pp. 707–717, 2005.

[40] J. M. Casacuberta and N. Santiago, "Plant LTR-retrotransposons and MITEs: control of transposition and impact on the evolution of plant genes and genomes," *Gene*, vol. 311, no. 1-2, pp. 1–11, 2003.

[41] J. A. George, P. G. DeBaryshe, K. L. Traverse, S. E. Celniker, and M.-L. Pardue, "Genomic organization of the *Drosophila* telomere retrotransposable elements," *Genome Research*, vol. 16, no. 10, pp. 1231–1240, 2006.

[42] N. L. Saccaro Jr., M.-A. Van-Sluys, A. de Mello Varani, and M. Rossi, "*MudrA*-like sequences from rice and sugarcane cluster as two *bona fide* transposon clades and two domesticated transposases," *Gene*, vol. 392, no. 1-2, pp. 117–125, 2007.

[43] M. Rossi, P. G. de Araujo, E. M. de Jesus, A. M. Varani, and M.-A. Van-Sluys, "Comparative analysis of *Mutator*-like transposases in sugarcane," *Molecular Genetics and Genomics*, vol. 272, no. 2, pp. 194–203, 2004.

[44] R. K. Cowan, D. R. Hoen, D. J. Schoen, and T. E. Bureau, "*MUSTANG* is a novel family of domesticated transposase genes found in diverse angiosperms," *Molecular Biology and Evolution*, vol. 22, no. 10, pp. 2084–2089, 2005.

[45] G. J. M. Beckers and S. H. Spoel, "Fine-tuning plant defence signalling: salicylate versus jasmonate," *Plant Biology*, vol. 8, no. 1, pp. 1–10, 2006.

[46] W. F. Broekaert, S. L. Delauré, M. F. C. De Bolle, and B. P. A. Cammue, "The role of ethylene in host-pathogen interactions," *Annual Review of Phytopathology*, vol. 44, pp. 393–416, 2006.

[47] J. Kahl, D. H. Siemens, R. J. Aerts, et al., "Herbivore-induced ethylene suppresses a direct defense but not a putative indirect defense against an adapted herbivore," *Planta*, vol. 210, no. 2, pp. 336–342, 2000.

[48] J. J. V. Cavalcante, C. Vargas, E. M. Nogueira, et al., "Members of the ethylene signalling pathway are regulated in sugarcane during the association with nitrogen-fixing endophytic bacteria," *Journal of Experimental Botany*, vol. 58, no. 3, pp. 673–686, 2007.

[49] K. Shiojiri, K. Kishimoto, R. Ozawa, et al., "Changing green leaf volatile biosynthesis in plants: an approach for improving plant resistance against both herbivores and pathogens," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 45, pp. 16672–16676, 2006.

[50] M. C. Falco, P. A. S. Marbach, P. Pompermayer, F. C. C. Lopes, and M. C. Silva-Filho, "Mechanisms of sugarcane response to herbivory," *Genetics and Molecular Biology*, vol. 24, no. 1–4, pp. 113–122, 2001.

[51] M. R. Lambais, "*In silico* differential display of defense-related expressed sequence tags from sugarcane tissues infected with diazotrophic endophytes," *Genetics and Molecular Biology*, vol. 24, no. 1–4, pp. 103–111, 2001.

[52] E. de Matos Nogueira, F. Vinagre, H. P. Masuda, et al., "Expression of sugarcane genes induced by inoculation with *Gluconacetobacter diazotrophicus* and *Herbaspirillum rubrisubalbicans*," *Genetics and Molecular Biology*, vol. 24, no. 1–4, pp. 199–206, 2001.

[53] F. Vinagre, C. Vargas, K. Schwarcz, et al., "SHR5: a novel plant receptor kinase involved in plant-N_2-fixing endophytic bacteria association," *Journal of Experimental Botany*, vol. 57, no. 3, pp. 559–569, 2006.

[54] O. Borrás-Hidalgo, B. P. H. J. Thomma, E. Carmona, et al., "Identification of sugarcane genes induced in disease-resistant somaclones upon inoculation with *Ustilago scitaminea* or *Bipolaris sacchari*," *Plant Physiology and Biochemistry*, vol. 43, no. 12, pp. 1115–1121, 2005.

[55] V. E. De Rosa Jr., F. T. S. Nogueira, M. Menossi, E. C. Ulian, and P. Arruda, "Identification of methyl jasmonate-responsive genes in sugarcane using cDNA arrays," *Brazilian Journal of Plant Physiology*, vol. 17, no. 1, pp. 173–180, 2005.

[56] Y. Wang and B. Fristensky, "Transgenic canola lines expresssing pea defense gene DRR206 have resistance to aggressive blackleg isolates and to *Rhizoctonia solani*," *Molecular Breeding*, vol. 8, no. 3, pp. 263–271, 2001.

[57] G. M. Souza, A. C. Q. Simoes, K. C. Oliveira, et al., "The sugarcane signal transduction (SUCAST) catalogue: prospecting signal transduction in sugarcane," *Genetics and Molecular Biology*, vol. 24, no. 1–4, pp. 25–34, 2001.

[58] Y.-Q. Gu, M. C. Wildermuth, S. Chakravarthy, et al., "Tomato transcription factors Pti4, Pti5, and Pti6 activate defense responses when expressed in *Arabidopsis*," *The Plant Cell*, vol. 14, no. 4, pp. 817–831, 2002.

[59] A. Arencibia, D. Prieto, P. Téllez, et al., "Transgenic sugarcane plants resistant to stem borer attack," *Molecular Breeding*, vol. 3, no. 4, pp. 247–255, 1997.

[60] D. P. V. Braga, E. D. B. Arrigoni, W. L. Burnquist, M. C. Silva-Filho, and E. C. Ulian, "A new approach for control of *Diatraea saccharalis* (Lepidoptera: Crambidae) through the expression of an insecticidal CryIa(b) protein in transgenic sugarcane," *Proceedings of the International Society of Sugar Cane Technology*, vol. 24, no. 2, pp. 331–336, 2001.

[61] K. A. Nutt, P. G. Allsopp, T. K. McGhie, et al., "Transgenic sugarcane with increased resistance to canegrubs," in *Proceedings of the Conference of the Australian Society of Sugar Cane Technologists*, vol. 21, pp. 171–176, Brisbane, Australia, April 1999.

[62] L.-X. Weng, H. Deng, J.-L. Xu, et al., "Regeneration of sugarcane elite breeding lines and engineering of stem borer resistance," *Pest Management Science*, vol. 62, no. 2, pp. 178–187, 2006.

[63] P. H. Chen, M. J. Lin, Z. P. Xue, and R. K. Chen, "A study on genetic transformation of GNA gene in sugarcane," *Acta*

Agriculturae Universitatis Jiangxiensis, vol. 26, no. 5, pp. 740–743, 2004.

[64] P. A. Joyce, R. B. McQualter, M. J. Bernard, and G. R. Smith, "Engineering for resistance to SCMV in sugarcane," *Acta Horticulturae*, vol. 461, pp. 385–391, 1998.

[65] I. L. Ingelbrecht, J. E. Irvine, and T. E. Mirkov, "Posttranscriptional gene silencing in transgenic sugarcane. Dissection of homology-dependent virus resistance in a monocot that has a complex polyploid genome," *Plant Physiology*, vol. 119, no. 4, pp. 1187–1197, 1999.

[66] R. B. McQualter, J. L. Dale, and G. R. Smith, "Virus derived transgenes confer resistance to Fiji disease in transgenic sugarcane plants," *Proceedings of the International Society of Sugar Cane Technology*, vol. 24, no. 2, pp. 584–585, 2001.

[67] M. C. Falco and M. C. Silva-Filho, "Expression of soybean proteinase inhibitors in transgenic sugarcane plants: effects on natural defense against *Diatraea saccharalis*," *Plant Physiology and Biochemistry*, vol. 41, no. 8, pp. 761–766, 2003.

[68] D. P. V. Braga, E. D. B. Arrigoni, M. C. Silva-Filho, and E. C. Ulian, "Expression of the Cry1Ab protein in genetically modified sugarcane for the control of *Diatraea saccharalis* (Lepidoptera: Crambidae)," *Journal of New Seeds*, vol. 5, no. 2-3, pp. 209–221, 2003.

[69] A. Soares-Costa, L. M. Beltramini, O. H. Thiemann, and F. Henrique-Silva, "A sugarcane cystatin: recombinant expression, purification, and antifungal activity," *Biochemical and Biophysical Research Communications*, vol. 296, no. 5, pp. 1194–1199, 2002.

[70] R. W. Kates and T. M. Parris, "Long-term trends and a sustainability transition," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 14, pp. 8062–8067, 2003.

[71] M. Riera, C. Valon, F. Fenzi, J. Giraudat, and J. Leung, "The genetics of adaptive responses to drought stress: abscisic acid-dependent and abscisic acid-independent signalling components," *Physiologia Plantarum*, vol. 123, no. 2, pp. 111–119, 2005.

[72] I. C. Macedo, "A energia da cana-de-açúcar. Doze estudos sobre a agroindústria da cana-de-açúcar no Brasil," São Paulo, 2005.

[73] Conab, Acompanhamento da safra brasileira, 2007.

[74] Y. Narusaka, M. Narusaka, M. Seki, et al., "Crosstalk in the responses to abiotic and biotic stresses in *Arabidopsis*: analysis of gene expression in *cytochrome P450* gene superfamily by cDNA microarray," *Plant Molecular Biology*, vol. 55, no. 3, pp. 327–342, 2004.

[75] H. Abe, K. Yamaguchi-Shinozaki, T. Urao, T. Iwasaki, D. Hosokawa, and K. Shinozaki, "Role of *Arabidopsis* MYC and MYB homologs in drought- and abscisic acid-regulated gene expression," *The Plant Cell*, vol. 9, no. 10, pp. 1859–1868, 1997.

[76] L.-S. P. Tran, K. Nakashima, Y. Sakuma, et al., "Isolation and functional analysis of *Arabidopsis* stress-inducible NAC transcription factors that bind to a drought-responsive *cis*-element in the *early responsive to dehydration stress 1* promoter," *The Plant Cell*, vol. 16, no. 9, pp. 2481–2498, 2004.

[77] K. Yamaguchi-Shinozaki and K. Shinozaki, "Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses," *Annual Review of Plant Biology*, vol. 57, pp. 781–803, 2006.

[78] M. Koizumi, K. Yamaguchi-Shinozaki, H. Tsuji, and K. Shinozaki, "Structure and expression of two genes that encode distinct drought-inducible cysteine proteinases in *Arabidopsis thaliana*," *Gene*, vol. 129, no. 2, pp. 175–182, 1993.

[79] M. F. Thomashow, "So what's new in the field of plant cold acclimation? Lots!," *Plant Physiology*, vol. 125, no. 1, pp. 89–93, 2001.

[80] Y.-C. Du, A. Nose, and K. Wasano, "Thermal characteristics of C_4 photosynthetic enzymes from leaves of three sugarcane species differing in cold sensitivity," *Plant and Cell Physiology*, vol. 40, no. 3, pp. 298–304, 1999.

[81] F. T. S. Nogueira, V. E. De Rosa Jr., M. Menossi, E. C. Ulian, and P. Arruda, "RNA expression profiles and data mining of sugarcane response to low temperature," *Plant Physiology*, vol. 132, no. 4, pp. 1811–1824, 2003.

[82] A. N. Olsen, H. A. Ernst, L. L. Leggio, and K. Skriver, "NAC transcription factors: structurally distinct, functionally diverse," *Trends in Plant Science*, vol. 10, no. 2, pp. 79–87, 2005.

[83] R. Vicentini and M. Menossi, "Pipeline for macro- and microarray analyses," *Brazilian Journal of Medical and Biological Research*, vol. 40, no. 5, pp. 615–619, 2007.

[84] S. G. Møller and N.-H. Chua, "Interactions and intersections of plant signaling pathways," *Journal of Molecular Biology*, vol. 293, no. 2, pp. 219–234, 1999.

[85] S. Gazzarrini and P. Mccourt, "Cross-talk in plant hormone signalling: what *Arabidopsis* mutants are telling us," *Annals of Botany*, vol. 91, no. 6, pp. 605–612, 2003.

[86] G. Hagen, A. Kleinschmidt, and T. Guilfoyle, "Auxin-regulated gene expression in intact soybean hypocotyl and excised hypocotyl sections," *Planta*, vol. 162, no. 2, pp. 147–153, 1984.

[87] J. P. Stafstrom, B. D. Ripley, M. L. Devitt, and B. Drake, "Dormancy-associated gene expression in pea axillary buds. Cloning and expression of PsDRM1 and PsDRM2," *Planta*, vol. 205, no. 4, pp. 547–552, 1998.

[88] M. Kasuga, S. Miura, K. Shinozaki, and K. Yamaguchi-Shinozaki, "A combination of the *Arabidopsis* DREB1A gene and stress-inducible *rd29A* promoter improved drought- and low-temperature stress tolerance in tobacco by gene transfer," *Plant & Cell Physiology*, vol. 45, no. 3, pp. 346–350, 2004.

[89] H. Hu, M. Dai, J. Yao, et al., "Overexpressing a NAM, ATAF, and CUC (NAC) transcription factor enhances drought resistance and salt tolerance in rice," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 35, pp. 12987–12992, 2006.

[90] J. P. Lynch and S. B. St.Clair, "Mineral stress: the missing link in understanding how global climate change will affect plants in real world soils," *Field Crops Research*, vol. 90, no. 1, pp. 101–115, 2004.

[91] K. G. Raghothama, "Phosphate acquisition," *Annual Review of Plant Physiology and Plant Molecular Biology*, vol. 50, pp. 665–693, 1999.

[92] F. Blackburn, *Sugarcane*, Longman, New York, NY, USA, 1984.

[93] R. Z. N. Vêncio and T. Koide, "HTself: self-self based statistical test for low replication microarray studies," *DNA Research*, vol. 12, no. 3, pp. 211–214, 2005.

[94] P. Tamayo, D. Slonim, J. Mesirov, et al., "Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96, no. 6, pp. 2907–2912, 1999.

[95] Y.-H. Wang, D. F. Garvin, and L. V. Kochian, "Rapid induction of regulatory and transporter genes in response to phosphorus, potassium, and iron deficiencies in tomato roots. Evidence for cross talk and root/rhizosphere-mediated signals," *Plant Physiology*, vol. 130, no. 3, pp. 1361–1370, 2002.

[96] J. P. Hammond, M. R. Broadley, and P. J. White, "Genetic responses to phosphorus deficiency," *Annals of Botany*, vol. 94, no. 3, pp. 323–332, 2004.

[97] C. Uhde-Stone, K. E. Zinn, M. Ramirez-Yáñez, A. Li, C. P. Vance, and D. L. Allan, "Nylon filter arrays reveal differential gene expression in proteoid roots of white lupin in response to phosphorus deficiency," *Plant Physiology*, vol. 131, no. 3, pp. 1064–1079, 2003.

[98] J. Wasaki, R. Yonetani, S. Kuroda, et al., "Transcriptomic analysis of metabolic changes by phosphorus stress in rice plant roots," *Plant, Cell & Environment*, vol. 26, no. 9, pp. 1515–1523, 2003.

[99] J. E. Lunn and R. T. Furbank, "Tansley Review No. 105: sucrose biosynthesis in C_4 plants," *New Phytologist*, vol. 143, no. 2, pp. 221–237, 1999.

[100] C. P. L. Grof and J. A. Campbell, "Sugarcane sucrose metabolism: scope for molecular manipulation," *Australian Journal of Plant Physiology*, vol. 28, no. 1, pp. 1–12, 2001.

[101] N. J. Brown, K. Parsley, and J. M. Hibberd, "The future of C_4 research—Maize, *Flaveria* or *Cleome*?" *Trends in Plant Science*, vol. 10, no. 5, pp. 215–221, 2005.

[102] L. Taiz and E. Zeiger, *Plant Physiology*, M S. Sunderland, Ed., pp. 214–215, Sinauer Associates, Sunderland, Mass, USA, 1998.

[103] A. J. McCormick, M. D. Cramer, and D. A. Watt, "Sink strength regulates photosynthesis in sugarcane," *New Phytologist*, vol. 171, no. 4, pp. 759–770, 2006.

[104] J. W. Patrick, "Phloem Unloading: sieve element unloading and post-sieve element transport," *Annual Review of Plant Physiology and Plant Molecular Biology*, vol. 48, pp. 191–222, 1997.

[105] S. Lalonde, D. Wipf, and W. B. Frommer, "Transport mechanisms for organic forms of carbon and nitrogen between source and sink," *Annual Review of Plant Biology*, vol. 55, pp. 341–372, 2004.

[106] A. L. Rae, J. M. Perroux, and C. P. L. Grof, "Sucrose partitioning between vascular bundles and storage parenchyma in the sugarcane stem: a potential role for the ShSUT1 sucrose transporter," *Planta*, vol. 220, no. 6, pp. 817–825, 2005.

[107] A. Reinders, A. B. Sivitz, A. Hsi, C. P. L. Grof, J. M. Perroux, and J. M. Ward, "Sugarcane ShSUT1: analysis of sucrose transport activity and inhibition by sucralose," *Plant, Cell & Environment*, vol. 29, no. 10, pp. 1871–1880, 2006.

[108] F. S. Papini-Terzi, J. M. Felix, F. R. Rocha, et al., "The sucestfun project: identifying genes that regulate sucrose content in sugarcane plants," *Proceedings of International Society of Sugar Cane Technologists*, vol. 26, 2007.

[109] M. Boudsocq and C. Laurière, "Osmotic signaling in plants. Multiple pathways mediated by emerging kinase families," *Plant Physiology*, vol. 138, no. 3, pp. 1185–1194, 2005.

[110] R. E. Casu, J. M. Manners, G. D. Bonnett, et al., "Genomics approaches for the identification of genes determining important traits in sugarcane," *Field Crops Research*, vol. 92, no. 2-3, pp. 137–147, 2005.

[111] R. E. Casu, C. M. Dimmock, S. C. Chapman, et al., "Identification of differentially expressed transcripts from maturing stem of sugarcane by in silico analysis of stem expressed sequence tags and gene expression profiling," *Plant Molecular Biology*, vol. 54, no. 4, pp. 503–517, 2004.

[112] R. E. Casu, J. M. Jarmey, G. D. Bonnett, and J. M. Manners, "Identification of transcripts associated with cell wall metabolism and development in the stem of sugarcane by Affymetrix GeneChip Sugarcane Genome Array expression profiling," *Functional & Integrative Genomics*, vol. 7, no. 2, pp. 153–167, 2007.

[113] D. A. Watt, A. J. McCormics, C. Govender, et al., "Increasing the utility of genomics in unraveling sucrose accumulation," *Field Crops Research*, vol. 92, no. 2-3, pp. 149–158, 2005.

[114] T. Calsa Jr. and A. Figueira, "Serial analysis of gene expression in sugarcane (*Saccharum* spp.) leaves revealed alternative C_4 metabolism and putative antisense transcripts," *Plant Molecular Biology*, vol. 63, no. 6, pp. 745–762, 2007.

[115] J. R. Bowyer and R. C. Leegood, "Photosynthesis," in *Plant Biochemistry*, P. M. Day and J. B. Harbone, Eds., pp. 49–110, Academic Press, San Diego, Calif, USA, 1997.

[116] T. Furumoto, S. Hata, and K. Izui, "cDNA cloning and characterization of maize phosphoenolpyruvate carboxykinase, a bundle sheath cell-specific enzyme," *Plant Molecular Biology*, vol. 41, no. 3, pp. 301–311, 1999.

[117] T. Furumoto, S. Hata, and K. Izui, "Isolation and characterization of cDNAs for differentially accumulated transcripts between mesophyll cells and bundle sheath strands of maize leaves," *Plant & Cell Physiology*, vol. 41, no. 11, pp. 1200–1209, 2000.

[118] J. M. Felix, "Análise da expressão gênica envolvida no metabolismo de sacarose em cana-de-açúcar (*Saccharum* spp.)," Campinas: Universidade Estadual de Campinas, 2006.

[119] D. Toroser, G. S. Athwal, and S. C. Huber, "Site-specific regulatory interaction between spinach leaf sucrose-phosphate synthase and 14-3-3 proteins," *FEBS Letters*, vol. 435, no. 1, pp. 110–114, 1998.

[120] S. C. Huber, D. Toroser, H. Winter, G. S. Athwal, and J. L. Huber, "Regulation of plant metabolism by protein phosphorylation. Possible regulation of sucrose-phosphate synthase by 14-3-3 proteins," in *Proceedings of the 11th International Photosynthesis Congress*, Kluwer Academic Publishers, Budapest, Hungary, August 1998.

[121] C. Sugden, P. G. Donaghy, N. G. Halford, and D. G. Hardie, "Two SNF1-related protein kinases from spinach leaf phosphorylate and inactivate 3-hydroxy-3-methylglutaryl-coenzyme a reductase, nitrate reductase, and sucrose phosphate synthase in vitro," *Plant Physiology*, vol. 120, no. 1, pp. 257–274, 1999.

[122] R. Bower and R. G. Birch, "Transgenic sugarcane plants via microprojectile bombardment," *The Plant Journal*, vol. 2, no. 3, pp. 409–416, 1992.

[123] M. Gallo-Meagher and J. E. Irvine, "Herbicide resistant transgenic sugarcane plants containing the bar gene," *Crop Science*, vol. 36, no. 5, pp. 1367–1374, 1996.

[124] M. C. Falco, A. Tulmann Neto, and E. C. Ulian, "Transformation and expression of a gene for herbicide resistance in a Brazilian sugarcane," *Plant Cell Reports*, vol. 19, no. 12, pp. 1188–1194, 2000.

[125] M. Manickavasagam, A. Ganapathi, V. R. Anbazhagan, et al., "*Agrobacterium*-mediated genetic transformation and development of herbicide-resistant sugarcane (*Saccharum* species hybrids) using axillary buds," *Plant Cell Reports*, vol. 23, no. 3, pp. 134–143, 2004.

[126] N. B. Leibbrandt and S. J. Snyman, "Stability of gene expression and agronomic performance of a transgenic herbicide-resistant sugarcane line in South Africa," *Crop Science*, vol. 43, no. 2, pp. 671–677, 2003.

[127] M. Butterfield, J. Irvine, M. Valdez Garza, and T. Mirkov, "Inheritance and segregation of virus and herbicide resistance transgenes in sugarcane," *Theoretical and Applied Genetics*, vol. 104, no. 5, pp. 797–803, 2002.

Hindawi Publishing Corporation
International Journal of Plant Genomics
Volume 2008, Article ID 793158, 22 pages
doi:10.1155/2008/793158

Review Article

Soybean Genomics: Developments through the Use of Cultivar "Forrest"

David A. Lightfoot

Department of Plant Soil and General Agriculture, Center for Excellence, The Illinois Soybean Center, Southern Illinois University at Carbondale, Carbondale, IL 62901-4415, USA

Correspondence should be addressed to David A. Lightfoot, ga4082@siu.edu

Received 25 April 2007; Accepted 26 December 2007

Recommended by P. Gupta

Legume crops are particularly important due to their ability to support symbiotic nitrogen fixation, a key to sustainable crop production and reduced carbon emissions. Soybean (*Glycine max*) has a special position as a major source of increased protein and oil production in the common grass-legume rotation. The cultivar "Forrest" has saved US growers billions of dollars in crop losses due to resistances programmed into the genome. Moreover, since Forrest grows well in the north-south transition zone, breeders have used this cultivar as a bridge between the southern and northern US gene pools. Investment in Forrest genomics resulted in the development of the following research tools: (i) a genetic map, (ii) three RIL populations ($96 > n > 975$), (iii) ~200 NILs, (iv) 115,220 BACs and BIBACs, (v) a physical map, (vi) 4 different minimum tiling path (MTP) sets, (vii) 25,123 BAC end sequences (BESs) that encompass 18.5 Mbp spaced out from the MTPs, and 2 000 microsatellite markers within them (viii) a map of 2,408 regions each found at a single position in the genome and 2104 regions found in 2 or 4 similar copies at different genomic locations (each of >150 kbp), (ix) a map of homoeologous regions among both sets of regions, (x) a set of transcript abundance measurements that address biotic stress resistance, (xi) methods for transformation, (xii) methods for RNAi, (xiii) a TILLING resource for directed mutant isolation, and (xiv) analyses of conserved synteny with other sequenced genomes. The SoyGD portal provides access to the data. To date these resources assisted in the genomic analysis of soybean nodulation and disease resistance. This review summarizes the resources and their uses.

1. INTRODUCTION

The soybean cultivar "Forrest," a product of a USDA breeding program, represents a determinate, Southern germplasm [1]. It was the first cultivar to possess soybean cyst nematode (SCN) resistance associated with high yield, and is believed to have played a key role in saving billions of US dollars during 1970s and 1980s that would have otherwise been lost, either due to SCN or due to the poor agronomic performance of earlier SCN resistant cultivars (see [2] and references therein). Forrest was an important parent of modern cultivars, "Hartwig," "Ina" and many others that have an improved SCN resistance gene from PI437654 introgressed into their genome [3–5]. Forrest was also central to an understanding of the genetics of resistance to sudden death syndrome, an important new disease of soybean [6–9].

Forrest is also one of the two cultivars (the other being "Williams 82"), providing the majority of genomic tools for soybean, available in USA (Figure 1) [10, 11]. These two cultivars provide models for soybean genomics research in the same way as are the cultivars *Col* and *Ler* in *Arabidopsis thaliana* or Mo17 and B73 in *Zea mays*. However, since the genomics of "Williams 82" was recently reviewed [11], its inclusion in this article would be repetitive. The other cultivars, which represent the worldwide germplasm variation for soybean genomics, include the following: (i) "Noir 1," a Korean plant introduction (PI) [12], (ii) "Misuzudaizu," a Japanese cultivar [13], and (iii) "Suinong14," a Chinese cultivar [14]. The soybean community is committed to advance the genomics of all these cultivars, which have been used in the past as resources for genomics research. However, the intent of this review is to present an overview of the genomic resources derived from Forrest; these genomics resources enable a wide range of analyses that address several fundamental questions, like the following: (i) what is the source of genetic variation in soybean improvement? [15];

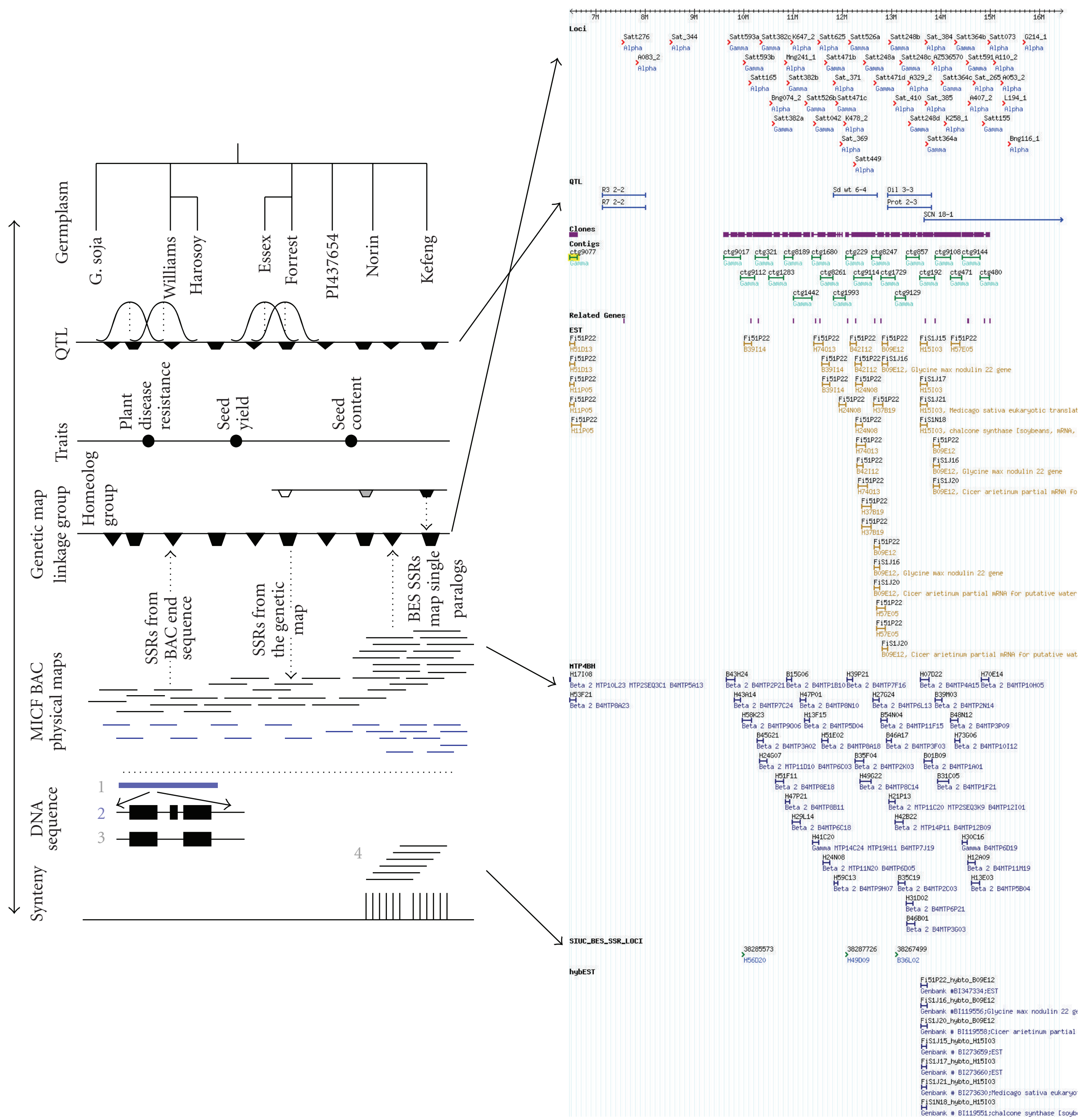


Figure 1: Soybean genomic resources and products schematic for Forrest (A) compared to the SoyGD representation (B). Panel A. Germplasm that are exemplars of soybean genetic diversity are shown. Selected germplasm encompass in mapped QTL a wide variety of traits placed on the composite genetic map. BAC libraries exist for many of the germplasm sources. Forrest BACs (shown in black) form the basis of an MICF physical map with 6-fold coverage. A region of conserved duplication (12-fold coverage) is shown on the right of the figure. In this region, fingerprinted clones from two homoeologous linkage groups coalesce. Genetic markers identified in, or derived from, BAC end sequences (BESs) will separate some of the duplicated conserved regions. Genetic markers anchored from map to BAC are of little use in conserved duplicated regions. BACs from diverse germplasm are shown as blue bars. There are 3 levels of DNA sequence envisioned. At level 1, BESs provide a sequence every 10–15 kbp with which to identify gene rich regions for later complete sequence determination (level 2). Arrayed BAC end sequences will be used to identify conserved syntenic regions in the genomes of model plant species. This information will also separate some of the duplicated conserved regions in soybean. Panel B. Shown are the chromosome (cursor), DNA markers (top row of features, red); QTL in the region (second row, blue); coalesced clones (purple) comprising the anchored contigs (third row, green); BAC end sequences (fourth row black); BESs encoding gene fragments (fifth row, puce); EST hybridizations to MTP2BH (sixth row gold); MTP4BH clones (seventh row, dark blue); BESs-derived SSR (eight row, green).

(ii) what is the role of variation in regions of genome duplication in paleopolyploid species? [16]; (iii) how does the nodulation of legumes work? [17]; (iv) why are protein and oil contents of seed inversely related? [18, 19]; (v) why are seed yield and disease resistance so hard to combine? [4, 5, 15, 20]; (vi) why is seed isoflavone content limited below 6 mg/kg? [18, 21–24]; (vii) how does partial resistance to disease work [6–9, 18]? It is believed that the development and use of genomics tools derived from Forrest will help soybean researchers to provide answers to these questions.

2. GENETIC VARIATION BETWEEN FORREST AND OTHER CULTIVARS

An important question that received the attention of soybean researchers in the past is how much sequence variation one can expect between Forrest and other cultivars, if many are to be sequenced. This variation is extensive (about 1 bp difference per 100–300 bp), when judged by using the criteria like the following: (i) the coefficient of parentage [25], (ii) the number of shared RFLP bands [26], (iii) polymorphism among microsatellite markers [27], and (iv) DNA sequence comparisons (Figure 2). In soybean, the degree of linkage disequilibria among loci is high, extending over distances that range from 50 kbp to 150 kbp [28]. Few meioses have occurred within these regions to reshuffle the gene or DNA sequences, because soybean is largely an inbreeding crop. In recent times, only seven or eight crosses have been made, starting from the time when the PIs were collected to the development of most modern US cultivars (Figure 3). Therefore, in different parts of the genome, LD encompasses large segments and sets of genes.

2.1. *The Essex × Forrest population*

A soybean recombinant inbred line (RIL) mapping population (Reg. no. MP-2, NSL 431663 MAP) involving Forrest was recently developed from the cross "Essex MAP" (PI 636326 MAP) × "Forrest MAP" (PI 636325 MAP) [10]. This RIL population was used for constructing a genetic map [9, 24, 30] that has been used extensively for an analysis of marker-trait associations [7–9, 24, 30–38]. The genetic marker data encompass thousands of polymorphic markers and tens of thousands of sequence-tagged site (STS) that were collected at SIUC by Dr. Lightfoot's group (Table 1) [10]. The genetic maps of E × F94 will continue to be enriched [27, 39]. The registration of this population [10] has allowed public access to the population and data generated from it worldwide.

A key feature of the above mapping population is that Essex (registered in 1973 [10]) was derived from the same southern US germplasm pool to which Forrest (registered in 1972 [1]) belongs. Consequently the RILs share identity across about 25% of their genomes, the portion that was monomorphic in both of the parents (Figure 3) [25, 26]. Further, the two cultivars were selected under similar conditions and, therefore, appear rather similar in most environments [6–10, 15–20, 30–38]. However, detailed records of maturity dates are important, since even a single day variation in maturity may influence the results of QTL analysis for many other traits [10, 41]. Since morphological and developmental traits differ very little in the population, the RILs have been used extensively to map those genes which control biochemical and physiological traits (Table 2). For example, the parents of the mapping population differ by resistance traits, which exhibit both qualitative and quantitative inheritance (Table 3).

A major limitation in using E × F population in genomics research is the small population size ($n = 100$) that could preclude fine mapping [10]. To overcome this problem, populations of near isogeneic lines (NILs; $n = 40$; Figure 3) were developed from each RIL [10, 37, 38, 43]. The NIL populations are listed in Table 1. The residual heterozygosis present in the F5 seed was largely fixed and captured in these NILs. The heterogeneity across the RILs has been measured to be 8%, which is more than the 6.25% expected among F5 lines [7, 24]. That increased heterogeneity appears to be caused by selection, since rare heterozygous plants still exist in some RILs and NILs [37, 38, 40]. Each locus that segregates in the RIL population is expected to segregate in about eight NIL populations. Therefore, each region in the genome will be segregating in about 420 lines (100 + 8 × 40), quite sufficient to create fine maps of 0.25 cM resolution (Table 4). A 0.25 cM interval represents 25–100 kbp on the physical map [16], sufficient for candidate gene identification [37, 38].

Consequent to the development of the NILs, the E × F population was used to study the genetics of a large number of quantitative traits (QTs), leading to the identification of quantitative trait loci (QTL; Table 2) underlying more than seventy different traits [24, 39, 40, 42, 44–46]. Biochemical and physiological traits included resistance to soybean sudden death syndrome (SDS) [caused by *Fusarium virguliforme*] in the US and Argentina, resistance to soybean cyst nematode (SCN; *Heterodera glycine* Ichinohe), seed yield, seed quality traits, agronomic traits, water use efficiency, manganese toxicity, aluminum toxicity, partial resistance to *Phytophthora sojae*, and insect herbivory. However, new opportunities abound because dozens of traits for resistance to pests and pathogens segregate in the population but were not yet mapped [10]. Further, the concentrations of many secondary metabolites among lines vary widely during development and among different organs [47]. Pesticide uptake, metabolism and degradation rates also vary among lines (unpublished). Preliminary studies have shown the link between the genome, proteome, and metabolome (the interactome), which can be further explored in these segregating populations [48]. Therefore, E × F will eventually be used to map thousands of QTL for hundreds of QT.

Importantly, the NILs that have been developed from each RIL for fine mapping also allow confirmation of QTL detected in the RIL population. For instance, *cqSDS001* was assigned to a QTL confirmed by NILs derived from Ripley [49], but earlier detected through RILs derived from Flyer [50] and "Pyramid" [6, 33]. The QTL have also been renamed under the new rules for QTL adopted by the Soybean Genetic Committee in 2006 [51], as a result of

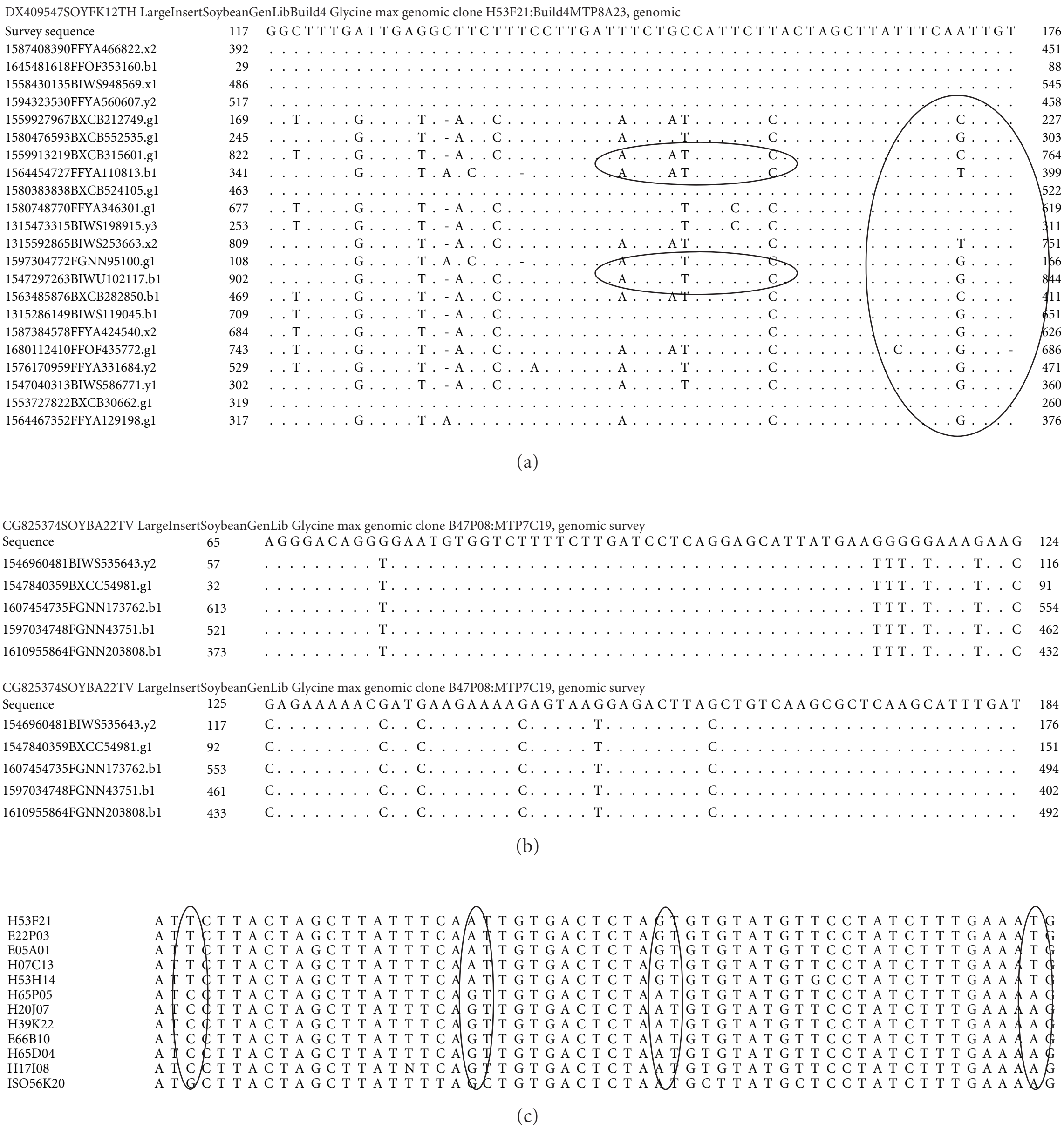


Figure 2: Comparison of MegaBlast analysis of an unduplicated region and a twice duplicated region as inferred by the fingerprint physical map (a). Analysis of the BESs from H53F21 in quadruplicated contig 9077. These BESs contained a very common repeat with 400 copies per haploid genome. Sequence analysis supported the inference of four copies of the region per haploid genome made from BAC fingerprint data (a). MegaBlast of H53F21 (Build4MTP8A23, gi89261445) against 7.3 million reads with repeated masking gave 7 identical matches among 24 homoeologous sequences. Cluster 1 was composed of traces ending in ...822,...160,...569,...607,...662,...749, and ...105 that shared A at position 172 (circled). Homoeolog specific variations (polymorphisms) were evident among the 4 clusters inferred. Cluster 2 was composed of clones ending in 749, 850, and 601 that shared C at position 172. Cluster 3 was composed of clones ending in 100, 117, and 535 that shared G at position 172. Cluster 4 also had G at that position. TreeCluster analysis showed the most similar homoeologs clustered into 4 separate sets as expected for regions duplicated twice (circled) (b). Analysis of the BESs from B47P08 in contig 321 from an unduplicated region. Sequence analysis supported the inferrence of an unduplicated region made from fingerprints at 90% sequence identity (c). The sequences found among BACs resequenced from contig 9077 showing a set of SNHs (HSVs) separated two groups of the four inferred to be present: the A cluster and the G cluster (adapted from [29]).

which *cqRfs1*, *cqRfs2*, and *cqRfs4* were renamed as *cqSDS003*, *cqSDS002*, and *cqSDS004*, respectively.

The molecular linkage map, the RILs, and the NILs were used during the positional cloning of *nts1, GmNARK* [50], *Rpg1* [17, 35], *Rhg1*, [38] *Rhg4*, [52], and *Rfs2* [37]. Many opportunities for further gene isolations exist. Tables 2 and 3 list some of the known phenotypes that differ between the parents and segregate among the lines and that are candidates for gene isolation. The RIL and NIL populations provide sets of recombination events that can be used to identify the

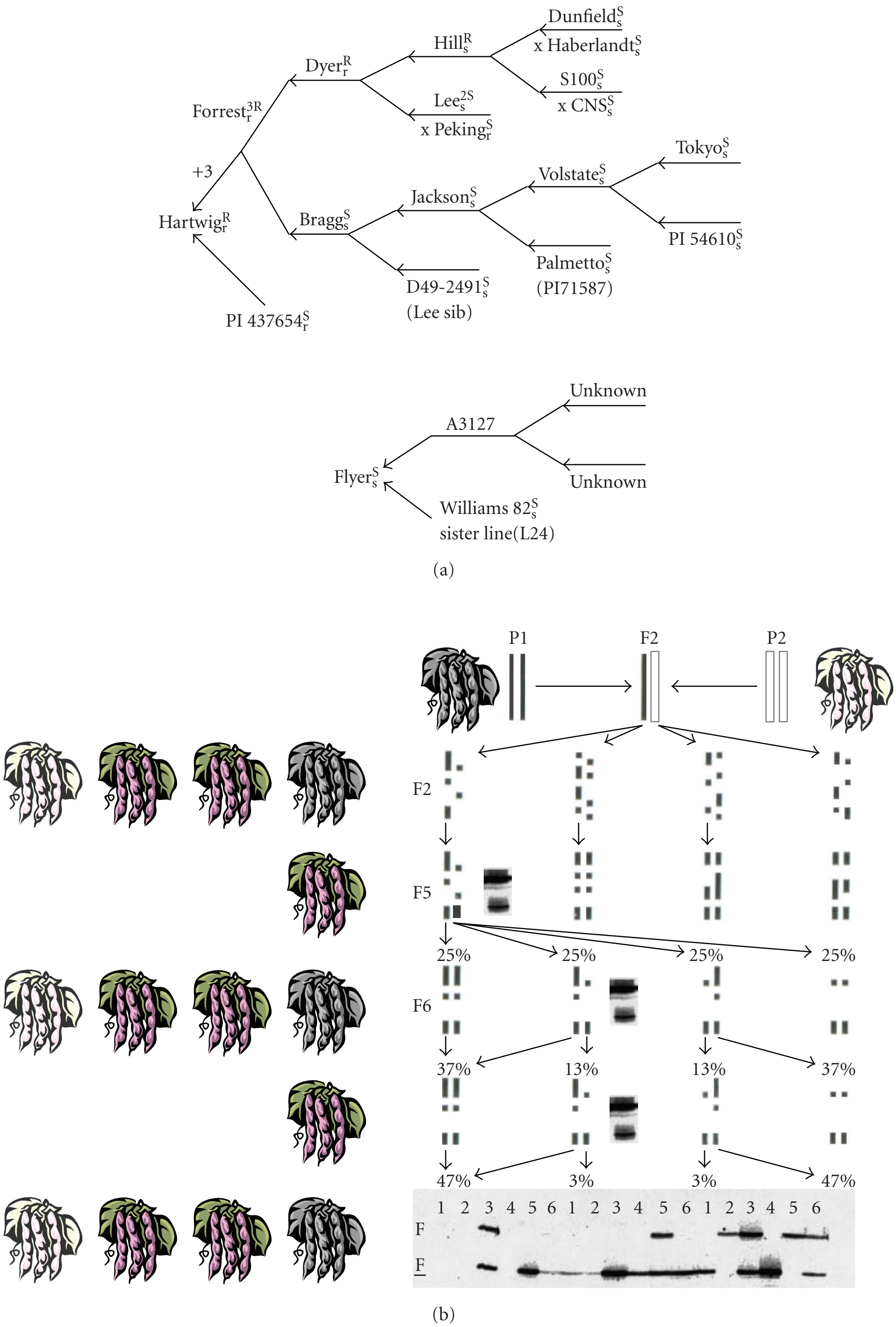


Figure 3: Genetic systems used with Forrest germplasm and the inbred soybean crop (a). The ancestry of Forrest and Hartwig showing the known cultivars that were crossed and the relationship between Flyer and Williams 82 (b). A diagram showing how NILs derived from RILs fix most loci but allow the continued segregation of heterozygous regions in inbred crops like soybean. The effect is to Mendelize a few of the loci contributing to QT while causing the majority to be fixed. A dark pod parent was crossed with a light colored pod parent; the F1 heterozygous type (shown as purple pods) was selfed; and F2 progeny was advanced to the F5 by selfing. A heterozygous plant at any time or heterogeneous RIL at $F_{5:7}$ or later identified is shown as purple pods. Single plants are extracted and seed increased. NILs that result may fix the heterogeneous region to the parent 1 allele, the parent 2 allele, or are still heterogeneous. Occasionally heterozygous plants are found within some heterogeneous NILs even at the $F_{5:15}$ and the progeny of such plants can be used to find new recombination events. Shown are the results with Satt309 and NIL11 plant 3 and eighteen of the progeny collected from it (adapted from [40]).

Table 1: Description of 20 linkage groups mapped in the Essex × Forrest mapping population. The map distances and markers distribution for the linkage groups were generated from analysis of the 100 F_5-derived progeny from E × F.

Linkage group	NIL[(a)] populations	Map Distance (cM)	No. of markers Total	SSR	RFLP	RAPD	BESs [(b)]	EST[(b)]	BESs [(c)] SSR
A1	6	73.8	14	4	3	7	458	13	4
A2	8	259.0	22	10	8	4	757	0	7
B1	4	164.0	16	11	2	3	234	7	5
B2	5	53.4	12	7	1	4	156	3	6
C1	4	150.1	13	10	0	3	136	0	9
C2	8	213.2	30	19	4	7	565	14	4
D1a + Q	9	140.0	17	14	0	3	625	30	3
D1b+W	8	87.4	14	8	1	5	124	1	3
D2	7	245.4	19	15	0	4	122	0	4
E	6	97.4	9	6	0	3	362	11	5
F	4	219.9	29	16	5	8	369	0	2
G	12	242.5	37	19	12	6	1126	33	5
H	8	98.3	9	6	1	2	427	9	4
I	9	116.9	16	11	0	5	192	6	3
J	7	40.7	7	3	1	3	577	3	2
K	9	150.9	18	13	0	5	590	1	4
L	8	103.8	12	9	0	3	91	3	2
M	6	105.2	10	6	1	3	87	9	4
N	3	145.1	21	9	2	10	156	0	3
O	2	116.4	13	10	0	3	566	9	0
Total	100	**2823.4**	**337**	**206**	**41**	90	7720	152	79
Unlinked	(2007)	**0**	**0**	**0**	**0**	0	10529	485	10

(a) NIL populations segregate for 2 or more regions on different chromosomes.
(b) ESTs and BESs may appear at 2 or more locations on the linkage map if they appear in homoeologous regions of different linkage groups.
(c) BESs-SSR placedon the genetic map, many more are placed in SoyGD by inference from marker anchored contigs.

positions of genes underlying QT [10]. Since all the lines self-fertilize, the populations can be used to provide an immortal resource, if seed germination ability can be regenerated every five years. This type of resource is particularly important for soybean because the draft genome sequence will be released in April 2008 (unpublished). Combining knowledge of locus positions with a comprehensive knowledge of gene content will lead to the rapid isolation of many new and economically important genes [16].

Selected lines from the E × F population that contrast for mapped QTL were also used for a variety of studies including the following: (i) to validate assays of pathogenicity [32, 53–55], (ii) to examine the effects of resistance genes on gene expression [34, 56, 57], (iii) to analyze components of drought tolerance [24, 31, 36, 42, 46, 58], (iv) to validate methods of marker assisted selection [6, 31, 59–62], and (v) to provide for germplasm releases (Figure 4) and cultivars [6, 63]. New cultivars and new methods for selection of improved soybean genotypes are among the most important spin-offs from the genomics research involving Forrest soybean. Among the selected lines, E × F78 later became LS-G96 [63] and then "Gateway 512" (Gateway Seeds, Nashville, Ill, USA). This line together with the line E × F55 was used as parents that combined moderate resistance (carrying resistance alleles at six loci) to SDS with high yield. The RIL E × F23 was released as SD-X for very high resistance to SDS [34] and good yield potential under license from Access Plant Technologies (Plymouth, Ind, USA), because it contained beneficial alleles at all eight known resistance loci. In contrast, E × F85 is susceptible to SDS as it contained no beneficial alleles at the known resistance loci. It makes a great entry for sentinel plots. For animal feed and human food, E × F52 has been used as a parent to provide very high phytoestrogen contents to progeny (unpublished), since it contained beneficial alleles at all the known loci underlying phytoestrogen content. Low phytoestrogen contents are also required for estrogen sensitive consumers; E × F89 and E × F92 were used as parents to provide parents for low phytoestrogen in the progeny (unpublished).

2.2. *Related populations flyer by hartwig (F × H) and Resnik by Hartwig (R × H)*

The F × H and R × H populations are integrated with E × F96 [10], since Forrest was the recurrent parent used to develop Hartwig (Figure 3) [62] and Essex shares many alleles with the Flyer and Resnik [15, 27]. Flyer and Resnik were sister lines derived from a cross between a Williams 82

Table 2: Ranges and means of selected mean traits measured across multiple locations and years using the RIL population and the "Essex" and "Forrest" parents. For traits 1–35 see [24]; traits 36–79 were from [39, 42] and or unpublished.

No. of trait and symbol	Unit	RIL population	
		Average	Range
1. SDS disease incidence	Score	48.5	4.4–94
2. SDS disease severity	Score	1.5	1.1–2.3
3. SDS disease index	Score	9.3	1.1–23.9
4. Soybean cyst nematode IP	(%)	53	0–100
5. Yield during SDS	$Kg\cdot ha^{-1}$	3.3	2.9–3.76
6. Seed daidzein content	$\mu g\cdot g^{-1}$	1314	874.5–2181
7. Seed genistein content	$\mu g\cdot g^{-1}$	996.8	695.5–1329
8. Seed glycitein content	$\mu g\cdot g^{-1}$	206.1	116–309
15. Total seed isoflavone content	$\mu g\cdot g^{-1}$	2516.8	1774.2–3759
21. Resistance to manganese toxicity	Scale 0–5	2.02	1.1–4.5
32. Seed yield	$Kg\ ha^{-1}$	3.44	2.64–4.13
33. Leaf trigonelline content (irrigated)	$\mu g\ g^{-1}$	98.85	59.87–126.96
34. Leaf trigonelline content (rain-fed)	$\mu g\cdot g^{-1}$	417.94	245.95–618.18
35. Flower color (white: purple)	color	43:47	na
38. Mean SDS DX in Argentina	Scale 1–10	1.6	0.1–3.1
43. Tolerance to aluminum toxicity	(%)	14	−20–37
47. Seed protein content	(%)	39.5	37.5–41.5
51. Seed oil content	(%)	18.9	18.0–20.1
55. Resistance to insect herbivory (IP)	(%)	22.3	13.0–32.5
60. Seedling root growth	mm	8.3	6–11

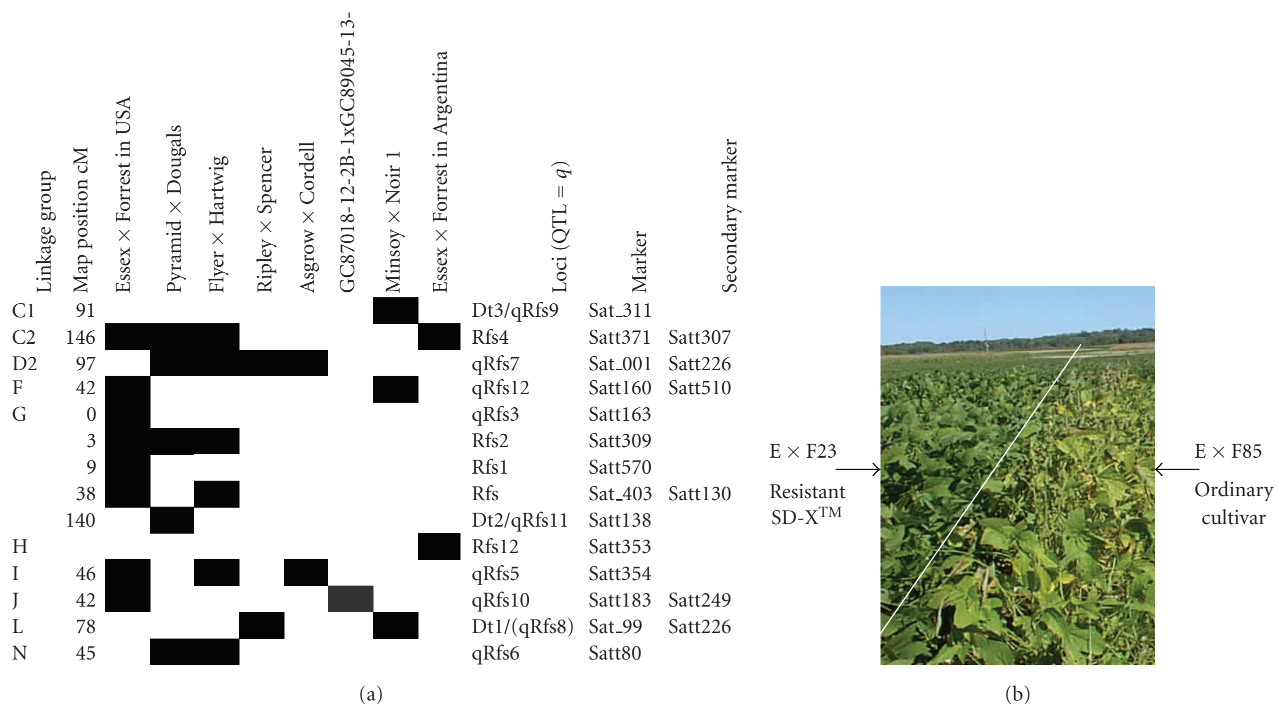


Figure 4: An example of the use of Forrest genomics resources for soybean germplasm improvement (a). Summary of the map locations of the known loci for resistance to SDS. A black rectangle indicates that the allele is segregating in that population. Nonsegregating alleles may be either fixed to the resistance or susceptibility forms (b). An example of quantitative variation for disease resistance identified in lines derived from Forrest. The resistant line RIL23, left of the line, has beneficial alleles for six QTL for resistance to *Fusarium virguliforme*. The leaf scorch associated with the fungal infection is evident in the neighboring RIL80 to the right of the white line.

Table 3: Disease resistance that segregates among the RIL and NIL population.

Disease resistance in	Causal agent
A. Forrest	
Soybean cyst nematode	*Heterodera glycines* HG type 0; races 3
Root-knot nematode	*Meloidogyne incognita*
Bacterial pustule	*Xanthomonas glycines*
Wildfire	*Pseudomonas syringae subsp. tabaci*
Target spot	*Alternaria* sp
Partial *Phytophthora* root rot	*Phytophthora sojae*
SDS root rot	*Fusarium virguliforme*
SDS leaf symptoms	Toxin
B. Essex	
Bacterial pustule	*Xanthomonas glycines*
Downy mildew	*Peronospora manshurica*
Frogeye leaf spot	*Cercospora sojina*
Purple seed stain disease	*Cercospora kikuchii*
Partial *Phytophthora* root rot	*Phytophthora sojae*
SDS leaf symptoms	Toxin
C. Hartwig	
Soybean cyst nematode	All HG Types from 1.2.3.4.5.6.7.
Root-knot nematode	*Meloidogyne incognita*
Reniform nematode	*Rotenlenchulus reniformis*
Bacterial pustule	*Xanthomonas axonopodis* pv. *glycines*
Wildfire	*Pseudomonas syringae* pv. *tabaci*
Target spot	*Corynespora cassiicol* a
Partial *Phytophthora* root rot	*Phytophthora sojae*
SDS root rot	*Fusarium virguliforme*
SDS leaf symptoms	Toxin
D. Flyer	
Powdery mildew	caused by *Microsphaera diffusa*
Purple seed strain disease	*Cercospora kikuchii*
Pod and stem blight	*Diaporthe phaselorum*
Multirace *Phytophthora* root rot	*Phytophthora sojae*
SDS leaf symptoms	Toxin

sister line and a commercial cultivar [64]. The F × H has 92 RILs and R × H has 952 RILs that have been used to confirm QTL detected in E × F96 and for fine mapping of these QTL [4, 5, 15, 50, 52]. Flyer and Resnik each contains many genes conferring resistance against *P. sojae*. Both these populations can be used to map genes underlying additional biochemical, physiological, and some agronomic traits that include the following: (i) resistance against *Phytophthora* root rot, soybean sudden death syndrome (SDS) caused by *F. virguliforme* and soybean cyst nematode (SCN), *Heterodera glycine* Ichinohe, (ii) seed yield [15, 50, 52], and (iii) seed quality traits. These RILs were also used to develop SSR markers that anchor contigs and sequence scaffolds (http://soybeangenome.siu.edu/) to the physical map [27].

3. PHENOTYPIC VARIATION BETWEEN FORREST AND OTHER CULTIVARS

One major limitation using the resources based on Forrest was the low amount of genetic variation detected in the populations based on this cultivar [65]. The implication was that the alleles detected in E × F would not be weaker variants of the major gene effects found in weedy plant introductions (PIs). It was hypothesized that, instead, the loci detected in the E × F population and in the material derived from this population perhaps represented other gene systems of lower hierarchical position and therefore lower value. Consideration of a few examples of the locations of QTL underlying phenotypic variation between Forrest and other cultivars has been informative regarding this issue. The results to date all infer that the alleles underlying QTs in Forrest are variations in the same genes as the PI alleles, if weaker in effects on QTs.

3.1. The genetics of phytoestrogen content

The phytoestrogen content of soybeans seed mainly consists of daidzein (60%) and genistein (∼30%) with small proportion of glycitein (∼10%) [66]. Analysis of germplasm and elite cultivars (18, 21–24, 67–69) indicated that phytoestrogen concentrations in some elite cultivars (∼2 mg/kg) were higher than those in many of the ancestors of cultivated soybean (∼1 mg/kg). Phytoestrogen content and profile varied with environment (year and location effect) and genotype. However, the final seed content was largely controlled by the genotype (40–60% of the variation) and is controlled by a set of about 6–12 loci [18, 24, 67]. If the content of each phytoestrogen component was controlled independently, improvements in content by genetic selection should be possible. For instance, raising glycitein content to the same amount as that of daidzein could double the total phytoestrogen content. However, because heritability of phytoestrogen content is moderate at about 40–60%, direct selection (without DNA markers) has not been very effective. Through marker-assisted selection (MAS), the phytoestrogen amounts were raised to 3.6 mg/kg, well above the amounts found in elite cultivars or weedy PIs. Here, the variation programmed by the alleles segregating in E × F population was greater than that among the entire germplasm collection.

Recently, crosses have been made betweenlines from southern Illinois and Canada having the highest phytoestrogen contents [23] and, separately, the lines having the lowest phytoestrogen content [67]. MAS exercised in the segregating populations (at F4 in 2007) should lead to improvement in phytoestrogen content. Opportunities for collaborative studies exist with sets of RILs in maturity groups that are not adapted to be grown in southern Illinois or Canada.

Table 4: Saturation mapping with markers on chromosome 18 in the 2–4 Mbp encompassing *Rhg1*, *Rfs1*, and *Rfs2* (SDS) loci with leaf and root phenotype classes shown.

Geno type	Satt214	Sat1	TMD1	Satt309	Sat185	CGG5	OI03	CTA13	Bng122	Leaf	Root
1	E	F	*E*	*E*	F	E	**F**	E	F	*S*	**R**
2	E	E	*E*	*E*	E	E	E	E	F	R	S
3	E	E	E	H	E	E	E	E	F	R	S
4	E	E	F	F	E	E	E	E	E	R	S
5	E	F	F	F	F	E	E	E	E	R	S
6	F	F	F	F	E	F	**F**	F	F	R	**R**
7	F	F	F	F	E	E	E	F	F	R	S
8	F	F	F	F	F	F	**F**	F	E	R	**R**

3.2. The genetics of seed yield, protein and oil content

The overall average increase of 1-2% per year in soybean yield witnessed during 1960–1999 was only half the yield advances achieved in corn and other out crossing crops, where genetic diversity was not limiting [68]. As one would expect, there are hundreds of loci controlling yield in soybean [69]. In view of this, half of the yield loci detected in E × F population were those which were earlier detected in other crosses [24]. These loci could each boost seed yield by 0.2 Mg/Ha. In contrast, substantial gains (0.9–1.1 Mg/Ha) can be made in soybean yield by identifying unique alleles in weedy PIs and introgressions into elite cultivars [70]. The nature of the genes altering seed yield will be an interesting product from fine map analysis and positional cloning.

The major components ofsoybean seed yield include the following: (i) protein (∼40%), (ii) oil (∼20%), (iii) structural carbohydrates (∼6%), (iv) water (∼13%), (v) soluble carbohydrates (∼14%), and (vi) other metabolites (∼7%) [71]. Metabolic changes during development driven by gene expression underlie the seed composition and yield [72]. Seed yield and composition are under polygenic control with different genes active at different stages of seed development. Seed traits are also associated with significant genotype × environment (G × E) interactions as observed in E × F population (see [15, 18, 19]). Again, the G × E interactions significantly reduce the effectiveness of visual selection based on the phenotype alone.

At harvest, seed protein content is inversely related to seed oil content and seed yield in E × F population [18, 19] as also in other germplasm (see [68]). While some loci are implicated in all the three traits, there are others which influence only one or two of the three traits. Several QTL underlying soybean yield, protein, and oil content have been mapped in both the E × F and the F × H RIL populations [5, 18]. They do correspond with loci detected in crosses between high protein weedy types and low protein adapted cultivars. Three QTL on linkage groups A1, A2 and linkage group E have been fine-mapped and localized within 0.25 cM using substitution mapping to identify the underlying genes. Isolation of these genes will partly explain the molecular basis of the genetic control of yield and its component traits. However, a danger here is that because different genes are active at different stages of seed development, one would generally map only a composite trait, based on a mean of the action of several loci. Isolation of genes by position would not be successful in this circumstance.

3.3. The genetics of Phytophthora root rot resistance

The annual soybean yield loss suffered from the root and stem rot disease caused by the oomycete pathogen, *Phytophthora sojae* is valued at about $273 million in the US [73]. Monogenic resistance due to a series of *Rps* genes has been providing a reasonable protection to the soybean crop against the pathogen over the last four decades [74]. Several mapped *Rps*-genes are known to occur in Flyer and Resnik [50, 64]. Partial, rate-reducing resistance to many races of *P. sojae* is found also in Forrest, Essex, and Hartwig. The loci providing this partial resistance were not mapped by 2007.

3.4. The genetics of SCN resistance

Soybean cyst nematodes (*Heterodera glycines* I.) are the most damaging pests of soybean worldwide [73]. Development of resistant cultivars is the only viable control measure [75]. Resistance genes have been found to be located on 17 of the 20 chromosomes by 2007. A combination of recessive genes is necessary to provide resistance against SCN populations because many are known to be capable of overcoming all known single resistance genes. SCN populations can be classified into 16 broad races or up to 1024 biotypes (HG Types) [76] based on the host responses of 8 weedy indicator lines. SCN resistance in many other adapted and weedy cultivars [9, 31] shared the same loci underlying bigeneic inheritance in E × F [20]. The E × F population was used to isolate candidate genes for those two loci (*rhg1* and *Rhg4* ; Table 4) that control resistance against SCN race 3 (HG Type 0). Alleles of the candidate genes were identified in many PIs through association studies [38, 77]. Paralogs of both these genes were found at new locations in BAC libraries and whole genome shotgun (WGS) sequences [78, 79]. They appear to be part of multigene families showing homoeology and intragenomic conserved synteny.

Three cultivars including Peking, PI437654, and Hartwig encoded 2–4 additional genes that provide additional resistances to SCN [52, 80, 81]. Peking has alleles for resistance to races 1 and 5 that were not transferred to Forrest [20].

Hartwig and PI437654 have complete resistance against all races of SCN except race 0, HG Type 1.2.3.4.5.6.7.8. The location of SCN resistance loci in F × H and R × H agreed with those found in crosses between PIs and adapted germplasm [81, 82]. Therefore, the resistance to SCN traits that are introgressed from PIs to Forrest-based germplasm is useful and the underlying genes can be isolated from Forrest.

3.5. The genetics of SDS resistance

Soybean sudden death syndrome caused by *Fusarium virguliforme* (e.g., *solani* f. sp. *glycines*) is among the most damaging syndrome of diseases affecting soybean in the US and worldwide [73]. The syndrome is composed of a root rot disease and a leaf scorch disease [53]. Development of resistant cultivars is the only viable control measure. Twelve resistance loci have now been found on 8 chromosomes (Figure 4), eight segregate in E × F [24, 44] and two additional loci segregate in F × H [5, 50]. A combination of loci is needed to provide resistance to both root rot (2 or more loci) and leaf scorch (all loci). Loci for resistance to SDS were named *Rfs* to *Rfs11* [39]. Using NILs (Table 4), a set of candidate genes for the *Rfs2* locus were identified [37]. Among these genes, a receptor like kinase [38] and a laccase [83] are being tested for their ability to provide resistance following transformation and mutation (unpublished). However, the presence of a pair of syntenic genes on linkage group O with similar DNA sequences (84%) and encoding nearly identical amino acid sequences (98%) complicates the analysis following reverse genetics approach.

One of the two loci underlying root rot resistance is encoded in the DNA sequence around marker $OI03_{514}$ that lies between AFLP derived SCARs, CGG5, and CTA13 on linkage group G [37]. However, the root rot resistance locus (Table 4) lay in a region not well represented among BAC libraries [84, 85], so that the gene isolation was delayed until the local genome sequence could be assembled. Transcript analysis showed that the fungus attempts to prevent gene transcription in the target roots [34, 55, 56]. Resistant cultivars prevent the poisoning of transcription by inducing stress and defense genes that produce fungicidal metabolites within 2 days of contact with the fungus. However, the induced genes do not appear to map to the loci that control the SDS resistance response [57]. Instead, genes of a higher hierarchical position in the interactome were found in this interval (unpublished). One of these genes is expected to underlie root resistance to SDS.

For the fungus, *F. virguliforme* causing SDS, no races are known so far in the US [86]. When lines from E × F have been used to look at variations in pathogenicity between strains, no convincing evidence for a host differential response was observed (unpublished). However, different Fusarium species that are capable of causing SDS are found in South America [86]. E × F was planted in Argentina since 2004, and it was shown that the SDS pathogen(s) invoked responses that mapped to different resistance loci [39]. Therefore, the fungus does have the potential to form races that vary in their pathogenicity. Hence, soybean breeders should be cautious in using the available resistance genes and should realize that stacking of all the twelve genes for full resistance would not be wise because it would select for mutants in the pathogen populations that could lead to the development of races.

In conclusion, a variety of approaches including QTL analysis, fine map development for some loci, and analysis of isolated genes have revealed that the alleles detected in E × F are variants of the same major genes found in weedy plant introductions (PIs) [5, 24, 41, 53]. Only few loci detected in the E × F population and in the other materials derived from this cross seem to represent other gene systems at a lower hierarchical position [57]. Identification of the lower tier of genetic control may require intercrosses among NILs or assays that relate to development, time, position, or cell type.

4. STRUCTURAL GENOMICS RESOURCES

Soybean (*Glycine max* L. Merr.) has a genome size of 1115 Mbp/1C [87]. The soybean genome is the product of a diploid ancestor ($n = 11$), that underwent aneuploid loss ($n = 10$), allo- and autopolyploidization events separated by millions of years ($n = 40$) with reversion to a lower ploidy after one of those two events ($n = 20$) [88]. Evidence that two genome duplications occurred, 40–50 MYA and 8–10 MYA, was supported by RFLP analysis suggesting 4–8 homoeologous loci for most probes [89] and discontinuous variation among paralagous EST sequences [90–92]. Even PCR-based markers that can amplify single loci from genomic DNA amplify multiple amplicons from BAC pool DNA (Figure 2). The duplicated regions have been segmented and reshuffled after the polyploidization events [16, 93–95].

Recently, a systematic measurement of DNA sequence divergence between homoeologous regions was made possible by comparing Forrest BAC end sequences with 7 million reads from the WGS sequences of Williams 82 [29, 93]. MegaBlast searches distinguished some regions, resolving up to 10% nonidentity between homoeologs over a 60 bp window (Figure 2). This implied that significant sequence divergence has occurred at about half the loci tested, as predicted from the gene-family size distribution observed in the physical map [57] (Figure 5). Conversely, highly conserved regions (>90% identity) exceeding about 150 kbp (the size of a large insert clone) have been inferred in certain regions [29]. Within these regions, 2 or 4 homoeologs can be distinguished by single nucleotide variants that correspond to the duplicated regions of a paleopolyploid genome or recently polyploid genome. These variants have been described as single nucleotide polymorphisms among homoeologs (SNHs) [93] though they are commonly called homoeologous sequence variants (HSVs) (see, e.g., [91]).

Overlain on the segmented regions found in 2 or 4 copies, the soybean genome is a composite of dispersed and contiguous euchromatic regions [88]. The short arms of four chromosomes are entirely heterochromatic, but in the remaining 16 chromosomes with potentially gene rich euchromatic arms, the heterochromatin is restricted to pericentromeric regions. Euchromatin represents 64% of the soybean genome, with a range of 40–85% on an

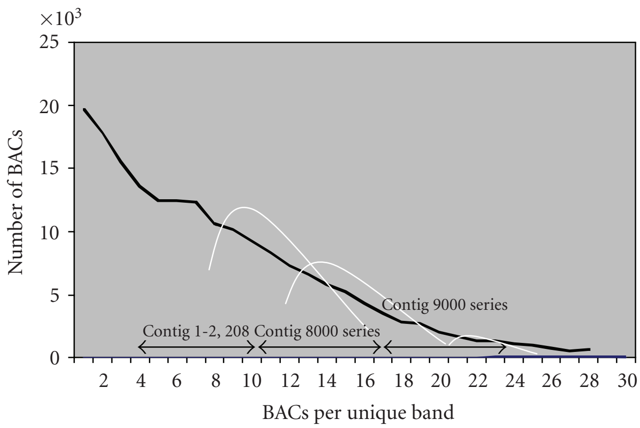


Figure 5: Quality estimate for the physical maps build 4 showing measurements of BAC clones per unique band. Three sets of distributions were inferred, representing the diverged DNA and the conserved DNA following the two genome duplications (shown as white lines). The 2208 single copy contigs (labeled 1–3500 after merges and splits) encompassed diverged DNA and are each inferred to contain clones from a single region. Contigs in the 8000 series are inferred to contain clones from two homoeologous regions. Contigs in the 9000 series are inferred to contain clones from four homoeologous regions. Clearly, some contigs in each set will be missplaced, hybrid contigs will occur, and ranges will overlap.

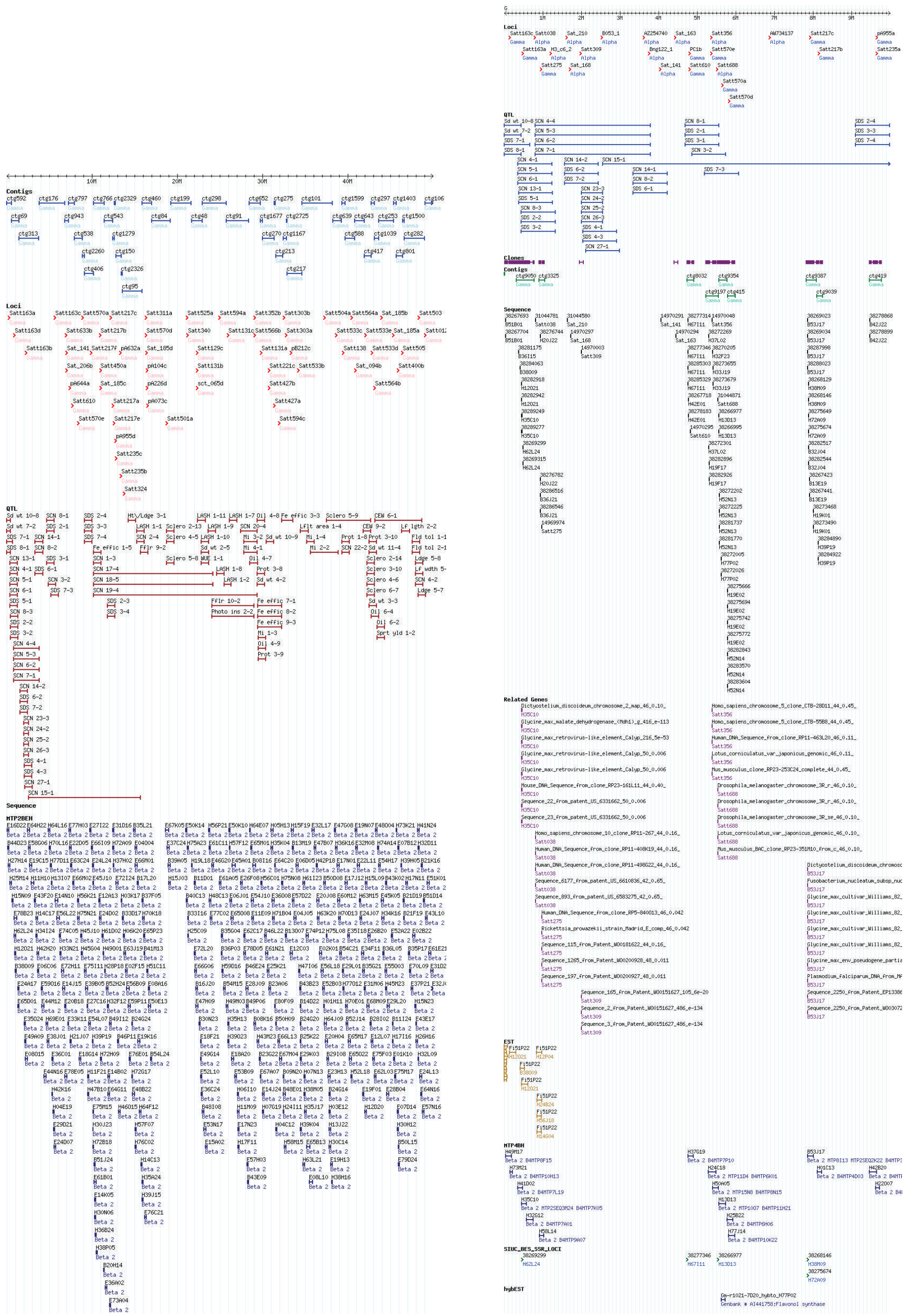
individual chromosome. Due to these features and the following other reasons, analysis of soybean genome has been a challenge: (i) large genome size, (ii) serial duplication of regions, (iii) small proportion of unique DNA, and (iv) highly conserved repeated DNA. One reasonable prediction would be that many of the duplicated regions would be silenced in heterochromatin. However, a comparison of the genetic map and physical map [93–95] has shown that duplicated segments are neither clustered nor restricted to heterochromatic arms. Further, the gene-rich islands are not separate from the duplicated regions. Therefore, new models to explain gene regulation that include duplicated conditions must be developed. Lessons learned from this exercise will help in the analysis of some legume and many dicotyledonous crop genomes, where genome duplication is believed to have often accompanied speciation. Breeders, who develop new cultivars through selection from the available variation within a cultivar, will also utilize this information and will develop new selection methods through an understanding of the effects and benefits of partial, segmented, genome duplication.

4.1. BAC libraries and physical maps

Construction of fingerprint-based physical maps in soybean relied on the availability of deep-coverage high-quality large insert genomic libraries, and a number of such public sector large insert libraries are available in four different plasmid vectors, providing >45-fold genome coverage. BAC libraries are available not only for Forrest and PI437654, but also for some *G. soja* PIs and the wild relatives of *G. max* [84, 85, 96, 97]. Among these libraries, there are three "Forrest" BAC libraries [84, 85], available in two different plasmid vectors with different *ori*s and different selectable markers (Table 5). Despite the availability of these rich BAC resources, there are still a few regions of the genome that are not well represented across the above set of BAC libraries. New libraries without involving restriction digestion may help solve this problem (unpublished).

A double-digest-based physical map for the soybean genome is now nearing completion. For this purpose, soybean BACs from five libraries belonging to three cultivars were fingerprinted and assembled [98] using a moderate information content fingerprint method (MICF) and FPC. The available BACs presently include 1182 Faribault BACs (~130 kbp, *Eco*RI inserts, 0.125x), 860 Williams 82 BACs (~130 kbp, *Hin*dIII inserts, 0.1x) and 78 001 Forrest BACs that were selected from the three libraries (125–157 kbp *Eco*RI, *Hin*dIII, and *Bam*HI inserts, 9x). Cultivar sequence variation did not appear to cause incorrect binning of BACs by FPC. However, the first release (build 3) [98] had many problems (Table 6), since many individual contigs appeared to contain noncontiguous genomic regions, and in some cases, different contigs contained the same region of the genome. Also, the available set of contigs encompassed a space that was 300 Mbp more than the size of the soybean genome. Clone contamination caused many of these problems, so that new methods to identify and eliminate contaminated clones were developed [99].

Subsequently, the publicly available soybean BAC fingerprint database was used to create build 4 [16] with the following specific aims: (i) to increase the number of genetic markers in the map, (ii) to reduce the frequency of clone contamination, (iii) to rebuild the physical map at high stringency, (iv) to examine clone density per contig, and (v) to examine the effectiveness of the generic genome browser in representing duplicated homoeologous regions (Table 6). Clones suspected of contamination were listed, fingerprints were examined, and contaminated clones removed from the FPC database. Many (7134 about 10%) well-to-well contaminated clones were removed from the fingerprint database. The edited database produced 2854 contigs and encompassed 1050 Mbp. In addition, homoeologous regions that might cause separate contigs to coalesce were detected in several ways. First, contigs with high clone density (23%) were inferred to represent two copy (240) or four copy (406) conserved genomic regions per haploid genome (Table 6). If the polyploid regions could all be split using HSVs (Figure 1) [29], there would be 1624 regions with two copies and 480 regions with four copies in the soybean genome. A second proof of this genome structure was that pairs of separate contigs that contained the same marker anchors (69%) were inferred to represent homoeologous but diverged genomic regions (Figure 6) [16]. A third proof came from EST hybridizations to BAC libraries where gene families with 1, 2, 4, and 8 members were more common than those with 3 or 5 members [57]. Finally, similarity search within the whole genome sequence at 90% similarity showed that the sequences that map to the contigs with duplicated regions do



(a) (b)

Figure 6: *Description of chromosome 18 resources at SoyGD (a).* The current GMOD representation of 50 Mbp of the 51.5 Mbp chromosome 18 (linkage group G) in SoyGD (a). shows the build 3 version of the chromosome (cursor), anchored contigs (top row, blue), DNA markers (second row of features, red), QTL in the region (third row, burgundy), MTP2 clones (B, H, and E fourth row, dark blue). Not shown here were BAC clones, ESTs, BAC end sequences, and gene models (b) shows the build 4 representation of 10 Mbp of the 51.5 Mbp chromosome 18 in SoyGD. Shown are the chromosome (cursor), DNA markers (top row of features, red); QTL in the region (second row, blue); coalesced clones (purple) comprising the anchored contigs (third row, green); BAC end sequences (fourth row black); BESs encoding gene fragments (fifth row, puce); EST hybridizations to MTP2BH (sixth row gold); MTP4BH clones (seventh row, dark blue); BESs derived SSR (eighth row, green); EST hybridizations inferred on build 4 from clones also in MTP2BH (ninth row, blue); WGS trace file matches from MegaBlast (tenth and last row, light blue). It is recommended for readers to visit updated site http://bioinformatics.siu.edu/ to see a full detailed color version and a build 5 view. The gaps between contigs will be filled in build 5 by contig merges suggested by BESs-SSRs and contig end overlap data.

Table 5: Progress in the soybean physical map builds 2 to 5.

	Automated Build 2 Sept. 2001	Manual Build 3 Oct 2002	Manual Build 4 Oct 2003 Total	Judged by BACs/unique band to be (pploid) [unique]	Manual Build 5 Jan 2008
BAC clones in FPC database	81,024	83,026	78,001		78,001
BACs used in contig assembly	75,568	78,001	72,942		72,837
Number of singletons	5,884	4954	27,1812		17,942
Marker anchored singletons	0	0	120		63
Clone in contigs (fold genome)	69,684	73,069	45,135		58,765
Fold genome in contigs	8.7	9.1	5.6		6.2
Number of contigs	5,488	2,907	2,854	(646)[2208]	521
Anchoring Markers	0	385	404	(280)[124]	1,523
Anchored Contigs	0	781	742	(181)[223]	455
Contlgs contain: >25 clone	220	921	477	(268)[209]	335
10–25 clones	3,038	920	1,458	(433)[1025]	110
3–9 clones	1,845	850	820	(0)[820]	43
2 clones	385	216	99	(0)[99]	33
Unique bands in the contigs	396,843	345,457	#258,240	(64,560)	257,356
Length of the contigs (Mb)	**1,667****	**1,451****	**1.037**	**(0.258) [0.769]**	**1.034**

*Based on 4.00 kbp per unique hand. # Based on 4.05 kbp per unique band, for 2,854 contigs containing ~68 unique bands in 15 clones, 264 duplicated region contigs containing ~68 unique bands in 30 clones I5,840 unique bands and 406 highly repeated region contigs containing ~68 unique bands is 60 clones, 48,720 unique bands.

Table 6: *Summary of sequence coverage of the three minimum tile paths (MTPs) used for BAC end sequencing made from three BAC libraries*. To calculate the percentage of the soybean genome covered by the clones (clone coverage) in our *Eco*RI-(MTP4E) and *Bam*HI or HindIII insert libraries (MTP2BH and MTP4BH), the genome size of soybean was assumed to be 1130 Mb. The BAC libraries were each constructed from DNA derived from twenty five seedlings of an inbred cultivar Forrest.

	MTP4E	MTP4BH		MTP2 BH	Totals
Vector	pBeloBAC11	pCLD04541			na
Insertion site	*Eco* RI	*Bam*HI or *Hind*III			na
Duplicates/region	1	1	2–4	1–4	na
Number of clones	3840	4608	576	8064	17 088
Mean insert size (kbp)	175 ± 7	173 ± 7	173 ± 7	140 ± 5	685
Clone coverage	0.7	0.8	0.2	1.4	3.1
BESs good reads	3 324	6772	924	13 473	25 123
BESs coverage (Mbp)	2.9	5.0	0.7	9.9	18.5

have homoeologs in the sequence, whereas sequences from single copy regions do not (Figure 2) [29, 93].

To deal with duplicated regions, SoyGD was adapted to distinguish homoeologous regions by showing each contig at all potential anchor points, spread laterally, rather than as overlapping [16]. Therefore, it should be realized that the genes in such regions have duplicates in other regions of the genome (Figure 6). This information will prove useful in future for gene isolation by positional cloning following a reverse genetics approach, where aneupleurotic pathways regularly cause wide-spread failures [100–102] due to inability to predict phenotypes reliably.

In build 5, DNA sequence scaffolds (unpublished) have been used to cluster groups of neighboring contigs. This, however, does not solve the problems faced due to genome duplication. In many cases, (60–80%), homoeologous variants may help separation of coalesced regions [29], but this would require BESs for every fingerprinted BAC clone. In a minority of regions (20–40%), sequences longer than BESs may be needed to correctly separate BAC clones into contigs.

4.2. Minimum tile paths

The creation of minimally redundant tile paths (MTP) from contiguous sets of overlapping clones (contigs) in physical maps is a critical step for structural and functional genomics [95]. The first minimum tiling path (MTP) developed (from builds 2 and 3) contained 2 fold redundancy of the haploid genome (2,100 Mbp). MTP2 was 14 208 clones (mean insert size 140 kbp) that were picked from the 5597 contigs of build

2. MTP2 was constructed from three BAC libraries (BamHI (B), HindIII (H) and EcoRI (E) inserts), encompassing the contigs of build 3 that were derived from build 2 by a series of contig merges, but does not distinguish regions by degree of duplication, so that many regions are redundant. The MTP2 is used in two parts, MTP2BH and MTP2E (Table 6) because they are largely redundant and overlap each other. Also, the vectors differ in the antibiotic resistance conferred. Consequently, only the MTP2BH was used for development of EST map [57].

The third and fourth MTPs, called MTP4BH and MTP4E (Table 6), were each based on build 4 [95]. Each was selected as a single path through each of the 2854 contigs. MTP4BH had 4608 clones with a mean size 173 kbp in the large (27.6 kbp) T-DNA vector pCLD04541, which is suitable for plant transformation and functional genomics. Plates 1–8 contained clones from the contigs belonging to the single copy regions of the genome. Plates 9 and 10 were picked from the duplicated and quadruplicated regions without redundancy, so that an individual clone represented either 2 or 4 regions per haploid genome. Plates 11 and 12 contained the marker anchored clones also used in MTP2BH. Plate 13 of MTP4BH was developed from just 6 contigs from regions with four copies by redundant picking. This set of clones should resolve into 48 regions, if methods to separate them can be developed as the genome sequencing is completed [93]. This set of 13 plates was used for HICF fingerprinting by the same methods that were used for Williams 82 [11] and PI437654 BACS [79, 96]. The BACs used for HICF will form a bridge to other physical maps and a resource to test the ability of HICF to correctly separate duplicated regions, particularly in the contigs in plate 13.

MTP4E was designed to be 4608 BAC clones with large inserts (mean 175 kbp) in the small (7.5 kbp) pECBAC1 vector [57, 85]. However, only 3840 clones were picked to date. Sequencing efficiency was low on this MTP and reracking will be needed [103]. The vector is suitable for DNA sequencing and these clones will be used for sequencing across gaps in the WGS sequence.

MTP4BH and MTP4E clones each encompassed about 800 Mbp before duplicate regions were considered. The single copy regions represented 700 Mbp [57]. In addition there were 50 Mbp from the duplicate and 50 Mbp from the quadruplicate regions in the MTP. Because those regions were duplicate and quadruplicate they encompass another 300 Mbp in total. MTP2BH, MTP4E, and MTP4BH were each used for BAC-end sequencing and microsatellite integration into the physical map [27, 39]. MTP2BH was used for EST integration to the physical map [16, 57]. MTP4BH was used for high information content fingerprinting for integration with the Williams 82 physical map [11, 104]. In conclusion, it appears like each MTP and the derived BESs will be useful to deconvolute and finish the whole genome shotgun sequence of soybean while the whole genome sequence will help complete the physical map. A complete MTP5BH would be a useful tool for functional genomics because clones from these libraries were constructed in a T-DNA vector and are ready for plant transformation. About four thousand transgenic lines made from BACs would be enough to transfer every soybean sequence to another plant.

4.3. BAC end sequences (BESs)

BAC end sequences (BESs) anchored to a robust physical map constitute an important tool for genome analysis, and have been developed from BACs belonging to three available MTPs including MTP2BH, MTP4BH, and MTPE4 [95, 103]. Therefore, three sets of BESs were available, of which the first set consisted of 13 474 good BESs derived from 8064 clones of MTP2BH (Table 5). Enquiries to GenBank nr and pat databases identified 7260 potentially geneic homologs, and an analysis of the locations of inferred genes suggested presence of gene-rich islands on each chromosome [37]. In addition, 42 BESs showed homology (extending over a length of 80–341 bp at e^{-30} to e^{-300}) with DNA markers (10 RFLPs, 20 microsatellites) that were already genetically mapped [95]. This amounts to homology with about 2% of the markers, whose sequences are available in GenBank. Available BESs also carried as many as 1053 new SSR markers [27, 37] that are described further in the next section.

The second set of BESs consisted of 7700 good BESs reads from clones of MTP4BH (Table 5) of which 4147 had homologs in the GenBank nr and pat databases [57]. The clones in plates 11 and 12 were resequenced and so have 2 records for each BAC end in GenBank. Resequenced clones help determine the sequence error rate and greatly facilitate SNP detection [18, 19]. Twenty additional genetic anchors were detected in this second set of BESs (6 RFLPs, 14 microsatellites), which represented about 1% of the soybean markers with sequences in GenBank. This second set of BESs carried 625 SSR markers [27, 37] that are described further in the next section. The third set of BESs from MTP4E have recently been released and are only partly analyzed (Table 6).

The above builds of physical map representing recently duplicated regions of the genome can be further improved with existing databases and tools. In particular, this can be achieved by increasing the number of reliable genetic anchors derived from BESs [27, 37] and separating BACs from homoeologous regions with diagnostic SNPs (Figure 2) before contigs were formed [93].

4.4. Genetic map and SSR markers derived from BESs

The molecular genetic map for soybean genome can be improved further through several approaches including (i) addition of BESs markers on the available genetic map [27, 37], (ii) bioinformatics analysis of contig data [16] and (iii) through the use of novel approaches to error detection [99]. The composite genetic map of soybean at SoyGD (in 2007) contained 3073 DNA markers [16, 27], which included 1019 class I SSRs, each with >10 di- or trinucleotide repeat motifs (BARC-SSR markers; Song et al., 2004), and a few class II SSRs with <10 di- to pentanucleotide repeats that were mostly SIUC-SSR markers. Forrest BESs helped in increasing the number of class I and II SSR markers for the soybean genome, and allowed integration of BAC clones into the soybean physical map.

SSRs were mined separately from the two sets of BESs described above. As mentioned above, the first set of 10 Mbp of BAC end sequences (BESs) derived from 13 474 reads of 7050 clones constituting MTP2BH, had 1053 SSRs (333 class I + 720 from class II), and the second set of 5.7 Mbp BESs derived from 7700 reads from 5184 clones constituting MTP4BH, had 620 SSRs (150 class I + 480 class II). Potential markers are shown on the MTP_SSR track at SoyGD (Figure 6). About 530 primer pairs were designed for both the sets of SSRs. These primers were 20–24 mers long with a T_m of 55 + 1°C, and provided amplicons that were 100–500 bp long. As many as 123 of these primers belonging to duplicated regions gave multiple amplified products, and therefore should be avoided.

Different possible motifs were not randomly distributed among the above SSRs, with AT rich motifs being more frequent [27]. Compound SSRs having tetranucleotide repeats clustered with di- and trinucleotide motifs were also found. About 75% of class I and 60% of class II SSR markers were polymorphic among the parents of four recombinant inbred line (RIL) populations. Most of the BESs-SSRs were located on the soybean genetic map in regions with few BARC-SSR markers [27, 39]. Therefore, BESs-SSRs represent a useful tool for the improvement of the genetic map of soybean.

4.5. SNP markers derived from BESs to WGS

The soybean genome has been shown to be composed of ~8000 short interspersed regions of one, two, or four copies per haploid genome, as shown by RFLP analysis, SSR anchors to BACs and by BAC fingerprints [16]. Recently, the genome has been sequenced by WGS sequencing of 4 kbp inserts in pUC18 [105]. When the extent and homogeneity of duplications within contigs was examined using BAC end sequences (BESs) derived from minimum tile paths (MTP2BH and MTP4BH; Figure 2) [29], a strong correlation was found between the fold of duplication inferred from fingerprinting and that inferred from WGS matches. Duplicated regions were identified by BAC fingerprint contig analysis using a criterion of less than 10% mismatch across a trace with a window size of 60 bp. Previously, simulations had predicted that fingerprints of clones from different regions would coalesce, if sequence variation was less than 2%. Hopefully, the HSVs among contigs from duplicated regions can be used to separate clone sets from different regions. Ironically, improvements for contig building methods will result from the whole genome sequence! However, many duplicated regions with less than 1% sequence divergence were found [29, 93]. The implication for bioinformatics and functional annotation of the soybean genome (and other paleopolyploid or polyploid genomes) is that reverse genetics with many genes will be nearly impossible without tools to simultaneously repress or mutate several gene family members.

5. FUNCTIONAL GENOMICS TOOLS

Unequivocal identification and map-based cloning of genes underlying quantitative traits have been a challenge for soybean genomics research. Gene redundancy, gene action, and low transformation efficiencies seriously hampered positional cloning [16]. Therefore, a variety of approaches need to be used for soybean functional genomics research. Two major areas of soybean genomics research include (i) annotation of genomic sequences (genes with unknown functions) and (ii) analysis of genome sequences of "Forrest" for synteny with the genomes of other dicotyledonous genera and with those of other soybean cultivars.

5.1. Annotation of genome sequences

The three methods that proved useful for annotation of the genome sequences of Forrest and related germplasm include (i) mutant complementation using transformation, (ii) gene silencing through RNAi, and (iii) targeted mutations. Each will be briefly discussed.

(i) *Mutant complementation using transformation*. A popular approach for the study of gene function is mutant complementation, which involves transformation of mutants with the wild alleles. Therefore, development of transformation protocols is an essential component of functional genomics research. In soybean, *A. tumefaciens* and *A. rhizogenes*-mediated transformation of cultured cells with Forrest BAC clones has been successfully achieved using previously described protocols involving the T-DNA vector pCLD04541 [84]. In this protocol, *npt II* gene is used as a plant selectable marker, and kanamycin as used as a selective agent [106–109]. Screenable markers are available in some BAC clones (Table 7). Whole BAC transformation is important because fine maps locating loci at genetic distance of 0.25 cM that is equivalent to 50–150 kbp were earlier prepared using RILs and NILs. The clones selected for transformation are listed in Table 7, and should provide for complementation of easily scoreable phenotypes in mutants. For instance, dominant mutant phenotypes of traits like pubescence, color, and disease resistances should be evident in the very first products of transformation. BAC transformation with sets of overlapping clones will be the best approach in situations where an individual locus represents a cluster of genes [37, 38].

(ii) *Gene silencing using RNAi*. The composite plant system for RNAi has been tested in NILs derived from Forrest, and has been validated by Dr. C. G. Taylor at the Danforth Center (St. Louis, Mo, USA) [110] through expression of gene-specific dsRNA constructs. Using this system, shoots from stable transgenic soybean plants showing constitutive expression of *uid*A (GUS) were transformed with dsRNA constructs (Figure 7) that were designed using a modified pKannibal vector [111], with the 35S promoter replaced by the figwort mosaic virus (FMV) promoter. The 600 bp homologous sequences of the GUS or green fluorescent protein (GFP) gene were introduced in an antisense and sense orientation separated by the pKannibal intron (spacer) sequence. These constructs were designed to produce transcripts with a stem loop secondary structure that would be recognized by the plant cell machinery and activate RNAi. The dsRNA constructs placed in a binary vector, introduced into *A. rhizogenes*, were used for composite plant production

Table 7: Some of the BACs, mutant and nonmutant soybean lines to be transformed for complementation.

BIBAC clone names	Phenotypes	Insert size kbp	Dominant?
Gm-SIU1-B100B10	*Rhg4* bigenic resistance to SCN[(a)]	140	Yes
Gm-SIU1-B73P06	*rhg1* bigenic resistance to SCN and *Rfs2* for SDS[(a)]	79	Co-
Gm-SIU2-H050N07	*Rpg1-b* resistance to bacterial pustule[(b)]	110	Yes
Gm-SIU1-B54E07	*T* tawny pubescence; flavonoid-3-monoxygenase[(c)]	82	Yes
Gm-SIU2-H04P03	*W1* White flower and black hila color[(d)]	153	No
Gm-SIU2-H82CO8	*Rfs1* root resistance to SDS	130	Yes
Gm-SIU1-TBD	*Rps4* resistance to *Phytophthora* root rot	120	Yes

[(a)]]*Rhg4* and *rhg1* each encodes transmembrane receptor-like kinase. Resistant and susceptible alleles differ by 3–6 amino acid changes and 23 base changes. There are mutant lines derived from Forrest.
[(b)]*Rpg1-b* encodes a nucleotide binding leucine rich repeat protein.
[(c)]*T* encodes flanonoid-3 monoxygenase (EC1.13.14.21). The recessive genes differ from the dominant by deletion of a single C nucleotide. There are mutant lines.
[(d)]*W* encodes an unknown enzyme, probably a glycosidase.

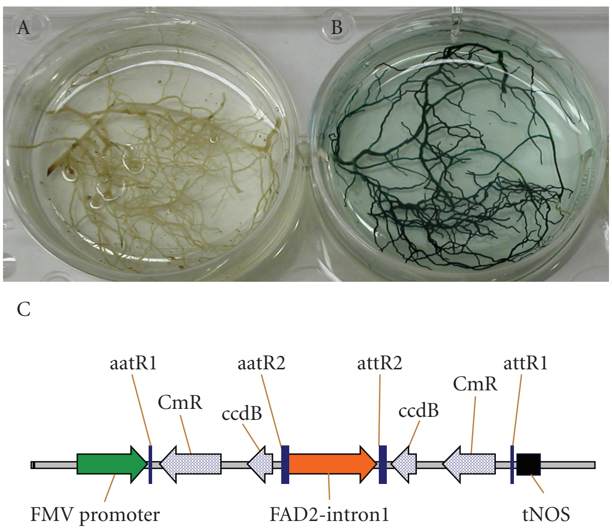


Figure 7: Evidence for RNAi silencing of GUS gene in 35S::GUS soybean plants. Panel A. GUS expression in composite plant roots expressing and RNAi from the gene encoding GFP. Panel B. GUS expression in composite plant roots expressing RNAi from the gene encoding GUS. Panel C. The transformation cassette used (thanks to Dr. C. G. Taylor, Danforth Center, unpublished data).

[112]. GUS-specific RNAi construct silenced, while non-GUS RNAi (GFP) construct failed to silence GUS expression in hairy roots produced on shoots of transgenic soybean plants. These results show that the hairy roots can be used to produce dsRNAs. Further, the RNAi machinery in soybean hairy roots is fully functional in a sequence-specific manner. Thus, RNAi technology will allow the rapid analysis of sets of candidate genes for alleles underlying variation [38].

(iii) *Study of gene function through TILLING.* Two soybean mutagenized M2 libraries are already available for TILLING [113], from which ~3000 of the 6000 available M2 lines were phenotyped visually. A soybean mutant database has been developed to track and sort these mutants (http://www.soybeantilling.org/). While developing a database that would allow search for "TILLED" genes a search engine was developed, so that the database can be searched for both phenotype and gene. The mutations occurred at a rate of ~1 mutation/170 kbp, so that a screening of 6150 M2 families may provide a series of up to 40 to 60 alleles within each 1.5 kbp fragment of a target gene. This approach led to the identification of a putative mutant for a soybean leucine rich repeat receptor like kinase gene *Gm-Clavata1A* (AF197946; Figure 8). In future, TILLING and crosses among TILLED mutants [100–102] will allow the testing of candidate genes and will provide new genetic variation that may lead to germplasm enhancement.

5.2. Analyses of conserved synteny

Forrest genome sequences have also been used for a study of their synteny with genomes of other dicotyledonous genera/species and also with the genomes of other soybean cultivars. For this purpose, cross-species transferable genetic markers are available in the data-based legumeDB1 [114], and can be used to compare the linear order of markers/genes, which are either species specific or conserved across genera [115–124]. For instance, genes for resistance to pathogens will often appear as new genes or gene clusters inserted in regions, which otherwise exhibit conserved synteny across genera [35, 115, 122]. Synteny extends beyond genes into repeat DNA, as exemplified by the distributions of 15 bp sequences that provide sequence-specific genome fingerprints [94]. Interestingly the fingerprints do not show the same patterns of relatedness between species found in gene sequence. Therefore, genome fingerprinting will help identify good candidates for cross species markers in repeat DNA such as microsatellite markers.

Conserved synteny has also been observed among the genomes involved in the constitution of the allo- and autotetraploids hypothesized for soybean. It has been shown that about 25–30% of the genome has extensive conservation of gene order in otherwise shuffled blocks of 150–300 kbp [16]. Consequently, blocks of 3–10 genes are repeated at 2 or 4 locations per haploid genome [38, 79]. There are also genomic regions, where synteny among genomes of different cultivars has been shown to break down. Several interesting

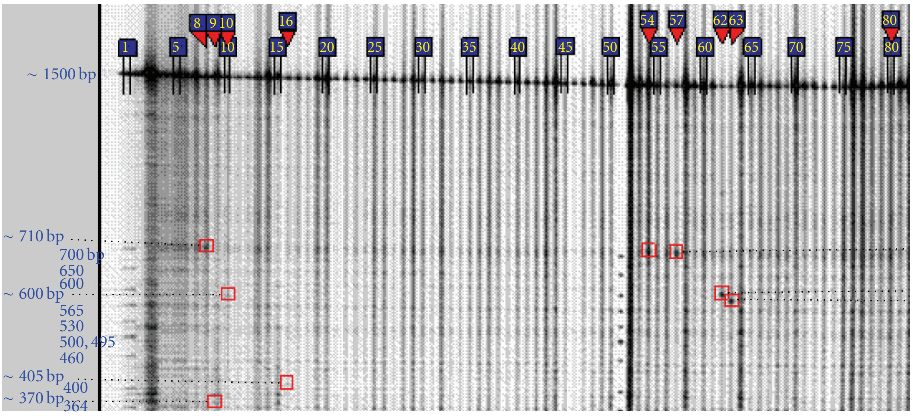


Figure 8: Soybean Tilling gel image of *Gm_clavata1* pool ps33 screening, representing 768 individuals, 8 individuals per pool (LI-COR 700 channel mutations are marked in red boxes; blue boxes represent lane numbers) from http://www.soybeantilling.org/ (thanks to Dr. K. Meksem and Dr. B. Liu SIUC, unpublished data).

features including the following have been observed in these nonsyntenic regions: (i) in some cases, a loss of conserved synteny between cultivars is associated with a gene introgressed from a Plant Introduction [38]. (ii) In another case, a moderately repeated sequence common in one cultivar is absent in another cultivar [29]. (iii) In still another case, a sequence inserted in one cultivar appears to alter the expression of a neighboring gene (unpublished). It is thus apparent that genome analysis involving study of an association of these nonsyntenic sequence tracts in otherwise syntenic regions, with phenotypes will be an active area of research, when genome sequences from a number of soybean cultivars are available.

6. CONCLUSIONS

The soybean genomics resources developed through the use of cultivar Forrest have been used and will continue to be used in future leading to significant advances in soybean genomics knowledge base. The soybean genome shows evidence of a paleopolyploid origin with regions, encompassing gene-rich islands that were highly conserved following duplication [16]. In fact, it was estimated that 25–30% of the genome was highly conserved after both duplications. Implications of this feature are profound. First, a map of homoeology and an associated map of duplicated regions had to be developed. Second, an estimate of sequence conservation among the duplicated regions was necessary. Third, the implications for functional genomics were considered. Given that all soybean genes have been duplicated twice during recent evolution, and that most plant genomes encode functionally redundant pathways, it is not surprising that TILLING, RNAi-mediated silencing and overexpression of several genes often did not lead to phenotypic changes [101, 102, 110, 113]. In future, the E × F population will continue to be used for (i) an analysis of functions of a number of gene families, (ii) patenting of inventions based on useful genes [6, 77, 124–126], (iii) manipulation of soybean seed composition including protein, oil [19] and bioactive factors [127–129], and (iv) an analysis of the protein interactome [130]. In summary, the newly released E × F population and the other associated genomic resources developed through the use of cultivar “Forrest” will provide tremendous opportunities for further research in the field of genomics research.

ACKNOWLEDGMENTS

This research summarized in this review was funded in part by grants from the NSF 9872635, ISA (95-122-04; 98-122-02; 02-127-03), and USB 2228–6228. The physical map was based upon work supported by the National Science Foundation under Grant no. 9872635. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. The continued support of SIUC, College of Agriculture, and Office of the Vice Chancellor for Research to the Genomics Facility is appreciated. The author thanks Dr. Q. Tao and Dr. H. B. Zhang for assistance with fingerprinting; Dr. C. Town and Dr. C. Foo at TIGR for the BESs; Dr. K. Meksem for the Tilling figure; Dr. C. G. Taylor for an RNAi figure; Dr. Z. Zhang for transformation of Forrest; and the SIUC team for their rigor in addressing a novel problem for genomics. All team members are thanked for their continued collaborations. Thanks are also due to P. K. Gupta, the guest editor, for critical readings, which led to significant improvement of the manuscript.

REFERENCES

[1] E. E. Hartwig and J. M. Epps, “Registration of ‘Forrest’ soybeans,” *Crop Science*, vol. 13, no. 2, p. 287, 1973.

[2] G. J. Carbone, L. O. Mearns, T. Mavromatis, E. J. Sadler, and D. Stooksbury, “Evaluating CROPGRO-soybean performance for use in climate impact studies,” *Agronomy Journal*, vol. 95, no. 3, pp. 537–544, 2003.

[3] O. Myers Jr. and S. C. Anand, “Inheritance of resistance and genetic relationships among soybean plant introductions to

races of soybean cyst nematode," *Euphytica*, vol. 55, no. 3, pp. 197–201, 1991.

[4] R. R. Prabhu, V. N. Njiti, B. Bell-Johnson, et al., "Selecting soybean cultivars for dual resistance to soybean cyst nematode and sudden death syndrome using two DNA markers," *Crop Science*, vol. 39, no. 4, pp. 982–987, 1999.

[5] S. Kazi, J. L. Shultz, R. Bashir, et al., "Identification of loci underlying seed yield and resistance to soybean cyst nematode race 2 in 'Hartwig'," *Theoretical and Applied Genetics*. In press.

[6] D. A. Lightfoot, K. Meksem, and P. T. Gibson, "Soybean sudden death syndrome resistant soybeans, soybean cyst nematode resistant soybeans and methods of breeding and identifying resistant plants," DNA markers, US Patent no. 6300541, October 2001.

[7] N. Hnetkovsky, S. J. C. Chang, T. W. Doubler, P. T. Gibson, and D. A. Lightfoot, "Genetic mapping of loci underlying field resistance to soybean sudden death syndrome (SDS)," *Crop Science*, vol. 36, no. 2, pp. 393–400, 1996.

[8] S. J. C. Chang, T. W. Doubler, V. Kilo, et al., "Two additional loci underlying durable field resistance to soybean sudden death syndrome (SDS)," *Crop Science*, vol. 36, no. 6, pp. 1684–1688, 1996.

[9] S. J. C. Chang, T. W. Doubler, V. Y. Kilo, et al., "Association of loci underlying field resistance to soybean sudden death syndrome (SDS) and cyst nematode (SCN) race 3," *Crop Science*, vol. 37, no. 3, pp. 965–971, 1997.

[10] D. A. Lightfoot, V. N. Njiti, P. T. Gibson, M. A. Kassem, J. M. Iqbal, and K. Meksem, "Registration of Essex x Forrest recombinant inbred line mapping population," *Crop Science*, vol. 45, no. 4, pp. 1678–1681, 2005.

[11] S. A. Jackson, D. Rokhsar, G. Stacey, R. C. Shoemaker, J. Schmutz, and J. Grimwood, "Toward a reference sequence of the soybean genome: a multiagency effort," *Crop Science*, vol. 46, supplement 1, pp. S55–S61, 2006.

[12] J. H. Orf, K. Chase, T. Jarvik, et al., "Genetics of soybean agronomic traits—I: comparison of three related recombinant inbred populations," *Crop Science*, vol. 39, no. 6, pp. 1642–1651, 1999.

[13] N. Yamanaka, S. Ninomiya, M. Hoshi, et al., "An informative linkage map of soybean reveals QTLs for flowering time, leaflet morphology and regions of segregation distortion," *DNA Research*, vol. 8, no. 2, pp. 61–72, 2001.

[14] Y. S. Dong, L. M. Zhao, B. Liu, Z. W. Wang, Z. Q. Jin, and H. Sun, "The genetic diversity of cultivated soybean grown in China," *Theoretical and Applied Genetics*, vol. 108, no. 5, pp. 931–936, 2004.

[15] J. Yuan, V. N. Njiti, K. Meksem, et al., "Quantitative trait loci in two soybean recombinant inbred line populations segregating for yield and disease resistance," *Crop Science*, vol. 42, no. 1, pp. 271–277, 2002.

[16] J. L. Shultz, D. Kurunam, K. Shopinski, et al., "The soybean genome database (SoyGD): a browser for display of duplicated, polyploid, regions and sequence tagged sites on the integrated physical and genetic maps of *Glycine max* ," *Nucleic Acids Research*, vol. 34, database issue, pp. D758–D765, 2006.

[17] P. M. Gresshoff, J. Stiller, T. L. D. McGuire, et al., "Integrating functional genomics to define the plants function in symbiotic nodulation," in *Nitrogen Fixation: Global Perspectives*, T. Finan, M. O'Brien, D. Layzell, K. Vessey, and W. Newton, Eds., pp. 95–98, CAB International, New York, NY, USA, 2001.

[18] C. R. Yesudas, J. L. Shultz, and D. A. Lightfoot, "Identification of loci underlying resistance to Japanese beetle herbivory, in soybean," *Theoretical and Applied Genetics*. In press.

[19] C. R. Yesudas, R. Bashir, J. Shultz, S. Kazi, and D. A. Lightfoot, "QTL for seed isoflavones, protein, oil and Japanese beetle (Popillia japonica, Newman) resistance in soybean [*Glycine max* (L.) Merr.]," in *Plant & Animal Genome XV Conference*, p. 411, San Diego, Calif, USA, January 2007.

[20] K. Meksem, P. Pantazopoulos, V. N. Njiti, L. D. Hyten, P. R. Arelli, and D. A. Lightfoot, "'Forrest' resistance to the soybean cyst nematode is bigenic: saturation mapping of the *Rhg1* and *Rhg4* loci," *Theoretical and Applied Genetics*, vol. 103, no. 5, pp. 710–717, 2001.

[21] V. N. Njiti, K. Meksem, D. A. Lightfoot, W. J. Banz, and T. A. Winters, "Molecular markers of phytoestrogen content in soybeans," *Journal of Medicine and Food*, vol. 2, pp. 165–167, 2000.

[22] K. Meksem, V. N. Njiti, W. J. Banz, et al., "Molecular markers of phytoestrogen content in soybeans," *Journal of Biomedicine and Biotechnology*, vol. 1, no. 1, pp. 38–44, 2001.

[23] M. A. Kassem, K. Meksem, M. J. Iqbal, et al., "Definition of soybean genomic regions that control seed phytoestrogen amounts," *Journal of Biomedicine and Biotechnology*, vol. 2004, no. 1, pp. 52–60, 2004.

[24] M. A. Kassem, J. L. Shultz, K. Meksem, et al., "An updated 'Essex' by 'Forrest' inkage map and first composite interval map of QTL underlying six soybean traits," *Theoretical and Applied Genetics*, vol. 113, no. 6, pp. 1015–1026, 2006.

[25] Z. Gizlice, T. E. Carter Jr., T. M. Gerig, and J. W. Burton, "Genetic diversity patterns in North American public soybean cultivars based on coefficient of parentage," *Crop Science*, vol. 36, no. 3, pp. 753–765, 1996.

[26] P. Keim, W. Beavis, J. Schupp, and R. Freestone, "Evaluation of soybean RFLP marker diversity in adapted germplasm," *Theoretical and Applied Genetics*, vol. 85, no. 2-3, pp. 205–212, 1992.

[27] J. L. Shultz, J. S. Kazi, R. Bashir, A. J. Afzal, and D. A. Lightfoot, "Development of microsatellite markers from the soybean sequence ready physical map," *Theoretical and Applied Genetics*, vol. 114, no. 6, pp. 1081–1090, 2007.

[28] Y. L. Zhu, Q. J. Song, D. L. Hyten, et al., "Single-nucleotide polymorphisms in soybean," *Genetics*, vol. 163, no. 3, pp. 1123–1134, 2003.

[29] D. A. Lightfoot, J. Shultz, and N. Saini, "Reannotation of the physical map of *Glycine max* for ploidy by BAC end sequence driven whole genome shotgun read assembly," in *Proceedings of the International Conference on Bioinformatics and Computational Biology (BIOCOMP '07)*, pp. 65–72, Las Vegas, Nev, USA, June 2007.

[30] M. J. Iqbal, K. Meksem, V. N. Njiti, M. A. Kassem, and D. A. Lightfoot, "Microsatellite markers identify three additional quantitative trait loci for resistance to soybean sudden-death syndrome (SDS) in Essex x Forrest RILs," *Theoretical and Applied Genetics*, vol. 102, no. 2-3, pp. 187–192, 2001.

[31] K. Meksem, E. Ruben, D. L. Hyten, M. E. Schmidt, and D. A. Lightfoot, "High-throughput genotyping for a polymorphism linked to soybean cyst nematode resistance gene *Rhg4* by using Taqman^{TM} probes," *Molecular Breeding*, vol. 7, no. 1, pp. 63–71, 2001.

[32] V. N. Njiti, J. E. Johnsona, T. A. Torto, L. E. Gray, and D. A. Lightfoot, "Inoculum rate influences selection for field resistance to soybean sudden death syndrome in the greenhouse," *Crop Science*, vol. 41, no. 6, pp. 1726–1731, 2001.

[33] V. N. Njiti, K. Meksem, M. J. Iqbal, et al., "Common loci underlie field resistance to soybean sudden death syndrome in Forrest, Pyramid, Essex, and Douglas," *Theoretical and Applied Genetics*, vol. 104, no. 2-3, pp. 294–300, 2002.

[34] M. J. Iqbal, S. Yaegashi, V. N. Njiti, R. Ahsan, K. L. Cryder, and D. A. Lightfoot, "Resistance locus pyramids alter transcript abundance in soybean roots inoculated with *Fusarium solani* f. sp. *glycines*," *Molecular Genetics and Genomics*, vol. 268, no. 3, pp. 407–417, 2002.

[35] T. Ashfield, A. Bocian, D. Held, et al., "Genetic and physical localization of the soybean *Rpg1-b* disease resistance gene reveals a complex locus containing several tightly linked families of NBS-LRR genes," *Molecular Plant-Microbe Interactions*, vol. 16, no. 9, pp. 817–826, 2003.

[36] M. A. Kassem, K. Meksem, C. H. Kang, et al., "Loci underlying resistance to manganese toxicity mapped in a soybean recombinant inbred line population of 'Essex' x 'Forrest'," *Plant and Soil*, vol. 260, no. 1-2, pp. 197–204, 2004.

[37] K. Triwitayakorn, V. N. Njiti, M. J. Iqbal, S. Yaegashi, C. Town, and D. A. Lightfoot, "Genomic analysis of a region encompassing *QRfs1* and *QRfs2*: genes that underlie soybean resistance to sudden death syndrome," *Genome*, vol. 48, no. 1, pp. 125–138, 2005.

[38] E. Ruben, J. Jamai, A. J. Afzal, et al., "Genomic analysis of the *Rhg1* locus: candidate genes that underlie soybean resistance to the cyst nematode," *Molecular Genetics and Genomics*, vol. 276, no. 6, pp. 503–516, 2006.

[39] R. Bashir, "Minimum tile derived microsatellite markers improve the physical map of the soybean genome and the Essex by Forrest genetic map," M.S. thesis, Southern Illinois University at Carbondale, Carbondale, III, USA, 2007.

[40] K. Meksem, T. W. Doubler, K. Chancharoenchai, et al., "Clustering among loci underlying soybean resistance to *Fusarium solani*, SDS and SCN in near-isogenic lines," *Theoretical and Applied Genetics*, vol. 99, no. 7-8, pp. 1131–1142, 1999.

[41] V. N. Njiti and D. A. Lightfoot, "Genetic analysis infers *Dt* loci underlie resistance to *Fusarium solani* f. sp. *glycines* in indeterminate soybeans," *Canadian Journal of Plant Science*, vol. 86, no. 1, pp. 83–90, 2006.

[42] A. Sharma, "The identification of QTL underlying resistance to aluminum toxicity in a soybean recombinant inbred line population of Essex x Forrest," M.S. thesis, Southern Illinois University at Carbondale, Carbondale, III, USA, 2007.

[43] V. N. Njiti, T. W. Doubler, R. J. Suttner, L. E. Gray, P. T. Gibson, and D. A. Lightfoot, "Resistance to soybean sudden death syndrome and root colonization by *Fusarium solani* f. sp. *glycine* in near-isogenic lines," *Crop Science*, vol. 38, no. 3, pp. 472–477, 1998.

[44] M. A. Kassem, K. Meksem, A. J. Wood, and D. A. Lightfoot, "A microsatellite map developed from late maturity germplasm 'Essex' by 'Forrest' detects four QTL for soybean seed yield expected from early maturing germplasm," *Reviews in Biology & Biotechnology*, vol. 6, no. 1, pp. 7–12, 2007.

[45] A. Alcivar, J. Jacobson, J. Rainho, K. Meksem, D. A. Lightfoot, and M. A. Kassem, "QTL underlying seedling root traits mapped in the 'Essex' by 'Forrest' soybean RIL population," *Annals of Botany*. In press.

[46] A. Alcivar, J. Jacobson, J. Rainho, K. Meksem, D. A. Lightfoot, and M. A. Kassem, "Genetic analysis of soybean plant height, hypocotyl and internode lengths," *Journal of Agricultural, Food, and Environmental Sciences*, vol. 1, no. 1, pp. 40–50, 2007.

[47] Y. Cho, V. N. Njiti, X. Chen, et al., "Quantitative trait loci associated with foliar trigonelline accumulation in *Glycine max* L," *Journal of Biomedicine and Biotechnology*, vol. 2, no. 3, pp. 1–7, 2002.

[48] J. A. Afzal, S. Fasi, R. Mungur, D. B. Goodenowe, and D. A. Lightfoot, "Comparisons of metabolic profiles by FT-ICR-MS and GC-MS using near-isolines that contrast for resistance to SCN and SDS," in *Plant & Animal Genome XV Conference*, San Diego, Calif, USA, January 2007.

[49] A. L. de Farias Neto, R. Hashmi, M. E. Schmidt, et al., "Mapping and confirmation of a new sudden death syndrome resistance QTL on linkage group D2 from the soybean genotypes 'PI 567374' and 'Ripley'," *Molecular Breeding*, vol. 20, no. 1, pp. 53–62, 2007.

[50] S. Kazi, V. N. Njiti, T. W. Doubler, et al., "Registration of the flyer x hartwig recombinant inbred line mapping population," *Journal of Plant Registrations*, vol. 1, no. 2, pp. 175–178, 2007.

[51] Soybean Genetics Committee Report, http://soybase.agron.iastate.edu/resources/QTL.php.

[52] S. Kazi, J. L. Shultz, R. Bashir, A. J. Afzal, V. N. Njiti, and D. A. Lightfoot, "Identification of loci underlying resistance to soybean sudden death syndrome in 'Hartwig' by 'Flyer'," *Theoretical and Applied Genetics*. Available online April 2008; doi:10.1007/s00122-008-0728-0.

[53] V. N. Njiti, R. J. Suttner, L. E. Gray, P. T. Gibson, and D. A. Lightfoot, "Rate-reducing resistance to *Fusarium solani* f. sp. *phaseoli* underlies field resistance to soybean sudden death syndrome," *Crop Science*, vol. 37, no. 1, pp. 132–138, 1997.

[54] Y. Luo, O. Myers Jr., D. A. Lightfoot, and M. E. Schmidt, "Root colonization of soybean cultivars in the field by *Fusarium solani* f. sp. *glycines* ," *Plant Disease*, vol. 83, no. 12, pp. 1155–1159, 1999.

[55] M. J. Iqbal, A. J. Afzal, S. Yaegashi, et al., "A pyramid of loci for partial resistance to *Fusarium solani* f. sp. *glycines* maintains myo-inositol-1-phoshate synthase expression in soybean roots," *Theoretical and Applied Genetics*, vol. 105, no. 8, pp. 1115–1123, 2002.

[56] M. J. Iqbal, S. Yaegashi, R. Ahsan, K. L. Shopinski, and D. A. Lightfoot, "Root response to *Fusarium solani* f. sp. *glycines*: temporal accumulation of transcripts in partially resistant and susceptible soybean," *Theoretical and Applied Genetics*, vol. 110, no. 8, pp. 1429–1438, 2005.

[57] K. L. Shopinski, M. J. Iqbal, J. L. Shultz, D. Jayaraman, and D. A. Lightfoot, "Development of a pooled probe method for locating small gene families in a physical map of soybean using stress related paralogues and a BAC minimum tile path," *Plant Methods*, vol. 2, pp. 20–28, 2006.

[58] Y. Cho, D. A. Lightfoot, and A. J. Wood, "Survey of trigonelline concentrations in salt-stressed leaves of cultivated *Glycine max* ," *Phytochemistry*, vol. 52, no. 7, pp. 1235–1238, 1999.

[59] K. Meksem, E. Ruben, D. Hyten, K. Triwitayakorn, and D. A. Lightfoot, "Conversion of AFLP bands into high-throughput DNA markers," *Molecular Genetics and Genomics*, vol. 265, no. 2, pp. 207–214, 2001.

[60] B. B. Bell-Johnson, G. R. Garvey, J. E. Johnson, K. Meksem, and D. A. Lightfoot, "Methods for high-throughput marker assisted selection for soybean," *Soybean Genetics Newsletter*, vol. 25, pp. 115–118, 1998.

[61] D. A. Lightfoot and M. J. Iqbal, "Marker assisted selection for soybean," in *Agricultural Biotechnology and Genomics*, V. Krishna, Ed., pp. 15–51, American Scientific, Stevensons Ranch, Calif, USA, 2003.

[62] M. J. Iqbal and D. A. Lightfoot, “Application of DNA markers: soybean improvement,” in *Molecular Marker Systems in Plant Breeding and Crop Improvement*, L. Horst and W. Gerhard, Eds., p. 475, Springer, New York, NY, USA, 2004.

[63] M. E. Schmidt, R. J. Suttner, J. H. Klein III, P. T. Gibson, D. A. Lightfoot, and O. Myers Jr., “Registration of LS-G96 soybean germplasm resistant to soybean sudden death syndrome (SDS) and soybean cyst nematode race 3,” *Crop Science*, vol. 39, no. 2, p. 598, 1999.

[64] B. A. McBlain, R. J. Fioritto, S. K. St. Martin, et al., “Registration of ‘Flyer’ soybean,” *Crop Science*, vol. 30, no. 2, p. 425, 1990.

[65] G. Stacey, A. Dorrance, H. Nguyen, et al., “SoyCAP: roadmap for soybean translational genomics,” white paper. USB, USDA, Beltsville, Md, USA, 2004.

[66] H.-J. Wang and P. A. Murphy, “Isoflavone composition of American and Japanese soybeans in Iowa: effects of variety, crop year, and location,” *Journal of Agricultural and Food Chemistry*, vol. 42, no. 8, pp. 1674–1677, 1994.

[67] V. S. Primomo, V. Poysa, G. R. Ablett, C.-J. Jackson, M. Gijzen, and I. Rajcan, “Mapping QTL for individual and total isoflavone content in soybean seeds,” *Crop Science*, vol. 45, no. 6, pp. 2454–2464, 2005.

[68] J. E. Specht, D. J. Hume, and S. V. Kumudini, “Soybean yield potential—a genetic and physiological perspective,” *Crop Science*, vol. 39, no. 6, pp. 1560–1570, 1999.

[69] J. E. Specht, “Report of yield QTL—coordination topic,” in *Soybean Breeders’ Workshop*, St. Louis, Mo, USA, February 2005.

[70] V. C. Concibido, B. La Vallee, P. Mclaird, et al., “Introgression of a quantitative trait locus for yield from *Glycine soja* into commercial soybean cultivars,” *Theoretical and Applied Genetics*, vol. 106, no. 4, pp. 575–582, 2003.

[71] J. W. Burton, “Quantitative genetics: results relevant to soybean breeding,” in *Soybeans: Improvement, Production, and Uses*, J. R. Wilcox, Ed., pp. 211–247, ASA, CSSA, and SSSA, Madison, Wis, USA, 2nd edition, 1987.

[72] D. Sun, Y. Du, Z. Zhang, et al., “GXE interactions and development influence loci underlying seed quality traits in soybean,” *Theoretical and Applied Genetics*. In press.

[73] J. A. Wrather, S. S. Koenning, and T. R. Anderson, “Effect of diseases on soybean yields in the United States and Ontario (1999 to 2002),” *Plant Health Progress*, pp. 1–16, 2003.

[74] D. Sandhu, H. Gao, S. Cianzio, and M. K. Bhattacharyya, “Deletion of a disease resistance nucleotide-binding-site leucine-rich-repeat-like sequence is associated with the loss of the Phytophthora resistance gene *Rps4* in soybean,” *Genetics*, vol. 168, no. 4, pp. 2157–2167, 2004.

[75] V. C. Concibido, B. W. Diers, and P. R. Arelli, “A decade of QTL mapping for cyst nematode resistance in soybean,” *Crop Science*, vol. 44, no. 4, pp. 1121–1131, 2004.

[76] T. L. Niblack, P. R. Arelli, G. R. Noel, et al., “A revised classification scheme for genetically diverse populations of *Heterodera glycines* ,” *Journal of Nematology*, vol. 34, no. 4, pp. 279–288, 2002.

[77] K. Meksem and D. A. Lightfoot, “Novel polynucleotides and polypeptides relating to loci underlying resistance to soybean cyst nematode and methods of use thereof,” US Patent no. 09772134, January 2001.

[78] A. J. Afzal and D. A. Lightfoot, “Soybean disease resistance protein RHG1-LRR domain expressed, purified and refolded from *Escherichia coli* inclusion bodies: preparation for a functional analysis,” *Protein Expression and Purification*, vol. 53, no. 2, pp. 346–355, 2007.

[79] N. S. Nelsen, A. L. Warner, D. A. Lightfoot, B. F. Matthews, and H. T. Knap, “Duplication of a chromosomal region from linkage group A2 involved in soybean cyst nematode resistance in soybean,” *Molecular Breeding*. In press.

[80] S. C. Anand, “Registration of ‘Hartwig’ soybean,” *Crop Science*, vol. 32, no. 4, pp. 1069–1070, 1992.

[81] D. M. Webb, B. M. Baltazar, A.P. Rao-Arelli, et al., “Genetic mapping of soybean cyst nematode race-3 resistance loci in the soybean PI 437.654,” *Theoretical and Applied Genetics*, vol. 91, no. 4, pp. 574–581, 1995.

[82] I. Schuster, R. V. Abdelnoor, S. R. R. Marin, et al., “Identification of a new major QTL associated with resistance to soybean cyst nematode (*Heterodera glycines*),” *Theoretical and Applied Genetics*, vol. 102, no. 1, pp. 91–96, 2001.

[83] R. Ahsan, M. J. Iqbal, A. J. Afzal, A. Jamai, K. Meksem, and D. A. Lightfoot, “Analysis of the activity of the soybean laccase encoded within the *Rfs2/Rhg1* locus,” *Theoretical and Applied Genetics*. In press.

[84] K. Meksem, K. K. Zobrist, E. Ruben, et al., “Two large-insert soybean genomic libraries constructed in a binary vector: applications in chromosome walking and genome wide physical mapping,” *Theoretical and Applied Genetics*, vol. 101, no. 5-6, pp. 747–755, 2000.

[85] C.-C. Wu, P. Nimmakayala, F. A. Santos, et al., “Construction and characterization of a soybean bacterial artificial chromosome library and use of multiple complementary libraries for genome physical mapping,” *Theoretical and Applied Genetics*, vol. 109, no. 5, pp. 1041–1050, 2004.

[86] T. Aoki, K. O’Donnell, Y. Homma, and A. R. Lattanzi, “Sudden-death syndrome of soybean is caused by two morphologically and phylogenetically distinct species within the *Fusarium solani* species complex: *F. virguliforme* in North America and *F. tucumaniae* in South America,” *Mycologia*, vol. 95, no. 4, pp. 660–684, 2003.

[87] K. Arumuganathan and E. D. Earle, “Estimation of nuclear DNA content of plants by flow cytometry,” *Plant Molecular Biology Reporter*, vol. 9, no. 3, pp. 229–241, 1991.

[88] R. J. Singh and T. Hymowitz, “The genomic relationship between *Glycine max* (L.) Merr. and *G. soja* Sieb. and Zucc. as revealed by pachytene chromosome analysis,” *Theoretical and Applied Genetics*, vol. 76, no. 5, pp. 705–711, 1988.

[89] R. C. Shoemaker, K. Polzin, J. Labate, et al., “Genome duplication in soybean (*Glycine* subgenus *soja*),” *Genetics*, vol. 144, no. 1, pp. 329–338, 1996.

[90] R. Shoemaker, P. Keim, L. Vodkin, et al., “A compilation of soybean ESTs: generation and analysis,” *Genome*, vol. 45, no. 2, pp. 329–338, 2002.

[91] G. Blanc and K. H. Wolfe, “Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes,” *The Plant Cell*, vol. 16, no. 7, pp. 1667–1678, 2004.

[92] A.-G. Tian, J. Wang, P. Cui, et al., “Characterization of soybean genomic features by analysis of its expressed sequence tags,” *Theoretical and Applied Genetics*, vol. 108, no. 5, pp. 903–913, 2004.

[93] C. R. Yesudas, J. L. Shultz, H. B. Zhang, G. K.-S. Wong, and D. A. Lightfoot, “A catalog of duplicated regions from marker amplicon homologs and BAC DNA sequence analysis in soybean, a paleopolyploid genome,” in *Plant & Animal Genome XIV Conference*, p. 33, San Diego, Calif, USA, January 2006.

[94] J. L. Shultz, J. D. Ray, and D. A. Lightfoot, “A sequence based synteny map between soybean and *Arabidopsis thaliana* ,” *BMC Genomics*, vol. 8, article 8, pp. 1–8, 2007.

[95] J. L. Shultz, C. R. Yesudas, S. Yaegashi, J. A. Afzal, S. Kazi, and D. A. Lightfoot, "Three minimum tile paths from bacterial artificial chromosome libraries of the soybean (*Glycine max* cv. 'Forrest'): tools for structural and functional genomics," *Plant Methods*, vol. 2, article 9, pp. 1–10, 2006.

[96] J. P. Tomkins, R. Mahalingam, H. Smith, J. L. Goicoechea, H. T. Knap, and R. A. Wing, "A bacterial artificial chromosome library for soybean PI 437654 and identification of clones associated with cyst nematode resistance," *Plant Molecular Biology*, vol. 41, no. 1, pp. 25–32, 1999.

[97] Y. Chen, Y. Wang, K. Meksem, and D. Wang, "Construction of a partial physical map for wild soybean," in *Plant & Animal Genome XI Conference*, p. 483, San Diego, Calif, USA, January 2003.

[98] C. Wu, S. Sun, P. Nimmakayala, et al., "A BAC- and BIBAC-based physical map of the soybean genome," *Genome Research*, vol. 14, no. 2, pp. 319–326, 2004.

[99] J. L. Shultz, K. Meksem, and D. A. Lightfoot, "Evaluating physical maps by clone location comparisons," *Genome Letters*, vol. 2, no. 3, pp. 109–118, 2003.

[100] N. Bouché and D. Bouchez, " *Arabidopsis* gene knockout: phenotypes wanted," *Current Opinion in Plant Biology*, vol. 4, no. 2, pp. 111–117, 2001.

[101] T. Nawy, J.-Y. Lee, J. Colinas, et al., "Transcriptional profile of the *Arabidopsis* root quiescent center," *The Plant Cell*, vol. 17, no. 7, pp. 1908–1925, 2005.

[102] K. Meksem, "Soybean tilling project," 2007, http://www.soybeantilling.org/tilling.jsp.

[103] J. L. Shultz, C. R. Yesudas, S. Yaegashi, et al., "Sequencing of BAC ends associated with an updated ECBAC1 minimum tile path of soybean (*Glycine max*)," GenBank ER962965 to ER966289 (3,324 sequences), 2007.

[104] M. C. Luo, J. Dvorak, J. L. Shultz, and D. A. Lightfoot, "Enhancement of the Forrest physical map with four color fluorescent fingerprints of a minimum tile path," In press.

[105] DOE Soybean Project, http://www.jgi.doe.gov/sequencing/cspseqplans2006.html.

[106] P. M. Olhoft, L. E. Flagel, C. M. Donovan, and D. A. Somers, "Efficient soybean transformation using hygromycin B selection in the cotyledonary-node method," *Planta*, vol. 216, no. 5, pp. 723–735, 2003.

[107] M. M. Paz, H. Shou, Z. Guo, Z. Zhang, A. K. Banerjee, and K. Wang, "Assessment of conditions affecting *Agrobacterium*-mediated soybean transformation using the cotyledonary node explant," *Euphytica*, vol. 136, no. 2, pp. 167–179, 2004.

[108] P. Zeng, D. A. Vadnais, Z. Zhang, and J. C. Polacco, "Refined glufosinate selection in *Agrobacterium*-mediated transformation of soybean [*Glycine max* (L.) Merrill]," *Plant Cell Reports*, vol. 22, no. 7, pp. 478–482, 2004.

[109] K. Triwitayakorn, "Dissection of gene clusters underlying resistance to *Fusarium solani* F. sp. *glycines* (*Rfs* loci) and *Heterodera glycines* (*Rhg* loci) in soybean," Ph.D. thesis, PLB, Southern Illinois University at Carbondale, Carbondale, Ill, USA, 2002.

[110] U. Z. Hammes, E. Nielsen, L. A. Honaas, C. G. Taylor, and D. P. Schachtman, "AtCAT6, a sink-tissue-localized transporter for essential amino acids in *Arabidopsis* ," *The Plant Journal*, vol. 48, no. 3, pp. 414–426, 2006.

[111] S. V. Wesley, C. A. Helliwell, N. A. Smith, et al., "Construct design for efficient, effective and high-throughput gene silencing in plants," *The Plant Journal*, vol. 27, no. 6, pp. 581–590, 2001.

[112] R. Collier, B. Fuchs, N. Walter, W. K. Lutke, and C. G. Taylor, "*Ex vitro* composite plants: an inexpensive, rapid method for root biology," *The Plant Journal*, vol. 43, no. 3, pp. 449–457, 2005.

[113] A. Jamai, L. Shiming, X. H. Liu, M. Goellner-Mitchum, H. Ishihara, and K. Meksem, "Molecular and functional analysis of the *Rhg4* locus conferring resistance to the soybean cyst nematode," in *Plant & Animal Genome XIV Conference*, p. 781, San Diego, Calif, USA, January 2007.

[114] P. Moolhuijzen, M. Cakir, A. Hunter, et al., "LegumeDB1 bioinformatics resource: comparative genomic analysis and novel cross-genera marker identification in lupin and pasture legume species," *Genome*, vol. 49, no. 6, pp. 689–699, 2006, erratum in *Genome*, vol. 49, no. 9, pp. 1206-1207, 2006.

[115] H. H. Yan, J. Mudge, D.-J. Kim, B. C. Shoemaker, D. R. Cook, and N. D. Young, "Comparative physical mapping reveals features of microsynteny between *Glycine max, Medicago truncatula*, and *Arabidopsis thaliana* ," *Genome*, vol. 47, no. 1, pp. 141–155, 2004.

[116] D. Grant, P. Cregan, and R. C. Shoemaker, "Genome organization in dicots: genome duplication in *Arabidopsis* and synteny between soybean and *Arabidopsis* ," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 8, pp. 4168–4173, 2000.

[117] J. A. Schlueter, P. Dixon, C. Granger, et al., "Mining EST databases to resolve evolutionary events in major crop species," *Genome*, vol. 47, no. 5, pp. 868–876, 2004.

[118] M. G. Francki and D. J. Mullan, "Application of comparative genomics to narrow-leafed lupin (*Lupinus angustifolius* L.) using sequence information from soybean and *Arabidopsis* ," *Genome*, vol. 47, no. 4, pp. 623–632, 2004.

[119] T.-Y. Hwang, J.-K. Moon, S. Yu, et al., "Application of comparative genomics in developing molecular markers tightly linked to the virus resistance gene *Rsv4* in soybean," *Genome*, vol. 49, no. 4, pp. 380–388, 2006.

[120] H.-W. Wang, J.-S. Zhang, J.-Y. Gai, and S.-Y. Chen, "Cloning and comparative analysis of the gene encoding diacylglycerol acyltransferase from wild type and cultivated soybean," *Theoretical and Applied Genetics*, vol. 112, no. 6, pp. 1086–1097, 2006.

[121] M. D. Gonzales, E. Archuleta, A. Farmer, et al., "The legume information system (LIS): an integrated information resource for comparative legume biology," *Nucleic Acids Research*, vol. 33, database issue, pp. D660–D665, 2005.

[122] J. Mudge, S. B. Cannon, P. Kalo, et al., "Highly syntenic regions in the genomes of soybean, *Medicago truncatula*, and *Arabidopsis thaliana* ," *BMC Plant Biology*, vol. 5, article 15, pp. 1–16, 2005.

[123] J. L. Shultz, S. Ali, L. Ballard, and D. A. Lightfoot, "Development of a physical map of the soybean pathogen *Fusarium virguliforme* based on synteny with *F. graminearum* genomic DNA," *BMC Genomics*, vol. 8, Article ID 262, 7 pages, 2007.

[124] V. N. Njiti, K. Meksem, D. A. Lightfoot, W. J. Banz, and T. A. Winters, "A method for breeding and genetically manipulating phytoestrogen content in soybeans," US Patent no. 10/008789, August 2000.

[125] D. A. Lightfoot, "Isolated polynucleotides and polypeptides relating to loci underlying seed protein and oil content and methods of use thereof," US Patent no. 07/2007, December 2007.

[126] D. A. Lightfoot, K. Meksem, and P. T. Gibson, "Method of determining soybean sudden death syndrome resistant in a soybean plant," US Patent no. 7288386, October 2007.

[127] W. J. Banz, M. P. Williams, D. A. Lightfoot, and T. A. Winters, "The effects of soy protein and soy phytoestrogens on symptoms associated with cardiovascular disease in rats,"

Journal of Medicine and Food, vol. 2, no. 3-4, pp. 271–273, 1999.

[128] J. A. Greer-Baney, W. J. Banz, D. A. Lightfoot, and T. A. Winters, "Dietary soy protein and soy isoflavones: histological examination of reproductive tissues in rats," *Journal of Medicinal Food*, vol. 2, no. 3-4, pp. 247–249, 1999.

[129] M. J. Iqbal, S. Yaegashi, R. Ahsan, D. A. Lightfoot, and W. J. Banz, "Differentially abundant mRNAs in rat liver in response to diets containing soy protein isolate," *Physiological Genomics*, vol. 11, no. 3, pp. 219–226, 2003.

[130] Y. Cho, V. N. Njiti, X. Chen, D. A. Lightfoot, and A. J. Wood, "Trigonelline concentration in field-grown soybean in response to irrigation," *Biologia Plantarum*, vol. 46, no. 3, pp. 405–410, 2003.

Hindawi Publishing Corporation
International Journal of Plant Genomics
Volume 2008, Article ID 256597, 11 pages
doi:10.1155/2008/256597

Review Article

Recent Advances in *Medicago truncatula* Genomics

Jean-Michel Ané,[1] Hongyan Zhu,[2] and Julia Frugoli[3]

[1] *Department of Agronomy, University of Wisconsin, Madison, WI 53706, USA*
[2] *Department of Plant and Soil Sciences, University of Kentucky, Lexington, KY 40546, USA*
[3] *Department of Genetics and Biochemistry, Clemson University, 100 Jordan Hall, Clemson, SC 29634, USA*

Correspondence should be addressed to Julia Frugoli, jfrugol@clemson.edu

Received 2 April 2007; Accepted 14 September 2007

Recommended by Yunbi Xu

Legume rotation has allowed a consistent increase in crop yield and consequently in human population since the antiquity. Legumes will also be instrumental in our ability to maintain the sustainability of our agriculture while facing the challenges of increasing food and biofuel demand. *Medicago truncatula* and *Lotus japonicus* have emerged during the last decade as two major model systems for legume biology. Initially developed to dissect plant-microbe symbiotic interactions and especially legume nodulation, these two models are now widely used in a variety of biological fields from plant physiology and development to population genetics and structural genomics. This review highlights the genetic and genomic tools available to the *M. truncatula* community. Comparative genomic approaches to transfer biological information between model systems and legume crops are also discussed.

1. INTRODUCTION

Legumes are usually defined by their typical flower structure and the ability of many of them to form root nodules in presence of symbiotic bacteria named rhizobia. With more than 18 000 species, legumes are found from the artic circle to the tropics and include many crops of agronomic importance for grain production, pasture, and forestry [1, 2]. The ability of more than 88% of legumes to obtain nitrogen from the air through root nodules was probably a major determinant in this evolutionary, ecological, and economical success [3]. Interestingly, the study of symbiotic associations with rhizobia as well as with arbuscular mycorrhizal (AM) fungi also drove the development of two model legumes: *Medicago truncatula* Gaertner and *Lotus japonicus* (Regel) K. Larsen.

While *M. truncatula* is an annual medic from the Trifolieae tribe and a close relative of alfalfa and clovers, *L. japonicus* belongs to the Loteae and is more distant from cultivated cool season legumes than *M. truncatula*. This phylogenetic distance to economically important crops is critical in the choice of *M. truncatula* by many researchers and support by numerous funding agencies. The use of both model legumes allows unique comparative genomic studies within the legume family as well as the comparison between two patterns of root nodule development: indeterminate with a persistent nodule meristem in the case of *M. truncatula* and determinate in *L. japonicus*. Unfortunately, these two models belong to the same cool season legumes (Galegoid clade), whereas soybean and common bean are tropical season legumes (Phaseolid clade). Soybean is therefore proposed as a third model legume for both its own economic weight and the phylogenetic proximity to other important crops [4, 5].

Research efforts on model legumes and especially on *M. truncatula* encompass a broad range of fields in plant biology from population biology [6–8] and plant development [9–16] to plant pathology [17–22], insect resistance [23–27], and biotechnology production [28]. The goal of this review is to provide an overview of the natural characteristics and genetic and genomic tools that make *M. truncatula* such a desirable experimental system for a growing number of plant biologists. We will highlight how information gained from *M. truncatula* can be transferred to other legumes crops through comparative genomics and we will share our vision of how *M. truncatula* can allow us to reach the goal of sustainable well-being through sustainable food and biofuel production.

2. *MEDICAGO TRUNCATULA* AS A MODEL LEGUME

Natural attributes of *M. truncatula* that make it a valuable genetic model include its annual habit and rapid life cycle, its diploid ($2n = 16$) and autogamous nature, its prolific seed production, and a relatively small genome of about 550 Mb. Jemalong A17 has been selected by the research community as a reference line for most genetic and genomic approaches and is derived from the major commercial cultivar. *M. truncatula* is native to the Mediterranean basin and is found in a wide range of habitats. It is therefore not surprising to find a high level of variation among and within natural populations [29]. Using microsatellite markers, a publicly available core-collection of 346 inbred lines was developed and thus represents the breadth of this natural diversity [30]. *M. truncatula* is used as a fodder crop in ley-farming systems in Australia, and a large and diverse collection is housed at the South Australian Research and Development Institute (SARDI) [8].

Like many higher plants, *M. truncatula* forms symbiotic associations with a wide array of arbuscular mycorrhizal (AM) fungi. As a legume, *M. truncatula* is also able to develop root nodules with *Sinorhizobium meliloti,* which is one of the best-characterized rhizobium species at the genetic level [31]. Cultivation-independent techniques have been used to sample the diversity of microbes associated with *M. truncatula* roots at various developmental stages and they reveal an extremely dynamic genetic structure of its rhizosphere [32].

Mutagenesis approaches using ethyl methane sulfonate (EMS), gamma rays, and fast neutron bombardment (FNB) have generated large mutant populations of *M. truncatula* from which mutants affected in symbiotic as well as developmental pathways have been identified [33–37]. T-DNA and *Tnt1* mutagenesis have been developed recently to generate tagged mutants for forward and reverse genetics purposes [38–40].

Several protocols have been optimized to transform *M. truncatula* using *Agrobacterium tumefaciens* [41–45]. These protocols are particularly efficient for the R108 and Jemalong 2HA lines but the regeneration efficiency still needs to be improved for Jemalong A17. This moderate efficiency as well as the time required for the regeneration steps is driving the preference of the *Medicago* community towards *Tnt1* versus T-DNA for gene tagging approaches [46] as well as the search for alternative transformation systems.

Hairy root transformation via *Agrobacterium rhizogenes* proved to be a rapid and efficient transformation system allowing the generation of transgenic roots in 2-3 weeks. Such "hairy roots" can be infected by rhizobia or AM fungi with symbiotic phenotypes indistinguishable from nontransgenic roots and are therefore an ideal system for plant-microbe symbiosis studies [47]. The development of DsRed as a visual reporter reduced the need for Kanamycin or Basta selection systems which were significantly decreasing nodulation efficiency. This hairy root transformation system is now used routinely to express protein fusions or RNA interference (RNAi) constructs [48–50]. The possibility to regenerate transgenic plants from hairy roots of the R108 line has been reported recently. This flexible approach should allow a rapid initial screening of phenotypes on hairy roots and a subsequent regeneration of transgenic plants if necessary [41]. An interesting ex vitro procedure that eliminates the need for labor-intensive in vitro culture will undoubtedly increase the throughput of hairy root transformations to a level compatible with genomic studies [51].

The *Medicago* community has therefore identified many ecotypes and developed a wide range of mutants and transgenic lines. A current goal of the International *Medicago truncatula* steering committee is to address the need for a stock center able to maintain, amplify, and distribute these lines to an ever growing community.

3. MAPPING THE GENOME OF *MEDICAGO TRUNCATULA*

Genetic and cytogenetic tools have been instrumental to the development of a "gene rich" genome sequence for *M. truncatula*. This project also required several bacterial artificial chromosomes (BAC) libraries that were developed using *Hin*dIII and *Eco*RI partial digests as well as a robust physical map (Figure 1).

Genetic maps have been developed from F2 populations and a wide array of genetic markers such as CAPS, AFLPs, RAPDs, and microsatellites (SSRs) [52–54]. One of them, based on a Jemalong A17 A17 $\times$ A20 F2 population, is currently used as a reference for the genome sequencing project (http://www.medicago.org/genome/map.php). Unfortunately, these F2 populations are either based on a limited amount of genomic DNA or require a labor-intensive vegetative propagation of F2 individuals. In order to provide sustainable tools to the community, genetic maps based on recombinant inbred lines (RILs) and highly polymorphic microsatellite markers are developed and will undoubtedly represent the future reference for *M. truncatula* genetics (T. Huguet, personal communication).

Cytogenetic maps based on fluorescence *in situ* hybridization (FISH) with interphase or metaphase chromosomes provide a quick access to the chromosomal location of BAC clones and repeated sequences [55–57]. Obtaining pachytene chromosomes is more labor intensive than metaphase chromosomes but provides an unequalled resolution all along the chromosome and particularly in euchromatic regions [58]. Information from such cytogenetic tools was instrumental for comparative genomics and map-based cloning projects but also allowed the determination that *M. truncatula* heterochromatin was mostly localized in pericentromeric regions. Genetic and cytogenetic markers corresponding to the borders of these regions have been developed [57, 59, 60]. Based on this unique chromosomal structure, it is therefore possible to predict through the genetic map if a BAC clone belongs to a euchromatic or a heterochromatic region. This observation as well as the possibility to select EST-rich BAC clones led the *M. truncatula* community to initiate the sequencing of euchromatic (gene-rich) regions via a BAC-by-BAC strategy (http://www.medicago.org/genome/).

Four centers share the sequencing effort of the 8 chromosomes: Bruce Roe et al. at the University of Oklahoma, Chris Town et al. at The Institute for Genomic Research (TIGR), Jane Rogers et al. at the Sanger Centre, and Francis

Quétier et al. at the Genoscope. A physical map grouping and ordering of BAC clones was developed by the laboratory of Douglas R. Cook by combining *Hin*dIII digestion fingerprints with BAC-end sequence data through the FPC software [52, 61, 62]. More than 1370 FPC contigs cover 466 Mbp (93% of the genome) and are used to determine the minimum tiling path of gene-rich regions for whole genome sequencing [52].

As of February 2007, 188 Mb of genome sequence from 1950 BAC clones are publicly available. About 10% of this information is redundant due to the overlap of BAC clones necessary to create a tiling path and more than 300 gaps between contigs need to be filled. These gaps are sized by FISH and covered with contigs by long-range PCR or classical chromosome walking [62].

Integration of genetic, cytogenetic, physical, and sequence maps allowed the development of pseudochromosomes and greatly facilitated comparative mapping [52, 58–60]. Annotating pseudochromosome sequences is classically achieved through gene prediction programs and comparison with EST databases (Figure 1). The IMGAG (International *Medicago* Genome Annotation Group) has developed a unique automated pipeline to predict gene structures and functions [63]. More than 25 000 genes have been predicted so far and techniques to test these predictions need to be developed.

Oligonucleotides covering the entire sequence of pseudochromosomes can be printed on glass slides to generate tiling arrays. These arrays can be used for a wide range of applications from gene identification and detection of alternative splicing to comparative genome hybridization (CGH) and chromatin immunoprecipitation on chips (ChIP chips) [64–67].

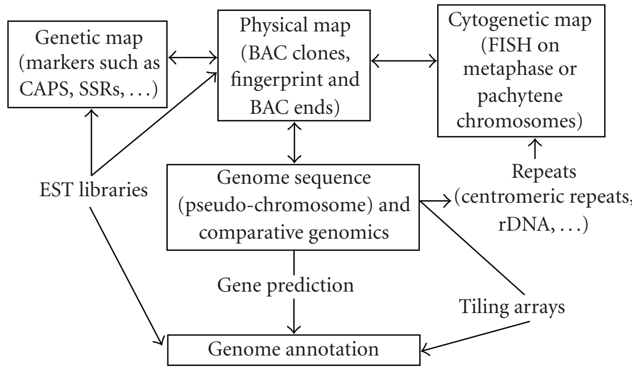


Figure 1: Integration of different maps and libraries to generate and annotate the genomic sequence of *M. truncatula.* Expressed sequence tags (EST) are used to generate genetic markers and to identify BAC clones in gene-rich regions as well as for gene identification. Repeats identified via genome sequencing and comparison with other species can be mapped via FISH on chromosome spreads.

4. SYSTEMS ANALYSIS

4.1. Transcriptomics

Large-scale EST sequencing is essential for functional genomics studies, permitting the direct identification of large gene collections and setting the stage for further analysis, such as those using DNA microarray technology. Several large EST projects have been completed [68–71]. The analysis of the almost 200 000 ESTs isolated from many different libraries constructed from diverse stages and treatments that came out of these projects is facilitated by searchable databases such as MtDB2 [72] and the TIGR Gene Index (http://www.tigr.org).

Both microarray and macroarray analyses of gene expression changes during symbiosis have been published [73–78]. These experiments ranged from analysis of a few thousand genes on filters during AM symbiosis [73] to almost 10 000 genes compared between wild type and nonnodulating mutants [76, 77] or between fix-mutants [79]. A dual symbiosis chip containing 10 000 *M. truncatula* genes and the entire *S. meliloti* prokaryotic genome allows side by side analysis of both partners in the symbiosis [80], and an Affymetrix chip with bioinformatically optimized oligonucleotides representing 48 000 genes is available (http://www.affymetrix.com/support/technical/datasheets/medicago_datasheet.pdf). As genome sequencing continues, following the expression of all *M. truncatula* genes under varying conditions should soon be possible. Affymetrix placed probe sets for 1850 *M. sativa* transcripts on these chips to facilitate the study of closely related species such as *M. sativa*. The use of *M. truncatula* arrays for analysis of *M. sativa* (crop alfalfa) gene expression has proven effective [81, 82].

Other effective genomic approaches to transcriptional analysis utilized to date in *M. truncatula* include suppressive subtractive hybridization (SSH) and serial analysis of gene expression (SAGE). In SSH, suppressive PCR is used to both normalize the abundance of transcripts in individual libraries and enrich for transcripts unique to the library by subtracting sequences common to several libraries, with rare sequences being enriched up to 1000 folds [83]. This method has been used to identify AM specific transcripts [84] and transcripts specifically involved in the *S. meliloti* symbiosis [85]. SAGE is a method for comprehensive analysis of gene expression patterns using short sequence tags obtained from a unique position within each transcript (10–14 bp) to uniquely identify a transcript. The expression level of the corresponding transcript is determined by quantifying of the number of times a particular tag is observed [86]. Although no publications have arisen yet, a project applying SAGE to *M. truncatula* is underway at the Center for *Medicago* Genomics Research at the Nobel Foundation (http://www.noble.org/medicago/GEP.html).

4.2. Proteomics

Another complementary approach to identify import gene products involved in interesting processes is to look at changes in the protein complement of a genome that vary by cell or treatment. In order for proteomic approaches to be useful in a system, a large sequence resource is necessary to match the sequences of peptides generated in tryptic digests to their proteins of origin. The growing sequence resource in *M. truncatula* allows identification of proteins by their mass spectra, making proteomics an effective approach for *M. truncatula* and proteomics approaches have become quite popular. A comprehensive review of considerations important in proteomics technology and applications in *M. truncatula* and *Arabidopsis* was recently published [87, 88]. Because small peptides have been shown to have roles in plant signaling, proteomics has been applied to identifying small protein/peptide components of certain *M. truncatula* tissues [89]. Proteomic approaches have also been applied to analyses of seed development [14, 16], pathogen interactions [90], symbiosome membranes [91], AM membranes [92], root microsomes [93], and other organ, tissue, and treatment-specific approaches [11, 94–100].

Most of the genes cloned thus far in the initial signal transduction pathway for nodulation are kinases [101]; suggesting global analysis of phosphoproteins is a way to identify important genes involved in signal transduction in *M. truncatula*. Unfortunately, phosphoproteins involved in cellular signaling are generally present in low abundance, creating new challenges for proteomics. By making adjustments the basic proteomics procedures, such as adding an enrichment step, a proof of concept experiment in *M. truncatula* phosphoproteomics, gives a taste of the potential of this approach [102].

4.3. Metabolomics

Alfalfa produces a number of secondary metabolites of great interest because of their contributions to human health and animal forage quality. The principle behind metabolomics is that metabolic profiling on a genomic scale offers a view of the metabolic status of an organism, which can lend insight to the study gene function or whole plant biology [103]. Successful attempts to link proteomics, transcriptomics, and metabolomics for cell cultures in *M. truncatula* have emerged from these studies [104, 105].

Metabolomics is a new and evolving science, and requires specialized equipment and multifaceted technical strategies. The Nobel Foundation employs a strategy that utilizes sequential or selective extraction followed by parallel analyses. The parallel analyses achieve a comprehensive view of the metabolome with high-performance liquid chromatography (HPLC), capillary electrophoresis (CE), gas chromatography (GC), mass spectrometry (MS), and various combinations of the above techniques such as GC/MS, LC/MS, and CE/MS. In addition to studying biological responses to biotic and abiotic elicitors in *M. truncatula* cellcultures, these techniques are being applied to the study of natural variants in *M. truncatula*, *M. truncatula* development, lignin biosynthesis, and legume-insect interactions.

Perhaps the most daunting aspect of metabolomic experiments is the analysis of the data. Early on, it became obvious that metabolomics required a standard similar to MIAME (minimum information about a microarray experiment) to allow comparison of data. A framework for the description of plant metabolomic experiments and their results has recently been developed. ArMet (architecture for metabolomics) is published and in accepted use [106, 107].

4.4. Phenomics

As more and more researchers use *M. truncatula* as a model, the need for a standardized method of describing phenotypes becomes acute. Since the timing and structure of vegetative and floral development in *M. truncatula* differ from *Arabidopsis*, adoption of standards such as those used for *Arabidopsis* [108] is inappropriate. Additionally, *M. truncatula* symbioses with AM fungi and *Sinorhizobium meliloti* add another dimension to developmental processes that require a standardized description of process stages and plant anatomy.

To date, a few attempts have been made to develop a standardized language for comparison. Vegetative growth parameters were carefully measured to provide a benchmark in [109], but the use of a glasshouse environment rather than a controlled light and temperature regime rendered the data not universally applicable. Likewise, flower development and response to vernalization have been documented in the same way [109], again in a glasshouse so the light intensity was uncontrolled. These experiments are progress toward a controlled standard for comparison of mutant phenotypes such as “late flowering” or “increased internodal distance.” Precision in phenotypic descriptions will be critical to genome scale mutant hunts.

There is no plant structural GO ontology terms for nodulation or nodule structures in the Plant Ontology Consortium site as of the February 2007 release (http://www.plantontology.org). The present plant ontology system provides terms for growth and developmental stages, as well as organs and tissues of *Arabidopsis*, maize and rice, but none of these plants nodulates, creating a problem for using GO annotation in *M. truncatula*.

4.5. Bioinformatics

All of the “omics” scale tools discussed above necessitate strong bioinformatics infrastructure for the species. A good place to begin is the *Medicago* Consortium website: http://www.medicago.org. In addition to a handbook of protocols for everything from growing and transforming *M. truncatula* to naming genes, links from this page lead to informatics tools such as ENSEMBL which allow a real time view of the annotation of the genome, tools allowing browsing of the genome for markers, genes, the location of BACs, the status of the sequencing project or the sequence status of any individual BAC. Users can also view the contigs assembled for sequencing, and make comparisons to other legumes through the legume information system [110] and the consensus legume database (www.legumes.org). Tools are also available through links from the medicago.org website for examining ESTs (TIGR, MtDB2, MENS), and in the future, examining microarray data. *In silico* approaches in *M. truncatula* have led to important insights, such as the identification of a large family of small legume-specific transcripts with conserved cysteine motifs whose function continues to be investigated [111–113].

But as genome scale biology is applied, the need to synthesize transcriptomics data, proteomics data, metabolomics data, and more becomes as important as the availability of informatics tools to analyze these data individually. Several steps in this direction have occurred within the *M. truncatula* community. Some of these integrated solutions are focused around a process, such as gene expression in mycorrhizal symbiosis [114]. Because of the nearly complete genome sequence and the cooperative nature of the sequencing and annotation of the genome, comprehensive integration of various data sources has been necessary from the beginning. Cannon et al. [115] provide a nice summary of the available sequence-based resources and how they interact. A freely available database of biochemical pathway data for *M. truncatula* (MedicCyc) contains more than 250 pathways with related genes, enzymes, and metabolites [116]. This provides the ability to not only visualize metabolomics data and integrate them with functional genomics data, but also allow comparison of *M. truncatula* pathways to those in other plants using the compatible AraCyc and RiceCyc databases.

5. REVERSE GENETICS

Reverse genetics approaches which identify mutants in a gene of interest based on sequence differences are critical genomic tools in a model system. A range of approaches are available, including retrotransposon tagging, T-DNA tagging, TILLING for EMS mutations, PCR screening for fast neutron mutations, and RNA-induced gene silencing (RNAi) [46]. Each method has advantages and disadvantages, and the choice of which method(s) to use will depend on the purpose of the investigator. In *M. truncatula*, RNAi, TILLING, and PCR screening of *Tnt1* insertion mutagenesis populations or fast-neutron generated deletion populations are reverse genetic approaches presently possible.

As noted above, the efficacy of RNAi in *M. truncatula* has been documented [48] including use in whole plants and in transformed roots. The combination of RNAi constructs and hairy root transformation is useful for large scale screening projects to identify genes of interest for further analysis. A large-scale project to identify gene function by silencing in *M. truncatula* is underway (http://www.cbs.umn.edu/labs/ganttlab/rnai.html). Initial results from this project include identification of a calcium-dependent protein kinase involved in nodule development, a gene that had not been identified through classical mutational analysis [117].

TILLING (Targeting Induced Local Lesions in Genomes) has proven useful in *Arabidopsis* and other plants (reviewed in [118]). Briefly, the sequence of a gene of interest is analyzed with a computer program that determines the consequences of all possible EMS mutations (primarily G to A transitions) on the amino acid sequence of the deduced protein. Regions are chosen for PCR amplification based on the concept that those regions most likely to result in altered protein function are highly conserved domains in proteins, and PCR primers are designed to amplify these regions of DNA from each plant. The PCR products are analyzed for single-base pair changes in a high throughput sequencing gel system using an enzyme that detects and cleaves single-base mismatches in DNA. The use of high throughput methodology and a well-characterized and curated population of mutagenized plants allows a plant containing a lesion in the gene of interest to be identified in days. The benefits of TILLING are not only the rapid identification of lesions, but the nature of the lesions themselves. The point mutations generated by EMS treatment allow the use of TILLING to generate an allelic series that includes both missense and nonsense mutations.

In *M. truncatula*, the Cook lab. at UC Davis developed a population of ~4000 curated EMS mutagenized plants for purpose of TILLING. This resource is currently unavailable as a community resource due to the absence of funding. To date genotypic screens for mutations in 15 genes of interest to the Cook lab or collaborators have been undertaken, and 143 mutants identified, with recovery rates of 9.89 alleles per kbp screened. Phenotypic characterization of one (of 23 unique) allele identified from one of the early genotypic screens for mutations in the *M. truncatula* arbuscule specific phosphate transporter *MtPT4* is described by Javot et al. [119]. Characterization of other mutants in this collection is currently ongoing or advanced to the stage where manuscripts are in preparation for submission (Douglas R. Cook and Varma Penmetsa, personal communication).

A reverse-genetics platform has been established in *Medicago truncatula* exploiting fast neutron (FN) mutagenesis and a highly sensitive PCR-based detection first documented in *Arabidopsis* [120]. The FN-based screening platform produces complete loss of function mutants by identifying large deletions in the targeted region. Central to this platform is the development of a detection strategy which allows a mutant amplicon, possessing an internal deletion, to be preferentially amplified in pools where genomic target sequence is present at a 20 000-fold excess. This detection sensitivity has been achieved through a combination of techniques for suppressing the amplification of the wild-type sequence and preferentially amplifying the mutant product. The population has been arrayed such that 12 000 M2 plants can be analyzed in 4 PCR reactions. These megapools can then be dissected using 25 PCR reactions on 3D pools, allowing identification of the individual seed lot containing the mutant. In comparison with the well-established TILLING method [121–123], which utilizes 8-fold PCR-based screening, FN alleles can be isolated at a fraction of the cost and avoid the problems associated with EMS mutagenesis of targeting small genes and the very high number of background mutations in isolated mutants. An initial characterization of the FN system analyzed 10 genes in a subpopulation of 60 000 M2 plants. Mutants were recovered for 4 target genes. A population of 180 000 M2 plants has now been established and should allow the recovery of mutants from a majority of targeted loci. Information for accessing this resource can be found at www.jicgenomelab.co.uk. (C. Rodgers and G. E. D. Oldroyd, personal communication).

Recently, researchers have identified a tobacco retrotransposon, *Tnt1*, that moves randomly in *M. truncatula* but only upon passage through tissue culture [39]. This retrotransposon can be used to generate a large population of plants with tagged mutation sites in tissue culture that become stable upon regeneration of whole plants, an important resource for both forward and reverse genetics. A population mutagenized by *Tnt1* can be used for reverse genetic screens by sequencing of tagged sites and forward genetic screens by observation of phenotypes. The isolation of the *M. truncatula pim* gene through this reverse genetics approach demonstrates the utility of the system for identifying mutants by sequence [40].

6. TRANSLATIONAL GENOMICS FROM MODEL TO CROP LEGUMES

The value of the model systems will be enhanced by the ability to connect model systems to crops at the structural and functional genome levels. For example, conserved genome structure (synteny) between model and crop species could allow the use of model species as a surrogate genome for map-based cloning of agronomically important genes in crops with complex genomes. Moreover, detailed knowledge of the molecular basis of conserved phenotypes in model species can be translated to great advantage for gene discovery in related species. Working with *M. truncatula* as a reference system, researchers have used comparative genomics tools to bridge model and crop legumes through comparative mapping of orthologous genes [54, 124, 125]. Alignment of linkage maps and sequenced orthologous regions reveals an extensive network of macro- and microsynteny between legume species [125–127]. In fact, the conserved genome organization between *M. truncatula* and crop legumes has allowed for cross-species prediction and isolation of several genes required for root symbiosis using *M. truncatula* as a surrogate [128, 129] and reviewed in [125]. Despite the emerging picture of substantial synteny between legumes, the level of conservation decreases as the evolutionary distance increases [124, 125]. Thus, comparisons within Galegoid or Phaseolid legumes tend to reveal chromosome-level synteny, while comparisons between the two clades tend to reveal large-segment synteny, which is also reflected in the differences in chromosome number between Galegoid and Phaseolid legumes [125, 127].The broad taxonomic distance separating the two clades warrants the development of one or two reference systems within each clade, *M. truncatula,* and *L. japonicus* for the cool-season legumes and soybean for the tropical-season legumes [130].

A significant effort has been undertaken in comparative genomic analysis of legume resistance gene homologs (RGHs). Most plant disease resistance genes identified to date belong to the nucleotide binding site (NBS) leucine rich repeat (LRR) family [131]. NBS-LRR genes can be further classified by the presence or absence of a toll/interleukin receptor (TIR) homology domain. In previous studies, researchers investigated the genomic architecture of RGHs in *M. truncatula* [21], and used phylogenetic methods to assess evolutionary trends in this large gene family in legumes and across the angiosperms [21, 132]. The results from these studies revealed several important insights into RGH gene evolution in plants. Despite the presence of the two major lineages of RGHs (i.e., TIR and non-TIR NBS-LRR genes) in all dicots, each of these lineages is populated by numerous family-specific or family-predominant clades [132]. For example, the major RGH clades that define legumes are absent from the Brassicaceae and Solanaceae, and vice versa. Thus, there are likely to be aspects of RGHs (including both structural and functional attributes) that are peculiar to individual plant families. When phylogenetic analyses were conducted within the legume family [21], it was found that all known major clades in legumes are represented by sequences from *M. truncatula*, providing evidence that the major RGH radiations predate the respective speciation events. There are also cases that cophyletic RGHs occupy syntenic positions between legumes. The availability of a nearly complete catalog of *M. truncatula* NBS-LRR genes is expected to greatly enable rapid and efficient characterization of RGHs in other closely related legumes. A legume genome project towards this effort has recently been funded by the NSF Plant Genome Research Program. The goal of this funded project was to develop genomic tools for five less-studied legume species (i.e., chickpea, pigeon pea, cowpea, peanut, and lupine), which are economically important in the developing countries of Africa and Asia (D. R. Cook, personal communication).

Forage legumes, such as alfalfa, red clover, and white clover, are an important component of animal and sustainable agriculture throughout the world. In addition to providing superior forage quality for animal production and improving soil fertility through nitrogen fixation, forage legumes also contribute to the improvement of soil structure and control of soil erosion. Alfalfa (*Medicago sativa*), for example, is grown on over 26 million acres and ranks third in acreage planted and dollar value in the US (USDA Crop Values Summary 2005). The true clovers (*Trifolium* spp.), which are often grown together with forage grasses, are also widely distributed. Despite serving as a major source of meat and milk products via animals, the economic importance of forages to food production and the agricultural economy of the US are not fully appreciated. Consequently, forage legumes suffer from poorly developed genetic and genomic infrastructure due to both limited federal funding and their intractable genetic system (e.g., polyploidy and self-incompatibility). The lack of such infrastructure limits the application of genomics-enabled technologies in the genetic improvement of forage legumes. Nevertheless, all these forage legumes are closely related to the model legume *M. truncatula*, a cool-season legume within the tribe Trifolieae. Therefore, forage legumes could be an immediate beneficiary of the study of *M. truncatula* genomics. As many of the pathogens of *M. truncatula* are also pathogens of closely related forage legumes, it should be possible to clone resistance genes that are active against pathogens of crop legume species in *M. truncatula*. In addition, due to the close relationship of resistance gene sequences between these species, it is likely that functional resistance genes can be moved across species boundaries by transgenic approaches.

Thus the genetic, genomic, and molecular tools available in *M. truncatula* allow not only investigation of basic processes important to legumes, but also transfer of that information to important crop species.

ACKNOWLEDGMENTS

The authors thank Dr. Kannan Iyyemperumal for his critical reading of the manuscript and useful discussion, and Giles E. D. Oldroyd, Christian Rodgers, Varma Penmetsa, and Douglas R. Cook for sharing unpublished information. Technical contribution no. 5319 of the Clemson University Experiment Station is highly acknowledged.

REFERENCES

[1] R. M. Polhill, "Classification of the Leguminosae," in *Phytochemical Dictionary of the Leguminosae*, F. A. Bisby, J. Buckingham, and J. B. Harborne, Eds., vol. 1, pp. 35–56, Chapman & Hall, New York, NY, USA, 1994.

[2] P. H. Graham and C. P. Vance, "Legumes: importance and constraints to greater use," *Plant Physiology*, vol. 131, no. 3, pp. 872–877, 2003.

[3] S. M. de Faria, G. P. Lewis, J. I. Sprent, and J. M. Sutherland, "Occurrence of nodulation in the Leguminosae," *New Phytologist*, vol. 111, no. 4, pp. 607–619, 1989.

[4] S. A. Jackson, D. Rokhsar, G. Stacey, R. C. Shoemaker, J. Schmutz, and J. Grimwood, "Toward a reference sequence of the soybean genome: a multiagency effort," *Crop Science*, vol. 46, supplement 1, pp. S-55–S-61, 2006.

[5] G. Stacey, L. Vodkin, W. A. Parrott, and R. C. Shoemaker, "National Science Foundation-sponsored workshop report. Draft plan for soybean genomics," *Plant Physiology*, vol. 135, no. 1, pp. 59–70, 2004.

[6] I. Bonnin, J.-M. Prospéri, and I. Olivieri, "Genetic markers and quantitative genetic variation in *Medicago truncatula* (Leguminosae): a comparative analysis of population structure," *Genetics*, vol. 143, no. 4, pp. 1795–1805, 1996.

[7] I. Bonnin, J. Ronfort, F. Wozniak, and I. Olivieri, "Spatial effects and rare outcrossing events in *Medicago truncatula* (Fabaceae)," *Molecular Ecology*, vol. 10, no. 6, pp. 1371–1383, 2001.

[8] S. R. Ellwood, N. K. D'Souza, L. G. Kamphuis, T. I. Burgess, R. M. Nair, and R. P. Oliver, "SSR analysis of the *Medicago truncatula* SARDI core collection reveals substantial diversity and unusual genotype dispersal throughout the Mediterranean basin," *Theoretical and Applied Genetics*, vol. 112, no. 5, pp. 977–983, 2006.

[9] N. Imin, M. Nizamidin, T. Wu, and B. G. Rolfe, "Factors involved in root formation in *Medicago truncatula*," *Journal of Experimental Botany*, vol. 58, no. 3, pp. 439–451, 2007.

[10] R. J. Rose, X.-D. Wang, K. E. Nolan, and B. G. Rolfe, "Root meristems in *Medicago truncatula* tissue culture arise from vascular-derived procambial-like cells in a process regulated by ethylene," *Journal of Experimental Botany*, vol. 57, no. 10, pp. 2227–2235, 2006.

[11] P. Holmes, R. Farquharson, P. J. Hall, and B. G. Rolfe, "Proteomic analysis of root meristems and the effects of acetohydroxyacid synthase-inhibiting herbicides in the root of *Medicago truncatula*," *Journal of Proteome Research*, vol. 5, no. 9, pp. 2309–2316, 2006.

[12] Q. C. B. Cronk, "Legume flowers bear fruit," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 13, pp. 4801–4802, 2006.

[13] B. Bucciarelli, J. Hanan, D. Palmquist, and C. P. Vance, "A standardized method for analysis of *Medicago truncatula* phenotypic development," *Plant Physiology*, vol. 142, no. 1, pp. 207–219, 2006.

[14] J. Boudet, J. Buitink, F. A. Hoekstra, et al., "Comparative analysis of the heat stable proteome of radicles of *Medicago truncatula* seeds during germination identifies late embryogenesis abundant proteins associated with desiccation tolerance," *Plant Physiology*, vol. 140, no. 4, pp. 1418–1436, 2006.

[15] N. Djemel, D. Guedon, A. Lechevalier, et al., "Development and composition of the seeds of nine genotypes of the *Medicago truncatula* species complex," *Plant Physiology and Biochemistry*, vol. 43, no. 6, pp. 557–566, 2005.

[16] K. Gallardo, C. Le Signor, J. Vandekerckhove, R. D. Thompson, and J. Burstin, "Proteomics of *Medicago truncatula* seed development establishes the time frame of diverse metabolic processes related to reserve accumulation," *Plant Physiology*, vol. 133, no. 2, pp. 664–682, 2003.

[17] B. Tivoli, A. Baranger, K. Sivasithamparam, and M. J. Barbetti, "Annual Medicago: from a model crop challenged by a spectrum of necrotrophic pathogens to a model plant to explore the nature of disease resistance," *Annals of Botany*, vol. 98, no. 6, pp. 1117–1128, 2006.

[18] F. Colditz, H.-P. Braun, C. Jacquet, K. Niehaus, and F. Krajinski, "Proteomic profiling unravels insights into the molecular background underlying increased *Aphanomyces euteiches* of *Medicago truncatula*," *Plant Molecular Biology*, vol. 59, no. 3, pp. 387–406, 2005.

[19] G. Vandemark and N. Grünwald, "Reaction of *Medicago truncatula* to *Aphanomyces euteiches* race 2," *Archives of Phytopathology and Plant Protection*, vol. 37, no. 1, pp. 59–67, 2004.

[20] O. Nyamsuren, F. Colditz, S. Rosendahl, et al., "Transcriptional profiling of *Medicago truncatula* roots after infection with *Aphanomyces euteiches* (oomycota) identifies novel genes upregulated during this pathogenic interaction," *Physiological and Molecular Plant Pathology*, vol. 63, no. 1, pp. 17–26, 2003.

[21] H. Zhu, S. B. Cannon, N. D. Young, and D. R. Cook, "Phylogeny and genomic organization of the TIR and non-TIR NBS-LRR resistance gene family in *Medicago truncatula*," *Molecular Plant-Microbe Interactions*, vol. 15, no. 6, pp. 529–539, 2002.

[22] F. Vailleau, E. Sartorel, M.-F. Jardinaud, et al., "Characterization of the interaction between the bacterial wilt pathogen Ralstonia solanacearum and the model legume plant *Medicago truncatula*," *Molecular Plant-Microbe Interactions*, vol. 20, no. 2, pp. 159–167, 2007.

[23] J. P. Klingler, O. R. Edwards, and K. B. Singh, "Independent action and contrasting phenotypes of resistance genes against spotted alfalfa aphid and bluegreen aphid in *Medicago truncatula*," *New Phytologist*, vol. 173, no. 3, pp. 630–640, 2007.

[24] L.-L. Gao, R. Horbury, R. M. Nair, K. B. Singh, and O. R. Edwards, "Characterization of resistance to multiple aphid species (Hemiptera: Aphididae) in *Medicago truncatula*," *Bulletin of Entomological Research*, vol. 97, no. 1, pp. 41–48, 2007.

[25] K. L. Korth, S. J. Doege, S.-H. Park, et al., "*Medicago truncatula* mutants demonstrate the role of plant calcium oxalate crystals as an effective defense against chewing insects," *Plant Physiology*, vol. 141, no. 1, pp. 188–195, 2006.

[26] J. C. Bede, R. O. Musser, G. W. Felton, and K. L. Korth, "Caterpillar herbivory and salivary enzymes decrease transcript levels of *Medicago truncatula* genes encoding early

enzymes in terpenoid biosynthesis," *Plant Molecular Biology*, vol. 60, no. 4, pp. 519–531, 2006.

[27] L.-L. Gao, J. P. Anderson, J. P. Klingler, R. M. Nair, O. R. Edwards, and K. B. Singh, "Involvement of the octadecanoid pathway in bluegreen aphid resistance in *Medicago truncatula*," *Molecular Plant-Microbe Interactions*, vol. 20, no. 1, pp. 82–93, 2007.

[28] R. Abranches, S. Marcel, E. Arcalis, F. Altmann, P. Fevereiro, and E. Stoger, "Plants as bioreactors: a comparative study suggests that *Medicago truncatula* is a promising production system," *Journal of Biotechnology*, vol. 120, no. 1, pp. 121–134, 2005.

[29] I. Bonnin, T. Huguet, M. Gherardi, J.-M. Prosperi, and I. Olivieri, "High level of polymorphism and spatial structure in a selfing plant species, *Medicago truncatula* (Leguminosae), shown using RAPD markers," *American Journal of Botany*, vol. 83, no. 7, pp. 843–855, 1996.

[30] J. Ronfort, T. Bataillon, S. Santoni, M. Delalande, J. L. David, and J.-M. Prosperi, "Microsatellite diversity and broad scale geographic structure in a model legume: building a set of nested core collection for studying naturally occurring variation in *Medicago truncatula*," *BMC Plant Biology*, vol. 6, 2006.

[31] F. Galibert, T. M. Finan, S. R. Long, et al., "The composite genome of the legume symbiont *Sinorhizobium meliloti*," *Science*, vol. 293, no. 5530, pp. 668–672, 2001.

[32] C. Mougel, P. Offre, L. Ranjard, et al., "Dynamic of the genetic structure of bacterial and fungal communities at different developmental stages of *Medicago truncatula* Gaertn. cv. Jemalong line J5," *New Phytologist*, vol. 170, no. 1, pp. 165–175, 2006.

[33] M. Sagan, D. Morandi, E. Tarenghi, and G. Duc, "Selection of nodulation and mycorrhizal mutants in the model plant *Medicago truncatula* (Gaertn.) after γ-ray mutagenesis," *Plant Science*, vol. 111, no. 1, pp. 63–71, 1995.

[34] V. Benaben, G. Duc, V. Lefebvre, and T. Huguet, "TE7, an inefficient symbiotic mutant of *Medicago truncatula* Gaertn. cv Jemalong," *Plant Physiology*, vol. 107, no. 1, pp. 53–62, 1995.

[35] J. M. Ané, G. B. Kiss, B. K. Riely, et al., "Medicago *truncatula DMI1* required for bacterial and fungal symbioses in legumes," *Science*, vol. 303, no. 5662, pp. 1364–1367, 2004.

[36] L. J. Bright, Y. Liang, D. M. Mitchell, and J. M. Harris, "The *LATD* gene of *Medicago truncatula* is required for both nodule and root development," *Molecular Plant-Microbe Interactions*, vol. 18, no. 6, pp. 521–532, 2005.

[37] R. V. Penmetsa and D. R. Cook, "Production and characterization of diverse developmental mutants of *Medicago truncatula*," *Plant Physiology*, vol. 123, no. 4, pp. 1387–1398, 2000.

[38] M. Scholte, I. D'Erfurth, S. Rippa, et al., "T-DNA tagging in the model legume *Medicago truncatula* allows efficient gene discovery," *Molecular Breeding*, vol. 10, no. 4, pp. 203–215, 2002.

[39] I. D'Erfurth, V. Cosson, A. Eschstruth, H. Lucas, A. Kondorosi, and P. Ratet, "Efficient transposition of the *Tnt1* tobacco retrotransposon in the model legume *Medicago truncatula*," *Plant Journal*, vol. 34, no. 1, pp. 95–106, 2003.

[40] R. Benlloch, I. D'Erfurth, C. Ferrandiz, et al., "Isolation of mtpim proves *Tnt1* a useful reverse genetics tool in *Medicago truncatula* and uncovers new aspects of *AP1* functions in legumes," *Plant Physiology*, vol. 142, no. 3, pp. 972–983, 2006.

[41] C. Crane, E. Wright, R. A. Dixon, and Z.-Y. Wang, "Transgenic *Medicago truncatula* plants obtained from *Agrobacterium tumefaciens* roots and *Agrobacterium rhizogenes* hairy roots," *Planta*, vol. 223, no. 6, pp. 1344–1354, 2006.

[42] X. Zhou, M. B. Chandrasekharan, and T. C. Hall, "High rooting frequency and functional analysis of GUS and GFP expression in transgenic *Medicago truncatula* A17," *New Phytologist*, vol. 162, no. 3, pp. 813–822, 2004.

[43] M. Chabaud, F. de Carvalho-Niebel, and D. G. Barker, "Efficient transformation of *Medicago truncatula* cv. Jemalong using the hypervirulent Agrobacterium tumefaciens strain AGL1," *Plant Cell Reports*, vol. 22, no. 1, pp. 46–51, 2003.

[44] K. Kamaté, I. D. Rodriguez-Llorente, M. Scholte, et al., "Transformation of floral organs with GFP in *Medicago truncatula*," *Plant Cell Reports*, vol. 19, no. 7, pp. 647–653, 2000.

[45] T. H. Trinh, P. Ratet, E. Kondorosi, et al., "Rapid and efficient transformation of diploid *Medicago truncatula* and Medicago sativa ssp. falcata lines improved in somatic embryogenesis," *Plant Cell Reports*, vol. 17, no. 5, pp. 345–355, 1998.

[46] M. Tadege, P. Ratet, and K. S. Mysore, "Insertional mutagenesis: a Swiss Army knife for functional genomics of *Medicago truncatula*," *Trends in Plant Science*, vol. 10, no. 5, pp. 229–235, 2005.

[47] A. Boisson-Dernier, M. Chabaud, F. Garcia, G. Bécard, C. Rosenberg, and D. G. Barker, "*Agrobacterium rhizogenes* roots of *Medicago truncatula* for the study of nitrogen-fixing and endomycorrhizal symbiotic associations," *Molecular Plant-Microbe Interactions*, vol. 14, no. 6, pp. 695–700, 2001.

[48] E. Limpens, J. Ramos, C. Franken, et al., "RNA interference in *Agrobacterium rhizogenes* roots of *Arabidopsis* and *Medicago truncatula*," *Journal of Experimental Botany*, vol. 55, no. 399, pp. 983–992, 2004.

[49] B. K. Riely, G. Lougnon, J. M. Ané, and D. R. Cook, "The symbiotic ion channel homolog DMI1 is localized in the nuclear membrane of *Medicago truncatula* roots," *Plant Journal*, vol. 49, no. 2, pp. 208–216, 2007.

[50] X. Huo, E. Schnabel, K. Hughes, and J. Frugoli, "RNAi phenotypes and the localization of a protein::GUS Fusion imply a role for *Medicago truncatulaPIN* genes in nodulation," *Journal of Plant Growth Regulation*, vol. 25, no. 2, pp. 156–165, 2006.

[51] R. Collier, B. Fuchs, N. Walter, W. K. Lutke, and C. G. Taylor, "*Ex vitro* composite plants: an inexpensive, rapid method for root biology," *Plant Journal*, vol. 43, no. 3, pp. 449–457, 2005.

[52] J.-H. Mun, D.-J. Kim, H.-K. Choi, et al., "Distribution of microsatellites in the genome of *Medicago truncatula*: a resource of genetic markers that integrate genetic and physical maps," *Genetics*, vol. 172, no. 4, pp. 2541–2555, 2006.

[53] P. Thoquet, M. Ghérardi, E.-P. Journet, et al., "The molecular genetic linkage map of the model legume *Medicago truncatula*: an essential tool for comparative legume genomics and the isolation of agronomically important genes," *BMC Plant Biology*, vol. 2, 2002.

[54] H.-K. Choi, D. Kim, T. Uhm, et al., "A sequence-based genetic map of *Medicago truncatula* and comparison of marker colinearity with *M. sativa*," *Genetics*, vol. 166, no. 3, pp. 1463–1502, 2004.

[55] M. Abirached-Darmency, E. Prado-Vivant, L. Chelysheva, and T. Pouthier, "Variation in rDNA locus number and position among legume species and detection of 2 linked rDNA loci in the model *Medicago truncatula* by FISH," *Genome*, vol. 48, no. 3, pp. 556–561, 2005.

[56] M. Cerbah, Z. Kevei, S. Siljak-Yakovlev, E. Kondorosi, A. Kondorosi, and T. H. Trinh, "FISH chromosome mapping

allowing karyotype analysis in *Medicago truncatula* lines Jemalong J5 and R-108-1," *Molecular Plant-Microbe Interactions*, vol. 12, no. 11, pp. 947–950, 1999.

[57] J. M. Ané, J. Lévy, P. Thoquet, et al., "Genetic and cytogenetic mapping of *DMI1*, *DMI2*, and *DMI3* genes of *Medicago truncatula* involved in Nod factor transduction, nodulation, and mycorrhization," *Molecular Plant-Microbe Interactions*, vol. 15, no. 11, pp. 1108–1118, 2002.

[58] O. Kulikova, G. Gualtieri, R. Geurts, et al., "Integration of the FISH pachytene and genetic maps of *Medicago truncatula*," *Plant Journal*, vol. 27, no. 1, pp. 49–58, 2001.

[59] O. Kulikova, R. Geurts, M. Lamine, et al., "Satellite repeats in the functional centromere and pericentromeric heterochromatin of *Medicago truncatula*," *Chromosoma*, vol. 113, no. 6, pp. 276–283, 2004.

[60] E. Schnabel, O. Kulikova, R. V. Penmetsa, T. Bisseling, D. R. Cook, and J. Frugoli, "An integrated physical, genetic and cytogenetic map around the *sunn* locus of *Medicago truncatula*," *Genome*, vol. 46, no. 4, pp. 665–672, 2003.

[61] F. W. Engler, J. Hatfield, W. Nelson, and C. A. Soderlund, "Locating sequence on FPC maps and selecting a minimal tiling path," *Genome Research*, vol. 13, no. 9, pp. 2152–2163, 2003.

[62] N. D. Young, S. B. Cannon, S. Sato, et al., "Sequencing the genespaces of *Medicago truncatula* and *Lotus japonicus*," *Plant Physiology*, vol. 137, no. 4, pp. 1174–1181, 2005.

[63] C. D. Town, "Annotating the genome of *Medicago truncatula*," *Current Opinion in Plant Biology*, vol. 9, no. 2, pp. 122–127, 2006.

[64] V. Stolc, M. P. Samanta, and W. Tongprasit, "Identification of transcribed sequences in *Arabidopsis thaliana* by using high-resolution genone tiling arrays," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 12, pp. 4453–4458, 2005.

[65] J. M. Johnson, S. Edwards, D. Shoemaker, and E. E. Schadt, "Dark matter in the genome: evidence of widespread transcription detected by microarray tiling experiments," *Trends in Genetics*, vol. 21, no. 2, pp. 93–102, 2005.

[66] K. Yamada, J. Lim, J. H. Dale, et al., "Empirical analysis of transcriptional activity in the *Arabidopsis Genome*," *Science*, vol. 302, no. 5646, pp. 842–846, 2003.

[67] P. Bertone, M. Gerstein, and M. Snyder, "Applications of DNA tiling arrays to experimental genome annotation and regulatory pathway discovery," *Chromosome Research*, vol. 13, no. 3, pp. 259–274, 2005.

[68] P. A. Covitz, L. S. Smith, and S. R. Long, "Expressed sequence tags from a root-hair-enriched *Medicago truncatula* cDNA library," *Plant Physiology*, vol. 117, no. 4, pp. 1325–1332, 1998.

[69] J. Györgyey, D. Vaubert, J. I. Jiménez-Zurdo, et al., "Analysis of *Medicago truncatula* nodule expressed sequence tags," *Molecular Plant-Microbe Interactions*, vol. 13, no. 1, pp. 62–71, 2000.

[70] C. J. Bell, R. A. Dixon, A. D. Farmer, et al., "The *Medicago* genome initiative: a model legume database," *Nucleic Acids Research*, vol. 29, no. 1, pp. 114–117, 2001.

[71] E.-P. Journet, D. van Tuinen, J. Gouzy, et al., "Exploring root symbiotic programs in the model legume *Medicago truncatula* using EST analysis," *Nucleic Acids Research*, vol. 30, no. 24, pp. 5579–5592, 2002.

[72] A.-F. J. Lamblin, J. A. Crow, J. E. Johnson, et al., "MtDB: a database for personalized data mining of the model legume *Medicago truncatula* transcriptome," *Nucleic Acids Research*, vol. 31, no. 1, pp. 196–201, 2003.

[73] J. Liu, L. A. Blaylock, G. Endre, et al., "Transcript profiling coupled with spatial expression analyses reveals genes involved in distinct developmental stages of an arbuscular mycorrhizal symbiosis," *Plant Cell*, vol. 15, no. 9, pp. 2106–2123, 2003.

[74] F. E. Yahyaoui, H. Küster, B. B. Amor, et al., "Expression profiling in *Medicago truncatula* identifies more than 750 genes differentially expressed during nodulation, including many potential regulators of the symbiotic program," *Plant Physiology*, vol. 136, no. 2, pp. 3159–3176, 2004.

[75] K. Manthey, F. Krajinski, N. Hohnjec, et al., "Transcriptome profiling in root nodules and arbuscular mycorrhiza identifies a collection of novel genes induced during *Medicago truncatula* root endosymbioses," *Molecular Plant-Microbe Interactions*, vol. 17, no. 10, pp. 1063–1077, 2004.

[76] R. M. Mitra and S. R. Long, "Plant and bacterial symbiotic mutants define three transcriptionally distinct stages in the development of the *Medicago truncatula/Sinorhizobium meliloti* Symbiosis," *Plant Physiology*, vol. 134, no. 2, pp. 595–604, 2004.

[77] R. M. Mitra, S. L. Shaw, and S. R. Long, "Six nonnodulating plant mutants defective for Nod factor-induced transcriptional changes associated with the legume-rhizobia symbiosis," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 27, pp. 10217–10222, 2004.

[78] D. P. Lohar, N. Sharopova, G. Endre, et al., "Transcript analysis of early nodulation events in *Medicago truncatula*," *Plant Physiology*, vol. 140, no. 1, pp. 221–234, 2006.

[79] C. G. Starker, A. L. Parra-Colmenares, L. Smith, R. M. Mitra, and S. R. Long, "Nitrogen fixation mutants of *Medicago truncatula* fail to support plant and bacterial symbiotic gene expression," *Plant Physiology*, vol. 140, no. 2, pp. 671–680, 2006.

[80] M. J. Barnett, C. J. Toman, R. F. Fisher, and S. R. Long, "A dual-genome Symbiosis Chip for coordinate study of signal exchange and development in a prokaryote-host interaction," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 47, pp. 16636–16641, 2004.

[81] N. Aziz, N. L. Paiva, G. D. May, and R. A. Dixon, "Transcriptome analysis of alfalfa glandular trichomes," *Planta*, vol. 221, no. 1, pp. 28–38, 2005.

[82] B. E. Deavours and R. A. Dixon, "Metabolic engineering of isoflavonoid biosynthesis in alfalfa," *Plant Physiology*, vol. 138, no. 4, pp. 2245–2259, 2005.

[83] L. Diatchenko, Y.-F. C. Lau, A. P. Campbell, et al., "Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 93, no. 12, pp. 6025–6030, 1996.

[84] A. Wulf, K. Manthey, J. Doll, et al., "Transcriptional changes in response to arbuscular mycorrhiza development in the model plant *Medicago truncatula*," *Molecular Plant-Microbe Interactions*, vol. 16, no. 4, pp. 306–314, 2003.

[85] L. Godiard, A. Niebel, F. Micheli, J. Gouzy, T. Ott, and P. Gamas, "Identification of new potential regulators of the *Medicago truncatula-Sinorhizobium meliloti* symbiosis using a large-scale suppression subtractive hybridization approach," *Molecular Plant-Microbe Interactions*, vol. 20, no. 3, pp. 321–332, 2007.

[86] V. E. Velculescu, L. Zhang, B. Vogelstein, and K. W. Kinzler, "Serial analysis of gene expression," *Science*, vol. 270, no. 5235, pp. 484–487, 1995.

[87] G. K. Agrawal, M. Yonekura, Y. Iwahashi, H. Iwahashi, and R. Rakwal, "System, trends and perspectives of proteomics in dicot plants—part III: unraveling the proteomes influenced by the environment, and at the levels of function and genetic relationships," *Journal of Chromatography B*, vol. 815, no. 1-2, pp. 137–145, 2005.

[88] G. K. Agrawal, M. Yonekura, Y. Iwahashi, H. Iwahashi, and R. Rakwal, "System, trends and perspectives of proteomics in dicot plants—part II: proteomes of the complex developmental stages," *Journal of Chromatography B*, vol. 815, no. 1-2, pp. 125–136, 2005.

[89] K. Zhang, C. McKinlay, C. H. Hocart, and M. A. Djordjevic, "The *Medicago truncatula* small protein proteome and peptidome," *Journal of Proteome Research*, vol. 5, no. 12, pp. 3355–3367, 2006.

[90] F. Colditz, O. Nyamsuren, K. Niehaus, H. Eubel, H.-P. Braun, and F. Krajinski, "Proteomic approach: identification of *Medicago truncatula* proteins induced in roots after infection with the pathogenic oomycete *Aphanomyces euteiches*," *Plant Molecular Biology*, vol. 55, no. 1, pp. 109–120, 2004.

[91] C. M. Catalano, W. S. Lane, and D. J. Sherrier, "Biochemical characterization of symbiosome membrane proteins from *Medicago truncatula* root nodules," *Electrophoresis*, vol. 25, no. 3, pp. 519–531, 2004.

[92] B. Valot, M. Dieu, G. Recorbet, M. Raes, S. Gianinazzi, and E. Dumas-Gaudot, "Identification of membrane-associated proteins regulated by the arbuscular mycorrhizal symbiosis," *Plant Molecular Biology*, vol. 59, no. 4, pp. 565–580, 2005.

[93] B. Valot, S. Gianinazzi, and D.-G. Eliane, "Sub-cellular proteomic analysis of a *Medicago truncatula* root microsomal fraction," *Phytochemistry*, vol. 65, no. 12, pp. 1721–1732, 2004.

[94] U. Mathesius, G. Keijzers, S. H. A. Natera, J. J. Weinman, M. A. Djordjevic, and B. G. Rolfe, "Establishment of a root proteome reference map for the model legume *Medicago truncatula* using the expressed sequence tag database for peptide mass fingerprinting," *Proteomics*, vol. 1, no. 11, pp. 1424–1440, 2001.

[95] B. S. Watson, V. S. Asirvatham, L. Wang, and L. W. Sumner, "Mapping the proteome of barrel medic (*Medicago truncatula*)," *Plant Physiology*, vol. 131, no. 3, pp. 1104–1123, 2003.

[96] N. Imin, F. De Jong, U. Mathesius, et al., "Proteome reference maps of *Medicago truncatula* embryogenic cell cultures generated from single protoplasts," *Proteomics*, vol. 4, no. 7, pp. 1883–1896, 2004.

[97] N. Imin, M. Nizamidin, D. Daniher, K. E. Nolan, R. J. Rose, and B. G. Rolfe, "Proteomic analysis of somatic embryogenesis in *Medicago truncatula*. Explant cultures grown under 6-benzylaminopurine and 1-naphthaleneacetic acid treatments," *Plant Physiology*, vol. 137, no. 4, pp. 1250–1260, 2005.

[98] Z. Lei, A. M. Elmer, B. S. Watson, R. A. Dixon, P. J. Mendes, and L. W. Sumner, "A two-dimensional electrophoresis proteomic reference map and systematic identification of 1367 proteins from a cell suspension culture of the model legume *Medicago truncatula*," *Molecular and Cellular Proteomics*, vol. 4, no. 11, pp. 1812–1825, 2005.

[99] N. Amiour, G. Recorbet, F. Robert, S. Gianinazzi, and E. Dumas-Gaudot, "Mutations in *DMI3* and *SUNN* modify the appressorium-responsive root proteome in arbuscular mycorrhiza," *Molecular Plant-Microbe Interactions*, vol. 19, no. 9, pp. 988–997, 2006.

[100] J. Prayitno, B. G. Rolfe, and U. Mathesius, "The ethylene-insensitive *sickle* mutant of *Medicago truncatula* shows altered auxin transport regulation during nodulation," *Plant Physiology*, vol. 142, no. 1, pp. 168–180, 2006.

[101] M. K. Udvardi and W.-R. Scheible, "GRAS genes and the symbiotic green revolution," *Science*, vol. 308, no. 5729, pp. 1749–1750, 2005.

[102] S. Laugesen, E. Messinese, S. Hem, et al., "Phosphoproteins analysis in plants: a proteomic approach," *Phytochemistry*, vol. 67, no. 20, pp. 2208–2214, 2006.

[103] R. N. Trethewey, A. J. Krotzky, and L. Willmitzer, "Metabolic profiling: a rosetta stone for genomics?" *Current Opinion in Plant Biology*, vol. 2, no. 2, pp. 83–85, 1999.

[104] C. D. Broeckling, D. V. Huhman, M. A. Farag, et al., "Metabolic profiling of *Medicago truncatula* cell cultures reveals the effects of biotic and abiotic elicitors on metabolism," *Journal of Experimental Botany*, vol. 56, no. 410, pp. 323–336, 2005.

[105] H. Suzuki, M. S.S. Reddy, M. Naoumkina, et al., "Methyl jasmonate and yeast elicitor induce differential transcriptional and metabolic re-programming in cell suspension cultures of the model legume *Medicago truncatula*," *Planta*, vol. 220, no. 5, pp. 696–707, 2005.

[106] H. Jenkins, N. Hardy, M. Beckmann, et al., "A proposed framework for the description of plant metabolomics experiments and their results," *Nature Biotechnology*, vol. 22, no. 12, pp. 1601–1606, 2004.

[107] H. Jenkins, H. Johnson, B. Kular, T. Wang, and N. Hardy, "Toward supportive data collection tools for plant metabolomics," *Plant Physiology*, vol. 138, no. 1, pp. 67–77, 2005.

[108] J. Bowman, *Arabidopsis: An Atlas of Morphology and Development*, Springer, New York, NY, USA, 1994.

[109] D. Moreau, C. Salon, and N. Munier-Jolain, "A framework for the analysis of genetic variability in *Medicago truncatula*," in *Proceedings of the 2nd International Conference on Legume Genomics and Genetics*, p. 286, Dijon, France, June 2004.

[110] M. D. Gonzales, E. Archuleta, A. Farmer, et al., "The legume information system (LIS): an integrated information resource for comparative legume biology," *Nucleic Acids Research*, vol. 33, pp. D660–D665, 2005.

[111] M. Fedorova, J. van de Mortel, P. A. Matsumoto, et al., "Genome-wide identification of nodule-specific transcripts in the model legume *Medicago truncatula*," *Plant Physiology*, vol. 130, no. 2, pp. 519–537, 2002.

[112] P. Mergaert, K. Nikovics, Z. Kelemen, et al., "A novel family in *Medicago truncatula* consisting of more than 300 nodule-specific genes coding for small, secreted polypeptides with conserved cysteine motifs," *Plant Physiology*, vol. 132, no. 1, pp. 161–173, 2003.

[113] M. A. Graham, K. A.T. Silverstein, S. B. Cannon, and K. A. VandenBosch, "Computational identification and characterization of novel genes from legumes," *Plant Physiology*, vol. 135, no. 3, pp. 1179–1197, 2004.

[114] H. Küster, A. Becker, C. Firnhaber, et al., "Development of bioinformatic tools to support EST-sequencing, *in silico* and microarray-based transcriptome profiling in mycorrhizal symbioses," *Phytochemistry*, vol. 68, no. 1, pp. 19–32, 2007.

[115] S. B. Cannon, J. A. Crow, M. L. Heuer, et al., "Databases and information integration for the *Medicago truncatula* genome and transcriptome," *Plant Physiology*, vol. 138, no. 1, pp. 38–46, 2005.

[116] E. Urbanczyk-Wochniak and L. W. Sumner, "MedicCyc: a biochemical pathway database for *Medicago truncatula*," *Bioinformatics*, vol. 23, no. 11, pp. 1418–1423, 2007.

[117] J. Liu, S. S. Miller, M. Graham, et al., "Recruitment of novel calcium-binding proteins for root nodule symbiosis in *Medicago truncatula*," *Plant Physiology*, vol. 141, no. 1, pp. 167–177, 2006.

[118] S. Henikoff, T. Furuyama, and K. Ahmad, "Histone variants, nucleosome assembly and epigenetic inheritance," *Trends in Genetics*, vol. 20, no. 7, pp. 320–326, 2004.

[119] H. Javot, R. V. Penmetsa, N. Terzaghi, D. R. Cook, and M. J. Harrison, "A *Medicago truncatula* phosphate transporter indispensable for the arbuscular mycorrhizal symbiosis," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 5, pp. 1720–1725, 2007.

[120] X. Li, Y. Song, K. Century, et al., "A fast neutron deletion mutagenesis-based reverse genetics system for plants," *Plant Journal*, vol. 27, no. 3, pp. 235–242, 2001.

[121] T. Colbert, B. J. Till, R. Tompa, et al., "High-throughput screening for induced point mutations," *Plant Physiology*, vol. 126, no. 2, pp. 480–484, 2001.

[122] C. M. McCallum, L. Comai, E. A. Greene, and S. Henikoff, "Targeting induced local lesions IN genomes (TILLING) for plant functional genomics," *Plant Physiology*, vol. 123, no. 2, pp. 439–442, 2000.

[123] C. M. McCallum, L. Comai, E. A. Greene, and S. Henikoff, "Targeted screening for induced mutations," *Nature Biotechnology*, vol. 18, no. 4, pp. 455–457, 2000.

[124] H.-K. Choi, J.-H. Mun, D.-J. Kim, et al., "Estimating genome conservation between crop and model legume species," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 43, pp. 15289–15294, 2004.

[125] H. Zhu, H.-K. Choi, D. R. Cook, and R. C. Shoemaker, "Bridging model and crop legumes through comparative genomics," *Plant Physiology*, vol. 137, no. 4, pp. 1189–1196, 2005.

[126] H. Zhu, B. K. Riely, N. J. Burns, and J. M. Ané, "Tracing nonlegume orthologs of legume genes required for nodulation and arbuscular mycorrhizal symbioses," *Genetics*, vol. 172, no. 4, pp. 2491–2499, 2006.

[127] S. B. Cannon, L. Sterck, S. Rombauts, et al., "Legume genome evolution viewed through the *Medicago truncatula* and *Lotus japonicus* genomes," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 40, pp. 14959–14964, 2006.

[128] G. Endre, P. Kaló, Z. Kevei, et al., "Genetic mapping of the non-nodulation phenotype of the mutant MN-1008 in tetraploid alfalfa (*Medicago sativa*)," *Molecular Genetics and Genomics*, vol. 266, no. 6, pp. 1012–1019, 2001.

[129] E. Limpens, C. Franken, P. Smit, J. Willemse, T. Bisseling, and R. Geurts, "LysM domain receptor kinases regulating rhizobial nod factor-induced infection," *Science*, vol. 302, no. 5645, pp. 630–633, 2003.

[130] P. Gepts, W. D. Beavis, E. C. Brummer, et al., "Legumes as a model plant family. Genomics for food and feed report of the cross-legume advances through genomics conference," *Plant Physiology*, vol. 137, no. 4, pp. 1228–1235, 2005.

[131] S. H. Hulbert, C. A. Webb, S. M. Smith, and Q. Sun, "Resistance gene complexes: evolution and utilization," *Annual Review of Phytopathology*, vol. 39, pp. 285–312, 2001.

[132] S. B. Cannon, H. Zhu, A. M. Baumgarten, et al., "Diversity, distribution, and ancient taxonomic relationships within the TIR and non-TIR NBS-LRR resistance gene subfamilies," *Journal of Molecular Evolution*, vol. 54, no. 4, pp. 548–562, 2002.

Hindawi Publishing Corporation
International Journal of Plant Genomics
Volume 2008, Article ID 582837, 9 pages
doi:10.1155/2008/582837

Review Article

Progress in Understanding and Sequencing the Genome of *Brassica rapa*

Chang Pyo Hong,[1] Soo-Jin Kwon,[2] Jung Sun Kim,[2] Tae-Jin Yang,[3] Beom-Seok Park,[2] and Yong Pyo Lim[1]

[1] *Department of Horticulture, College of Agriculture and Life Science, Chungnam National University, Daejeon 305764, South Korea*
[2] *Brassica Genomics Team, National Institute of Agricultural Biotechnology (NIAB), Rural Development Administration (RDA), Suwon 441707, South Korea*
[3] *Department of Plant Science, College of Agriculture and Life Sciences, Seoul National University, Seoul 151921, South Korea*

Correspondence should be addressed to Yong Pyo Lim, yplim@cnu.ac.kr

Received 13 June 2007; Accepted 21 November 2007

Recommended by Yunbi Xu

Brassica rapa, which is closely related to *Arabidopsis thaliana*, is an important crop and a model plant for studying genome evolution via polyploidization. We report the current understanding of the genome structure of *B. rapa* and efforts for the whole-genome sequencing of the species. The tribe *Brassicaceae*, which comprises ca. 240 species, descended from a common hexaploid ancestor with a basic genome similar to that of *Arabidopsis*. Chromosome rearrangements, including fusions and/or fissions, resulted in the present-day "diploid" *Brassica* species with variation in chromosome number and phenotype. Triplicated genomic segments of *B. rapa* are collinear to those of *A. thaliana* with InDels. The genome triplication has led to an approximately 1.7-fold increase in the *B. rapa* gene number compared to that of *A. thaliana*. Repetitive DNA of *B. rapa* has also been extensively amplified and has diverged from that of *A. thaliana*. For its whole-genome sequencing, the *Brassica rapa* Genome Sequencing Project (*Br*GSP) consortium has developed suitable genomic resources and constructed genetic and physical maps. Ten chromosomes of *B. rapa* are being allocated to *Br*GSP consortium participants, and each chromosome will be sequenced by a BAC-by-BAC approach. Genome sequencing of *B. rapa* will offer a new perspective for plant biology and evolution in the context of polyploidization.

1. IMPORTANCE OF *BRASSICA* GENOMICS

The genus *Brassica* is one of the core genera in the tribe *Brassicaceae* and includes a number of crops with wide adaptation under a variety of agroclimatic conditions. Economically, *Brassica* species are important sources of vegetable oil, fresh, preserved vegetables, and condiments. *B. napus, B. rapa*, *B. juncea,* and *B. carinata* provide about 12% of the worldwide edible vegetable oil supply [1]. The *B. rapa* and *B. oleracea* subspecies represent many of the vegetables in our daily diet. In particular, *B. rapa* ssp. *pekinensis* (Chinese cabbage), on which this article focuses, is one of the most widely used vegetable crops in northeast Asia. Moreover, *Brassica* species are important sources of dietary fiber, vitamin C, and anticancer compounds [2].

The genetic relationships among the different diploid and amphidiploid *Brassica* species are described by the U's triangle [3]. Of the six widely cultivated species of *Brassica*, *B. rapa* (AA, $2n = 20$), *B. nigra* (BB, $2n = 16$), and *B. oleracea* (CC, $2n = 18$) are monogenomic diploids. The remaining three species, *B. juncea* (AABB, $2n = 36$), *B. napus* (AACC, $2n = 38$), and *B. carinata* (BBCC, $2n = 34$) exhibit stable diploid genetics, but are allotetraploids, which have evolved via hybridization between differing monogenomic diploids [3]. The diploid *Brassica* genomes range from 1.1 pg/2C (529 Mbp/1C) for *B. rapa* to 1.4 pg/2C (696 Mbp/1C) for *B. oleracea* (see Figure 1) [4]. The genomes of the allotetraploids range from 2.2 pg/2C (1,068 Mbp/1C) for *B. juncea* to 2.6 pg/2C (1,284 Mbp/1C) for *B. carinata* (see Figure 1).

The genus *Brassica* is characterized by morphological diversity with regard to inflorescences, leaves, stems, roots, and terminal or apical buds [5]. For example, such morphological diversity can be easily observed in subspecies of *B. oleracea*: the enlarged inflorescences of cauliflower (*B. oleracea* ssp. *botrytis*) and broccoli (*B. oleracea* ssp. *italica*); the enlarged stems of kohlrabi (*B. oleracea* ssp. *gongylodes*) and marrowstem kale (*B. oleracea* ssp. *medullosa*); and the many axillary buds of Brussels sprout (*B. oleracea* ssp. *gemmifera*)

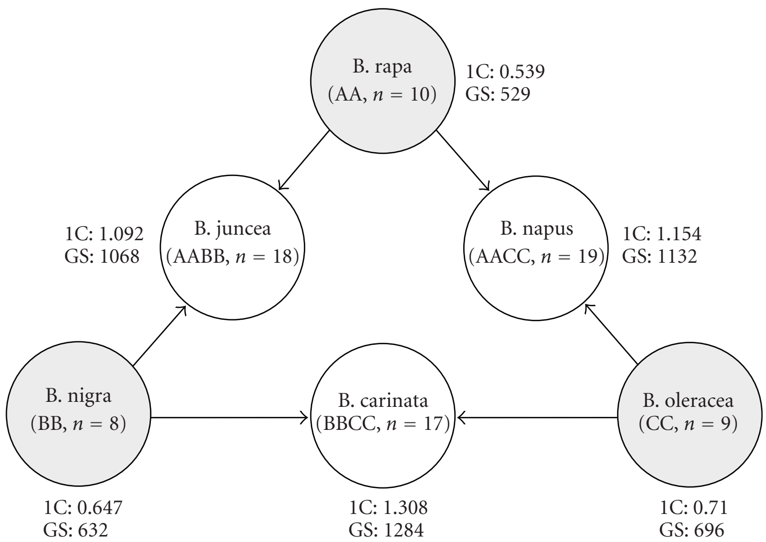


Figure 1: Genetic relationship of the different diploid and amphidiploid *Brassica* species. 1C, 1C nuclear DNA content (pg); GS, genome size (Mbp) [3, 4].

[5]. The morphological diversity in *Brassica* species may be linked to genomic changes associated with polyploidization [6]. The polyploidization in *Brassica* species has brought about triplication of genomic segments and subsequent rearrangements such as inversions, insertions, deletions, and substitutions [7–16], and these genetic variations may cause novel phenotypic variations for traits among these species [5, 6]. Thus, *Brassica* genomics will provide us with an understanding of the rapid phenotypic evolution of polyploid plants. Additionally, it will help us to understand genomic changes and how they shape the allotetrapolyploid *Brassica* species. For example, a study has been done looking at rapid genomic changes and the effect of nuclear-cytoplasm interaction in synthetic allotetrapolyploid species [17].

Because of the high economic value of *Brassica* species throughout the world and their potential to be models for the study of polyploidization, genome sequencing projects for *Brassica* species, especially *B. rapa* and *B. oleracea*, have recently been initiated (http://www.brassica.info) [18–20]. In particular, *B. rapa* ssp. *pekinensis* inbred line Chiffu-401-42, discussed in this article, has been selected for *Brassica*-A genome sequencing in the *Brassica rapa* Genome Sequencing Project (*Br*GSP) (http://www.brassica.info), a component of the consortium of the Multinational *Brassica* Genome Project, with the goal of completely sequencing this genome through a BAC-by-BAC approach. The *Br*GSP consortium has developed genomic resources for this purpose and is proceeding with whole-genome sequencing.

2. CURRENT UNDERSTANDING OF THE GENOME STRUCTURE OF *B. RAPA*

2.1. *Karyotype of B. rapa*

Karyotyping is the starting point for understanding the genome structure of a species. Moreover, it provides insight into genome evolution. Most of the karyotypic analyses in *B. rapa* have been performed on mitotic metaphase chromosomes [21–24]. However, the analyses are limited in what they can reveal about the cytological structure of the genome because of the low resolution of the technique. For example, different measurements of chromosome lengths and rDNA loci are obtained by this method. Recently, the high-resolution karyotype for the *B. rapa* ssp. *pekinensis* inbred line Chiffu was determined on pachytene chromosomes by using 4′-6-diamino-2-phenylindole dihydrochloride (DAPI) staining and fluorescence in situ hybridization (FISH) of rDNAs and pericentromeric satellite repeats [25]. By DAPI analysis, the mean lengths of ten pachytene chromosomes ranged from 23.7 μm to 51.3 μm, with a total of 385.3 μm, a total length which is 11.9- $\sim$ 17.5-fold longer than that of the mitotic metaphase chromosomes reported by Lim et al. [24] and Koo et al. [25]. In comparison, pachytene chromosome length of *A. thaliana*, *Medicago truncatula*, and tomato was estimated to be about 7.4%, 15%, and 24% of the total pachytene chromosome length, respectively (reviewed in Koo et al. [25]). In *B. rapa*, the pachytene karyotype consists of two metacentric (chromosomes 1 and 6), five submetacentric (chromosomes 3, 4, 5, 9, and 10), two subtelomeric (chromosomes 7 and 8), and one acrocentric chromosome (chromosome 2), with the corresponding centromeric index ranges of 38.8–41.0%, 29.5–36.7%, 17.4–20.2%, and 9.38%, respectively [25]. In the chromosomal structure at pachytene, the total length of pericentromeric heterochromatin regions was estimated to be 38.2 μm, which is approximately 10% of the total chromosome length [25]. In conjunction with chromosomal structure and characteristics, 5S rDNA loci were located on pericentromeric regions of the short arms of chromosomes 2 and 7 as well as the long arm of chromosome 10, while 45S rDNA loci were located on the short arms of chromosomes 1, 2, 4, and 5 as well as the long arm of chromosome 7 [24, 25].

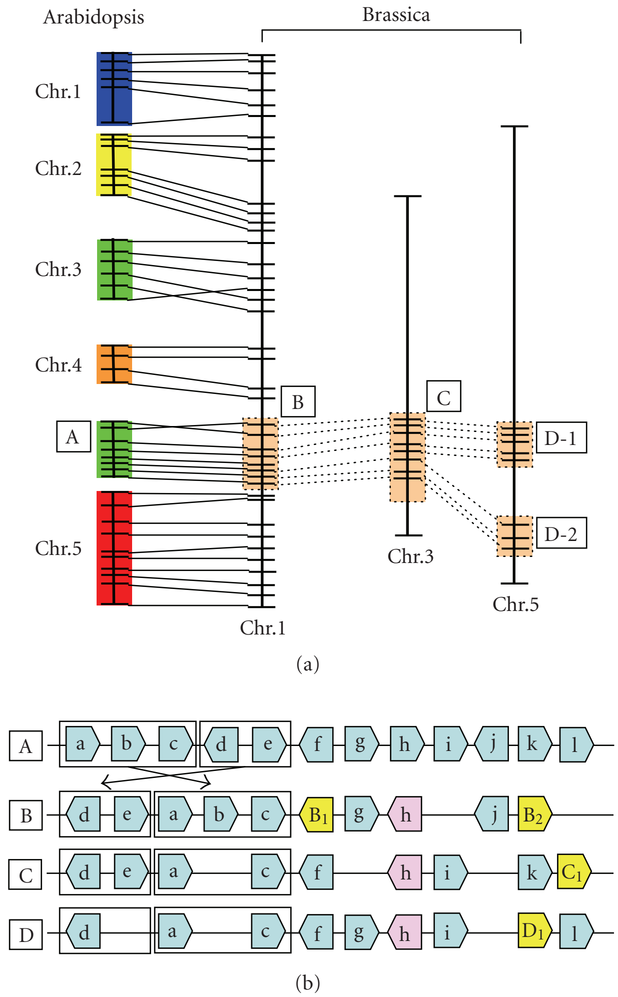


Figure 2: An example of a comparative map of *Arabidopsis* and *Brassica*. (a) Collinearity between genomic segments of the two species and genome triplication of *Brassica* revealed by comparative genetic mapping. (b) Synteny of genes in a triplicated genomic region of *Brassica*.

2.2. Collinearity between genomic segments of Arabidopsis and Brassica

Brassica species are closely related to *A. thaliana*, having diverged 17-18 million years ago (MYA) from their common ancestor [16]. *A. thaliana*, which has been completely sequenced, has a rather small genome (about 146 Mb) with relatively little repetitive DNA and a high gene density [26, 27]. Protein-coding regions of the genomes of *Brassica* species show high sequence conservation with those of *A. thaliana*, with nucleotide sequence similarity in exons between *B. oleracea* and *A. thaliana* ranging from 75% to 90%, compared to <70% for intronic regions [28]. This similarity allows the identification of sets of candidate genes in *Brassica* species and the studying of their genome structures through comparative genomics [29]. Comparative studies between *Arabidopsis* and *Brassica* have revealed the presence of collinear chromosome segments (see Figure 2). Comparative genetic mapping between diploid *Brassica* species and *A. thaliana* to identify homologous loci have revealed many conserved blocks in their genomes [7, 8, 14, 30]. Comparative physical mapping between *Arabidopsis* and *Brassica* further corroborated the findings. A set of six bacterial artificial chromosomes (BACs), representing a 431-kb contiguous region of *Arabidopsis* chromosome 2, was mapped on chromosomes and DNA fibers of *B. rapa* [31]. Moreover, studies on a 222-kb gene-rich region of *A. thaliana* chromosome 4 and its homologous counterparts in *B. rapa* or *B. oleracea* revealed the collinearity of genes in homologous segments [9, 11, 13]. This finding was supported by sequence analysis of specific homologous genomic segments [15, 16]. However, many structural rearrangements differentiate the *Brassica* and *Arabidopsis* chromosomes (see Figure 2). Comparative genetic mapping between *B. nigra* and *A. thaliana* species revealed that the average length of conserved segments between the two species was estimated at about 8 cM, which corresponds to ~90 rearrangements since the divergence of the two species [7]. In addition, it was found that gene contents in their homologous genomic segments were also variable with interstitial gene losses and insertions [9, 11, 13, 15, 16].

2.3. Genome triplication of diploid Brassica species

Most of the comparative studies mentioned above demonstrated that *Brassica* species contain extensively triplicated counterparts of the corresponding homologous segments of the *A. thaliana* genome (see Figure 2), thereby suggesting that diploid *Brassica* species may have been derived from a hexaploid ancestor: the genome which was similar to *Arabidopsis*. Consistent with the nature of genome triplication, Yang et al. [16] reported that paralogous subgenomes of diploid *Brassica* species triplicated 13 ~ 17 MYA, very soon after the *Arabidopsis* and *Brassica* divergence that occurred at 17 ~ 18 MYA. In addition, it was reported that after the *Brassica* genomes had triplicated, their subgenomes were rearranged by inversions, translocations [7, 12, 32], extensive interspersed gene loss, as well as gene insertions occurred relative to the inferred structure of the ancestral genome (see Figure 2). Additionally, such genome triplication was extensively found across the tribe *Brassicaceae* [12]. In comparison with the genome of *A. thaliana*, the genome triplication in *Brassica* species has clearly led to an increase in the genome size, resulting in a 3- to 5-fold inflation.

Genome triplication events in *Brassica* species may also have an effect on gene expression of multicopy genes, leading to such phenomena as pseudogenization, subfunctionalization, or neofunctionalization in species [33–38]. For example, the MADS-box transcription factor family, whose members control key aspects of plant vegetative and reproductive development, shapes genetic systems by subfunctionalization [37]. It appears that after polyploid formation, considerable and sometimes very rapid changes in genome structure and gene expression have occurred. Researchers have hypothesized that genomic triplication in *Brassica* species permits mutations in loci that are normally under tight selective constraints in *Arabidopsis*, and may thus result in the observed

Table 1: Comparison of gene, TE, and SSR abundances in *B. rapa* and *A. thaliana*.

Contents	*B. rapa*	*A. thaliana*	References
Genome size (Mb)	529	146	[4, 27]
Gene number	4300 ~ 63000	26,207	[16, 39, 43]
TE abundance (%)[1]	13.8	6 ~ 7	[39, 44]
SSR number	≈110,000 (one SSR/4.8 kb)	≈36,756 (one SSR/3.2 kb)	[39, 45]

[1] Coverage of TEs in the genome.

greater phenotypic plasticity in *Brassica* [5]. Studies on expression of duplicated genes in *Brassica* species will provide insight into the role of polyploidization in the phenotypic divergence of the plant genus.

2.4. Survey of the B. rapa genome revealed by BAC-end sequence analysis

The *B. rapa* genome was surveyed via the analysis of its 12,017 *Hin*dIII BAC-end sequences (Table 1) [39]. Analyses of BAC-end sequence or genome survey sequences assist in understanding whole genome structure [39–41]. It was estimated that the *B. rapa* genome might contain about 43000 genes (covering 16.8% of the genome), 1.6 times more than the *A. thaliana* genome. Recently, Yang et al. [16] also estimated the gene content of *B. rapa* to range from 49,000 to 63,000, based on predictions from microsynteny studies. It has been suggested that chromosomal triplication events in *Brassica* have led to an increase in gene number with subsequent gene loss [15, 16, 39, 42].

Transposable elements (TEs) with a predominance of retrotransposons were estimated to occupy approximately 14% of the genome (covering approximately 74 Mb), 8.2 times greater than that observed previously in *A. thaliana* [44]. Zhang and Wessler [44] reported that TEs in *B. oleracea* constituted 20% of the genome, slightly more than what was predicted for the *B. rapa* genome. Of the predicted TEs, LTR retrotransposon families were the most abundant (69.9%), followed by non-LTR retrotransposons (13.4%), DNA transposons (11.4%), and other retrotransposons (5.3%). In particular, *Ty1/copia*-like and *Ty3/gypsy*-like retrotransposons occupied 39.5% and 30.2% of LTR retrotransposon families, respectively. The amplification of TEs in *B. rapa*, especially retrotransposons, may have played a crucial role in both evolution and genomic expansion.

Simple sequence repeats (SSRs) have been estimated to occur with a frequency of approximately one per 4.8 kb within the *B. rapa* genome, as compared to approximately one per 3.2 kb within the *A. thaliana* genome [39]. Of SSRs identified, trinucleotides were the most abundant repeat type, constituting about 37% of all SSRs, a percentage similar to those reported in other plant genomes [39, 45]. Comparison of SSR densities in different genomic regions demonstrated that SSR density was greatest immediately in 5′-flanking regions of predicted genes [45]. SSRs were also preferentially associated with gene-rich regions, with pericentromeric heterochromatin SSRs mostly associated with retrotransposons [45], suggesting that the distribution of SSRs in the genome is nonrandom [39, 45].

2.5. Structure of (peri)centromeres of B. rapa

The centromere is a dynamic and rapidly evolving structure and consists largely of highly repetitive DNA sequences, especially tandem satellite repeats and retrotransposons [46, 47]. Centromeric repeats characterized in plant genomes are composed of 155 ~ 180-bp tandem repeat motifs, including the 180-bp pAL1 satellite in *A. thaliana* [48–50], the 155 ~ 165-bp CentO satellite in rice [51, 52], the 156-bp CentC satellite in maize [53] and the 169-bp satellite in *Medicago truncatula* [54, 55]. Centromeric satellite repeats of *Brassica* species, except for those of *B. nigra*, are represented by the 176-bp CentBr [24, 25, 56–59]. The CentBr repeats in the *B. rapa* genome belong to two classes which have 82% sequence similarity. The two classes are chromosome-specific, with CentBr1 found on eight chromosomes (chromosomes 1, 3, and 5–10) and CentBr2 on two chromosomes (chromosomes 2 and 4) [24, 25, 39]. Such distribution of the CentBr family may reflect the predominance of CentBr1 in the *Brassica* genome [39]. The CentBr repeats have also undergone rapid evolution within the *B. rapa* genome and have diverged among the related species of *Brassicaceae* [39]. Recently, Lim et al. [59] identified and characterized the major repeats in centromeric and pericentromeric heterochromatin of *B. rapa*. The region contains CentBr arrays, 238-bp degenerate tandem repeat (TR238) arrays, rDNAs, centromere-specific retrotransposons of *Brassica* (CRB), and pericentromere-specific retrotransposons (PCRBr). In particular, CRB was a major component of all centromeres in three diploid *Brassica* species and their allotetraploid relatives, and PCRBr and TR238 were A-genome-specific [59].

3. PROGRESS OF B. RAPA GENOME SEQUENCING

3.1. Genomic resources

The development of genomic resources is a prerequisite to undertaking genome sequencing in any crop species. Genomic resources, including reference mapping populations, DNA libraries, and DNA sequences have been developed for *B. rapa* ssp. *pekinensis* inbred line Chiffu-401-42 (Table 2). Two reference mapping populations were derived from two *B. rapa* ssp. *pekinensis* inbred lines, Chiifu-401-42 and Kenshin-402-43 (CK), and comprise 78 double haploid (DH) lines (the CKDH population) and 201 recombinant inbred (RI) lines (the CKRI population). These mapping populations have been used for construction of reference genetic maps for genome sequencing [20]. The bacterial artificial chromosome (BAC) system, commonly used for

Table 2: Genomic resources for whole-genome sequencing of *B. rapa*.

Genomic resources	Source material	Number
Mapping populations		
DH line	Chiffu-401-42 × Kenshin-402-43	78 lines (F_2 generation)
RI line	Chiffu-401-42 × Kenshin-402-43	201 lines (F_8 generation)
BAC libraries		
*Hin*dIII (KBrH)	Chiffu-401-42	56592 clones (115 kb[(1)])
*Bam*HI (KBrB)	Chiffu-401-42	50688 clones (124 kb[(1)])
*Sau*3AI (KBrS)	Chiffu-401-42	55296 clones (100 kb[(1)])
cDNA libraries		
22 cDNA libraries	Different tissues of Chiffu-401-42 and Jangwon including leaves, roots, cotyledons, stems, seedlings, ovules, siliques, anthers	—
BAC-end sequences	KBrH, KBrB, and KBrS clones	200017 sequences
ESTs	22 cDNA clones	129928 sequences
BAC shotgun sequences	KBrH, KBrB, and KBrS clones	on-going[(2)]

[(1)] Average insert size (kb).
[(2)] Of BACs sequenced, 511 BACs have been deposited in GenBank.

Table 3: Genetic linkage maps of *B. rapa* developed since 1990.

Mapping population	Population type	Population size	No. of loci	Type of markers	Total length of map (average interval)	References
Michihili × Spring broccoli	F_2	95	280	RFLP	1850 cM (6.6 cM)	[62]
Per (winter turnip rape)× R500 (spring yellow sarson)	F_2	91	139	RFLP	1785 cM (13.5 cM)	[63]
Per (winter turnip rape)× R500 (spring yellow sarson)	F_6RI	87	144	RFLP	890 cM (6.0 cM)	[64]
Developed from Chinese cabbage F_1 cultivar Jangwon	F_2	134	545	RFLP, SSR	1287 cM (2.4 cM)	[65]
G004 (CR[(a)] DH line)× A9709 (CS[(b)] DH line)(cultivars of Chinese cabbage)	F_2	94	262	RFLP, SSR, RAPD	1005 cM (3.7 cM)	[66]
Chiffu-401-42× Kenshin-402-43	DH	78	556	AFLP, SSR, RADP, ESTP, STS, CAPS	1182 cM (2.83 cM)	[67]

developing large-insert DNA libraries, is an invaluable resource for structural and functional genomics. Three Chiffu BAC libraries were constructed by using restriction enzymes: *Hin*dIII, *Bam*HI, and *Sau*3AI, and designated as KBrH, KBrB, and KBrS. These libraries consist of 56592, 50688, and 55296 clones with an average insert size of 115 kb, 124 kb, and 100 kb, respectively. These BAC libraries cover approximately 36 genome equivalents, assuming that the genome size of Chinese cabbage is 529 Mb. Using these BAC clones, the *Br*GSP community has recently generated a total of 200017 BAC-end sequences. In combination with BAC fingerprinting data, the BAC-end sequences will give insight into the structure of the genome, be a resource for development of genetic markers, and aid in finding the BAC clones that correspond to the minimal tilling paths in genome sequencing [19, 60, 61]. For functional genomics of *B. rapa*, 22 cDNA libraries from different tissues, including leaves, roots, cotyledons, stems, seedlings, ovules, siliques, and anthers of Chiffu, have been constructed, and a total of 128582 expressed sequence tags (ESTs) have been generated from these cDNA libraries (GenBank accession number CO749247 ~ CO750684 and EX015357 ~ EX142500). Currently, the ESTs have been used for construction of *B. rapa* unigene set and gene expression microarray (http://www.brassica-rapa.org).

3.2. Genetic and physical mapping

Some genetic linkage maps of *B. rapa*, on which genetic markers were distributed over ten linkage groups, have been constructed since 1990 [62–67] (summarized in Table 3). The distances of genetic linkage maps ranged from 890 cM to 1850 cM. However, the genetic linkage maps may not provide direct and accurate genetic information for the Chiffu genome sequencing because of genetic variation between the mapping populations. For that reason, the *Br*GSP community has constructed the CK genetic linkage map. Using the 78 CKDH lines, a reference genetic linkage map has been constructed [67]. The map consists of a total of 556 markers, including 278 AFLPs, 235 SSRs, 25 RAPDs, and 18 ESTPs/STS/CAPS markers. Ten linkage groups were

TABLE 4: The correspondence between genetic linkage groups of *B. rapa* ssp. *pekinensis* based on *B. napus* reference linkage maps.

Genetic linkage map of *B. napus* [68]	Choi et al. [67]	Kim et al. [65]	Suwabe et al. [66]
A1 (N1)	R1	R1	LG6
A2 (N2)	R2	R2	LG8
A3 (N3)	R3	R3	LG1
A4 (N4)	R4	R4	LG10
A5 (N5)	R5	R5	LG3
A6 (N6)	R6	R6	LG2
A7 (N7)	R7	R7	LG4
A8 (N8)	R8	R8	LG7
A9 (N9)	R9	R9	LG5
A10 (N10)	R10	R10	LG9

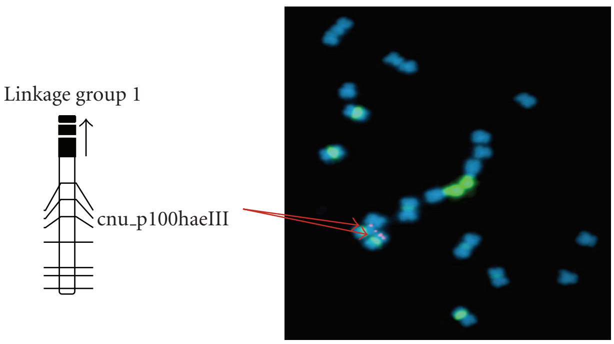


FIGURE 3: An example of an alignment of linkage group 1 in the reference genetic map to the corresponding chromosome 5 through FISH using locus-specific BAC clones.

identified and designated as R1 to R10 via mapping with SSR markers derived from the reference linkage map of *B. napus* reported previously [68] (Table 4). The total length of the linkage map was 1182 cM with an average interval of 2.83 cM between adjacent loci. Recently, for high-resolution genetic mapping, the community has set a goal of developing more than 1,000 SSR markers derived from BAC-end sequences, ESTs, and BACs. Moreover, based on the sequence-tagged site (STS) markers, a CKRI genetic linkage map has been constructed to be complementary to the CKDH one. The linkage groups in these genetic maps may not correspond to the chromosomes assigned in the cytogenetic map. Therefore, it is important to align the linkage groups on the genetic map with chromosomes of the cytogenetic map. All ten linkage groups of a reference genetic map of *B. rapa* are being assigned to the corresponding chromosomes through fluorescence in situ hybridization (FISH) using locus-specific BAC clones as probes (see an example in Figure 3, unpublished data).

The fingerprinted BAC map (so-called "physical map") makes it possible to select clones for sequencing that would ensure comprehensive coverage of the genome and reduce sequencing redundancy [69]. In addition, the clone-based map also enables the identification of large segments of the genome that are repeated, thereby simplifying the sequence assembly. To construct a deep-coverage BAC physical map of the *B. rapa* genome, all BAC clones from the three BAC libraries were fingerprinted using restriction enzyme digestion and SNaPshot [70] methodologies, and then BAC contigs have been assembled by FingerPrinted Contigs (FPC) software (http://www.agcol.arizona.edu/software/fpc/). This data will be open to the *Brassica rapa* genome sequencing consortium.

3.3. *Approach to genome sequencing*

Seed BACs for genome sequencing have been selected through in silico allocation of *B. rapa* BAC-end sequences onto counterpart locations of *Arabidopsis* chromosomes [19]. Of 91000 BAC-end sequences, a total of 45232 showed significant sequence similarity with unique *Arabidopsis* sequences, and 4317 BAC clones were allocated on *Arabidopsis* chromosomes by significant matching with both ends within 30–500 kb intervals, which span 93 Mb of *Arabidopsis* euchromatin regions (covering 78.2% of the *Arabidopsis* genome). However, approximately 9.4 Mb of euchromatin regions and 16.6 Mb pericentromeric heterochromatin regions of the *Arabidopsis* genome were not covered by the *B. rapa* BAC span (span is considered by best hit of paired ends). Based on the physical map of *B. rapa* and the in silico comparative map of its BAC-ends onto *Arabidopsis* chromosomes, 629 seed BACs have been selected spanning 86 Mb of *Arabidopsis* euchromatin regions and scattered throughout the *B. rapa* genome (http://www.brassica-rapa.org), and the BACs have been mapped on *B. rapa* chromosomes by STS mapping and FISH analysis. The seed BACs which are anchored and sequenced will be used as stepping stones for sequencing of the ten chromosomes.

Considering the large genome size and the possibility of international cooperation, a chromosome-based approach was suggested. Of ten chromosomes (or linkage groups), eight have been allocated to the participating countries as follow: Korea (R3 and R9), Canada (R2 and R10), UK and China (R1 and R8), USA (R6), and Australia (R7). However, R4 and R5 have remained unassigned. Progress of chromosome sequencing will be reported soon by each country.

4. CONCLUSIONS

Brassica species are economically important crops and serve as model plants for studying phenotypic evolution associated with polyploidization. The *Brassica* genomes have extensively triplicated and undergone subsequent genome rearrangements with sequence variations. This has significantly affected their genome structure and may underline phenotypic diversity. Genome sequencing of *B. rapa* can pave the way for elucidation of the relationship between genome evolution and phenotypic diversity. Moreover, it enables us to search for genes and develop molecular markers associated with agricultural traits, thereby establishing a molecular breeding system contributing to improvement of *Brassica* species economically.

ACKNOWLEDGMENTS

This research was supported by grants from Rural Development Administration (BioGreen 21 Program), National Institute of Agricultural Biotechnology (Project no. 04-1-12-2), and Korean Science and Engineering Foundation (R21-2004-000-10010-0), South Korea.

REFERENCES

[1] K. S. Labana and M. L. Gupta, "Importance and origin," in *Breeding Oilseed Brassicas*, K. S. Labana, S. S. Banga, and S. K. Banga, Eds., Spinger, Berlin, Germany, 1993.

[2] J. Fahey and P. Talalay, "The role of crucifers in cancer chemoprotection," in *Phytochemicals and Health*, D. L. Gustine and H. E. Florens, Eds., pp. 87–93, American Society of Plant Physiologists, Rockvill, Md, USA, 1995.

[3] N. U, "Genome analysis in Brassica with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization," *Japanese Journal of Botany*, vol. 7, no. 3, pp. 389–452, 1935.

[4] J. S. Johnston, A. E. Pepper, A. E. Hall, et al., "Evolution of genome size in Brassicaceae," *Annals of Botany*, vol. 95, no. 1, pp. 229–235, 2005.

[5] A. H. Paterson, T. H. Lan, R. Amasino, T. C. Osborn, and C. Quiros, "Brassica genomics: a complement to, and early beneficiary of, the *Arabidopsis* sequence," *GenomeBiology*, vol. 2, no. 3, pp. 1011.1–1011.4, 2001.

[6] L. N. Lukens, P. A. Quijada, J. Udall, J. C. Pires, M. E. Schranz, and T. C. Osborn, "Genome redundancy and plasticity within ancient and recent Brassica crop species," *Biological Journal of the Linnean Society*, vol. 82, no. 4, pp. 665–674, 2004.

[7] U. Lagercrantz, "Comparative mapping between *Arabidopsis thaliana* and *Brassica nigra* indicates that *Brassica* genomes have evolved through extensive genome replication accompanied by chromosome fusions and frequent rearrangements," *Genetics*, vol. 150, no. 3, pp. 1217–1228, 1998.

[8] T. H. Lan, T. A. Delmonte, K. P. Reischmann, et al., "An EST-enriched comparative map of *Brassica oleracea* and *Arabidopsis thaliana*," *Genome Research*, vol. 10, no. 6, pp. 776–788, 2000.

[9] C. M. O'Neill and I. Bancroft, "Comparative physical mapping of segments of the genome of *Brassica oleracea var. alboglabra* that are homoeologous to sequenced regions of chromosomes 4 and 5 of *Arabidopsis thaliana*," *The Plant Journal*, vol. 23, no. 2, pp. 233–243, 2000.

[10] D. Babula, M. Kaczmarek, A. Barakat, M. Delseny, C. F. Quiros, and J. Sadowski, "Chromosomal mapping of *Brassica oleracea* based on ESTs from *Arabidopsis thaliana*: complexity of the comparative map," *Molecular Genetics and Genomics*, vol. 268, no. 5, pp. 656–665, 2003.

[11] D. Rana, T. van den Boogaart, C. M. O'Neill, et al., "Conservation of the microstructure of genome segments in *Brassica napus* and its diploid relatives," *The Plant Journal*, vol. 40, no. 5, pp. 725–733, 2004.

[12] M. A. Lysak, M. A. Koch, A. Pecinka, and I. Schubert, "Chromosome triplication found across the tribe *Brassiceae*," *Genome Research*, vol. 15, no. 4, pp. 516–525, 2005.

[13] J. Y. Park, D. H. Koo, C. P. Hong, et al., "Physical mapping and microsynteny of *Brassica rapa* ssp. *pekinensis* genome corresponding to a 222 kbp gene-rich region of *Arabidopsis* chromosome 4 and partially duplicated on chromosome 5," *Molecular Genetics and Genomics*, vol. 274, no. 6, pp. 579–588, 2005.

[14] I. A. P. Parkin, S. M. Gulden, A. G. Sharpe, et al., "Segmental structure of the *Brassica napus* genome based on comparative analysis with *Arabidopsis thaliana*," *Genetics*, vol. 171, no. 2, pp. 765–781, 2005.

[15] C. D. Town, F. Cheung, R. Maiti, et al., "Comparative genomics of *Brassica oleracea* and *Arabidopsis thaliana* reveal gene loss, fragmentation, and dispersal after polyploidy," *The Plant Cell*, vol. 18, no. 6, pp. 1348–1359, 2006.

[16] T. J. Yang, J. S. Kim, S. J. Kwon, et al., "Sequence-level analysis of the diploidization process in the triplicated FLOWERING LOCUS C region of *Brassica rapa*," *The Plant Cell*, vol. 18, no. 6, pp. 1339–1347, 2006.

[17] L. N. Lukens, J. C. Pires, E. Leon, R. Vogelzang, L. Oslach, and T. Osborn, "Patterns of sequence loss and cytosine methylation within a population of newly resynthesized *Brassica napus* allopolyploids," *Plant Physiology*, vol. 140, no. 1, pp. 336–348, 2006.

[18] M. Ayele, B. J. Haas, N. Kumar, et al., "Whole genome shotgun sequencing of *Brassica oleracea* and its application to gene discovery and annotation in *Arabidopsis*," *Genome Research*, vol. 15, no. 4, pp. 487–495, 2005.

[19] T. J. Yang, J. S. Kim, K. B. Lim, et al., "The Korea *Brassica* genome project: a glimpse of the *Brassica* genome based on comparative genome analysis with *Arabidopsis*," *Comparative and Functional Genomics*, vol. 6, no. 3, pp. 138–146, 2005.

[20] Y. P. Lim, P. Plaha, S. R. Choi, et al., "Towards unraveling the structure of *Brassica rapa* genome," *Physiologia Plantarum*, vol. 126, no. 4, pp. 585–591, 2006.

[21] B. F. Cheng, W. K. Heneen, and B. Y. Chen, "Mitotic karyotypes of *Brassica campestris* and *Brassica alboglabra* and identification of the *B.alboglabra* chromosome in an addition line," *Genome*, vol. 38, no. 2, pp. 313–319, 1995.

[22] K. Fukui, S. Nakayama, N. Ohmido, H. Yoshiaki, and M. Yamabe, "Quantitative karyotyping of three diploid *Brassica* species by imaging methods and localization of 45s rDNA loci on the identified chromosomes," *Theoretical and Applied Genetics*, vol. 96, no. 3-4, pp. 325–330, 1998.

[23] R. J. Snowdon, T. Friedrich, W. Friedt, and W. Köhler, "Identifying the chromosomes of the A- and C-genome diploid *Brassica* species *B. rapa* (*syn. campestris*) and *B. oleracea* in their amphidiploid *B. napus*," *Theoretical and Applied Genetics*, vol. 104, no. 4, pp. 533–538, 2002.

[24] K. B. Lim, H. de Jong, T. J. Yang, et al., "Characterization of rDNAs and tandem repeats in the heterochromatin of *Brassica rapa*," *Molecules and Cells*, vol. 19, no. 3, pp. 436–444, 2005.

[25] D. H. Koo, P. Plaha, Y. P. Lim, Y. Hur, and J.-W. Bang, "A high-resolution karyotype of *Brassica rapa* ssp. *pekinensis* revealed by pachytene analysis and multicolor fluorescence in situ hybridization," *Theoretical and Applied Genetics*, vol. 109, no. 7, pp. 1346–1352, 2004.

[26] The Arabidopsis Genome Initiative, "Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*," *Nature*, vol. 408, no. 6814, pp. 796–815, 2000.

[27] M. Bevan and S. Walsh, "The *Arabidopsis* genome: a foundation for plant research," *Genome Research*, vol. 15, no. 12, pp. 1632–1642, 2005.

[28] C. F. Quiros, F. Grellet, J. Sadowski, T. Suzuki, G. Li, and T. Wroblewski, "*Arabidopsis* and *Brassica* comparative genomics: sequence, structure and gene content in the *ABI1-Rps2-Ck1* chromosomal segment and related regions," *Genetics*, vol. 157, no. 3, pp. 1321–1330, 2001.

[29] J. A. Kim, T. J. Yang, J. S. Kim, et al., "Isolation of circadian-associated genes in *Brassica rapa* by comparative genomics

with *Arabidopsis thaliana*," *Molecules and Cells*, vol. 23, no. 2, pp. 145–153, 2007.

[30] S. P. Kowalski, T. H. Lan, K. A. Feldmann, and A. H. Paterson, "Comparative mapping of *Arabidopsis thaliana* and *Brassica oleracea* chromosomes reveals islands of conserved organization," *Genetics*, vol. 138, no. 2, pp. 499–510, 1994.

[31] S. A. Jackson, Z. Cheng, M. L. Wang, H. M. Goodman, and J. Jiang, "Comparative fluorescence in situ hybridization mapping of a 431-kb *Arabidopsis* thaliana bacterial artificial chromosome contig reveals the role of chromosomal duplications in the expansion of the *Brassica rapa* genome," *Genetics*, vol. 156, no. 2, pp. 833–838, 2000.

[32] P. A. Ziolkowski, M. Kaczmarek, D. Babula, and J. Sadowski, "Genome evolution in *Arabidopsis/Brassica*: conservation and divergence of ancient rearranged segments and their breakpoints," *The Plant Journal*, vol. 47, no. 1, pp. 63–74, 2006.

[33] M. Lynch and J. S. Conery, "The evolutionary fate and consequences of duplicate genes," *Science*, vol. 290, no. 5494, pp. 1151–1155, 2000.

[34] A. Lawton-Rauh, "Evolutionary dynamics of duplicated genes in plants," *Molecular Phylogenetics and Evolution*, vol. 29, no. 3, pp. 396–409, 2003.

[35] Z. Gu, S. A. Rifkin, K. P. White, and W.-H. Li, "Duplicate genes increase gene expression diversity within and between species," *Nature Genetics*, vol. 36, no. 6, pp. 577–579, 2004.

[36] K. L. Adams and J. F. Wendel, "Novel patterns of gene expression in polyploid plants," *Trends in Genetics*, vol. 21, no. 10, pp. 539–543, 2005.

[37] R. C. Moore and M. D. Purugganan, "The evolutionary dynamics of plant duplicate genes," *Current Opinion in Plant Biology*, vol. 8, no. 2, pp. 122–128, 2005.

[38] W.-H. Li, J. Yang, and X. Gu, "Expression divergence between duplicate genes," *Trends in Genetics*, vol. 21, no. 11, pp. 602–607, 2006.

[39] C. P. Hong, P. Plaha, D.-H. Koo, et al., "A survey of the *Brassica rapa* genome by BAC-end sequence analysis and comparison with *Arabidopsis thaliana*," *Molecules and Cells*, vol. 22, no. 3, pp. 300–307, 2006.

[40] C. P. Hong, S. J. Lee, J. Y. Park, et al., "Construction of a BAC library of Korean ginseng and initial analysis of BAC-end sequences," *Molecular Genetics and Genomics*, vol. 271, no. 6, pp. 709–716, 2004.

[41] J. Messing, A. K. Bharti, W. M. Karlowski, et al., "Sequence composition and genome organization of maize," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 40, pp. 14349–14354, 2004.

[42] J. E. Bowers, B. A. Chapman, J. Rong, and A. H. Paterson, "Unraveling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events," *Nature*, vol. 422, no. 6930, pp. 433–438, 2003.

[43] B. J. Haas, J. R. Wortman, C. M. Ronning, et al., "Complete reannotation of the *Arabidopsis* genome: methods, tools, protocols and the final release," *BMC Biology*, vol. 3, p. 7, 2005.

[44] X. Zhang and S. R. Wessler, "Genome-wide comparative analysis of the transposable elements in the related species *Arabidopsis thaliana* and *Brassica oleracea*," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 15, pp. 5589–5594, 2004.

[45] C. P. Hong, Z. Y. Piao, T. W. Kang, et al., "Genomic distribution of simple sequence repeats in *Brassica rapa*," *Molecules and Cells*, vol. 23, no. 3, pp. 349–356, 2007.

[46] M. Ventura, N. Archidiacono, and M. Rocchi, "Centromere emergence in evolution," *Genome Research*, vol. 11, no. 4, pp. 595–599, 2001.

[47] J. Jiang, J. A. Birchler, W. A. Parrott, and R. K. Dawe, "A molecular view of plant centromeres," *Trends in Plant Science*, vol. 8, no. 12, pp. 570–575, 2003.

[48] K. Nagaki, P. B. Talbert, C. X. Zhong, R. K. Dawe, S. Henikoff, and J. Jiang, "Chromatin immunoprecipitation reveals that the 180-bp satellite repeat is the key functional DNA element of *Arabidopsis thaliana* centromeres," *Genetics*, vol. 163, no. 3, pp. 1221–1225, 2003.

[49] E. K. Round, S. K. Flowers, and E. J. Richards, "*Arabidopsis thaliana* centromere regions: genetic map positions and repetitive DNA structure," *Genome Research*, vol. 7, no. 11, pp. 1045–1053, 1997.

[50] H. Thompson, R. Schmidt, A. Brandes, J. S. Heslop-Harrison, and C. Dean, "A novel repetitive sequence associated with the centromeric regions of *Arabidopsis thaliana* chromosomes," *Molecular Genetics and Genomics*, vol. 253, no. 1-2, pp. 247–252, 1996.

[51] Z. Cheng, F. Dong, T. Langdon, et al., "Functional rice centromeres are marked by a satellite repeat and a centromere-specific retrotransposon," *The Plant Cell*, vol. 14, no. 8, pp. 1691–1704, 2002.

[52] Y. Zhang, Y. Huang, L. Zhang, et al., "Structural features of the rice chromosome 4 centromere," *Nucleic Acids Research*, vol. 32, no. 6, pp. 2023–2030, 2004.

[53] E. V. Ananiev, R. L. Phillips, and H. W. Rines, "Chromosome-specific molecular organization of maize (*Zea mays* L.) centromeric regions," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, no. 22, pp. 13073–13078, 1998.

[54] O. Kulikova, G. Gualtieri, R. Geurts, et al., "Integration of the FISH pachytene and genetic maps of *Medicago truncatula*," *The Plant Journal*, vol. 27, no. 1, pp. 49–58, 2001.

[55] O. Kulikova, R. Geurts, M. Lamine, et al., "Satellite repeats in the functional centromere and pericentromeric heterochromatin of *Medicago truncatula*," *Chromosoma*, vol. 113, no. 6, pp. 276–283, 2004.

[56] X. Xia, G. Selvaraj, and H. Bertrand, "Structure and evolution of a highly repetitive DNA sequence from *Brassica napus*," *Plant Molecular Biology*, vol. 21, no. 2, pp. 213–224, 1993.

[57] X. Xia, P. S. Rocha, G. Selvaraj, and H. Bertrand, "Genomic organization of the canrep repetitive DNA in *Brassica juncea*," *Plant Molecular Biology*, vol. 26, no. 3, pp. 817–832, 1994.

[58] G. E. Harrison and J. S. Heslop-Harrison, "Centromeric repetitive DNA sequences in the genus *Brassica*," *Theoretical and Applied Genetics*, vol. 90, no. 2, pp. 157–165, 1995.

[59] K. B. Lim, T. J. Yang, Y. J. Hwang, et al., "Characterization of the centromere and peri-centromere retrotransposons in *Brassica rapa* and their distribution in related *Brassica* species," *The Plant Journal*, vol. 49, no. 2, pp. 173–183, 2007.

[60] J. C. Venter, H. O. Smith, and L. Hood, "A new strategy for genome sequencing," *Nature*, vol. 381, no. 6581, pp. 364–366, 1996.

[61] G. G. Mahairas, J. C. Wallace, K. Smith, et al., "Sequence-tagged connectors: a sequence approach to mapping and scanning the human genome," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96, no. 17, pp. 9739–9744, 1999.

[62] K. M. Song, J. Y. Suzuki, M. K. Slocum, P. M. Williams, and T. C. Osborn, "A linkage map of *Brassica rapa* (syn. *campestris*) based on restriction fragment length polymorphism loci,"

Theoretical and Applied Genetics, vol. 82, no. 3, pp. 296–304, 1991.

[63] R. A. Teutonico and T. C. Osborn, "Mapping of RFLP and qualitative trait loci in *Brassica rapa* and comparison to the linkage maps of *B.napus, B.oleracea*, and *Arabidopsis thaliana*," *Theoretical and Applied Genetics*, vol. 89, no. 7-8, pp. 885–894, 1994.

[64] C. Kole, P. Kole, R. Vogelzang, and T. C. Osborn, "Genetic linkage map of a *Brassica rapa* recombinant inbred population," *Journal of Heredity*, vol. 88, no. 6, pp. 553–557, 1997.

[65] J. S. Kim, T. Y. Chung, G. J. King, et al., "A sequence-tagged linkage map of *Brassica rapa*," *Genetics*, vol. 174, no. 1, pp. 29–39, 2006.

[66] K. Suwabe, H. Tsukazaki, H. Iketani, et al., "Simple sequence repeat-based comparative genomics between *Brassica rapa* and *Arabidopsis thaliana*: the genetic origin of clubroot resistance," *Genetics*, vol. 173, no. 1, pp. 309–319, 2006.

[67] S. R. Choi, G. R. Teakle, P. Plaha, et al., "The reference genetic linkage map for the multinational *Brassica rapa* genome sequencing project," *Theoretical and Applied Genetics*, vol. 115, no. 6, pp. 777–792, 2007.

[68] I. A. Parkin, A. G. Sharpe, D. J. Keith, and D. J. Lydiate, "Identification of the A and C genomes of amphiploid *Brassica napus* (oilseed rape)," *Genome*, vol. 38, pp. 1122–1133, 1995.

[69] The International Human Genome Mapping Consortium, "A physical map of the human genome," *Nature*, vol. 409, no. 6822, pp. 934–941, 2001.

[70] M. C. Luo, C. Thomas, F. M. You, et al., "High-throughput fingerprinting of bacterial artificial chromosomes using the SNaPshot labeling kit and sizing of restriction fragments by capillary electrophoresis," *Genomics*, vol. 82, no. 3, pp. 378–389, 2003.

Hindawi Publishing Corporation
International Journal of Plant Genomics
Volume 2008, Article ID 742304, 20 pages
doi:10.1155/2008/742304

Review Article

Recent Advances in Cotton Genomics

Hong-Bin Zhang,[1] Yaning Li,[2] Baohua Wang,[3,4] and Peng W. Chee[3]

[1] *Department of Soil and Crop Sciences, Texas A&M University, College Station, TX 77843, USA*
[2] *Department of Plant Pathology, Biological Control Center of Plant Diseases and Plant Pests of Hebei Province, Agricultural University of Hebei, Baoding 071001, China*
[3] *Molecular Cotton Breeding Laboratory, University of Georgia, Tifton, GA 31793, USA*
[4] *School of Life Sciences, Nantong University, Nantong 226007, China*

Correspondence should be addressed to Hong-Bin Zhang, hbz7049@tamu.edu

Received 27 June 2007; Accepted 10 October 2007

Recommended by P. K. Gupta

Genome research promises to promote continued and enhanced plant genetic improvement. As a world's leading crop and a model system for studies of many biological processes, genomics research of cottons has advanced rapidly in the past few years. This article presents a comprehensive review on the recent advances of cotton genomics research. The reviewed areas include DNA markers, genetic maps, mapped genes and QTLs, ESTs, microarrays, gene expression profiling, BAC and BIBAC libraries, physical mapping, genome sequencing, and applications of genomic tools in cotton breeding. Analysis of the current status of each of the genome research areas suggests that the areas of physical mapping, QTL fine mapping, genome sequencing, nonfiber and nonovule EST development, gene expression profiling, and association studies between gene expression and fiber trait performance should be emphasized currently and in near future to accelerate utilization of the genomics research achievements for enhancing cotton genetic improvement.

1. INTRODUCTION

Cottons (*Gossypium spp.*) belong to the genus *Gossypium* of the family Malvaceae. *Gossypium* consists of 45–50 species, with 40–45 being diploids ($2n = 26$) and 5 being allotetraploids ($2n = 52$). The species are grouped into eight genome groups, designated A through G and K, on the basis of chromosome pairing affinities [1]. At the tetraploid level, there are five species, designated $(AD)_1$ through $(AD)_5$ for their genome constitutions. Phylogenetic analyses clustered the diploid species of *Gossypium* into two major lineages, including the 13 D-genome species lineage and the 30~32 A-, B-, E-, F-, C-, G-, and K-genome species lineage, and the polyploid species into one lineage, that is, the 5 AD-genome species lineage (Figure 1; [2]).

Of the *Gossypium* species, four are cultivated in agriculture, including two allotetraploids (*G. hirsutum* and *G. barbadense*) and two diploids (*G. herbaceum* and *G. arboreum*). *Gossypium hirsutum*, also known as Upland cotton, Long Staple Cotton, or Mexican Cotton, produces over 90% of the world's cotton; *G. barbadense*, also known as Sea Island Cotton, Extra Long Staple Cotton, American Pima, or Egyptian Cotton, contributes 8% of the world's cotton; and *G. herbaceum*, also known as Levant Cotton, and *G. arboreum*, also known as Tree Cotton, together provide 2% of the world's cotton.

Cottons are not only a world's leading textile fiber and oilseed crop, but also a crop that is of significance for foil energy and bioengergy production. Although cottons are native to tropics and subtropics naturally, including the Americas, Africa and Asia, they are cultivated in nearly 100 countries. India, China, USA, and Pakistan are the top four cotton growing countries, accounting for approximately 2/3 of the world's cotton (http://www.ers.usda.gov/Briefing/Cotton/trade.htm). According to the Food and Agriculture Organization (FAO) of the United Nations (http://www.fao.org), the cotton planting area reached about 35 million hectares and the total world's cotton production had a record of about 23 million metric tones in 2004/2005. Cotton products include fibers and seeds that have a variety of uses. Cotton fibers sustain one of the world's largest industries, the textile industry, for wearing

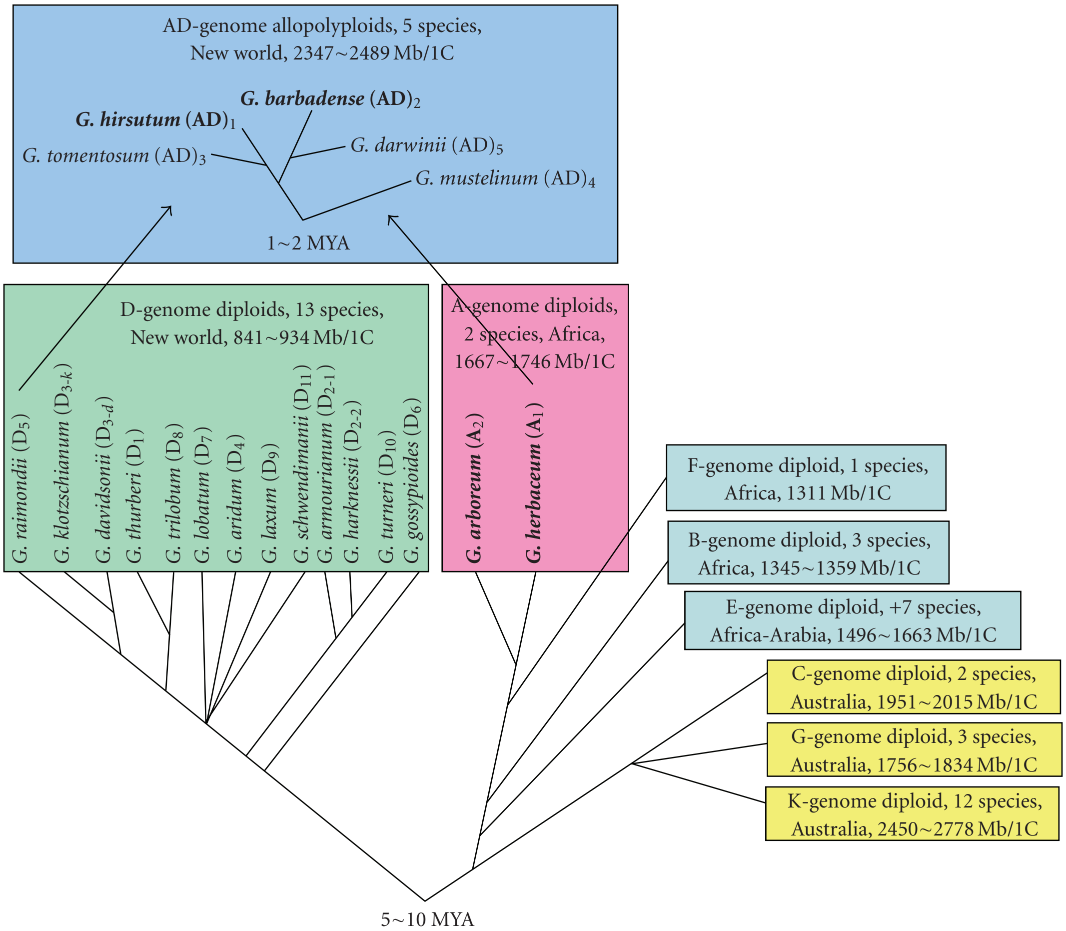


FIGURE 1: Phylogeny and evolution of *Gossypium* species. The phylogenetic data is from Wendel and Cronn [2], the genome sizes are from Hendrix and Stewart [3], and genomic designations follow Endrizzi et al. [4] and Percival [5]. The species in bold face are cultivated. MYA: million years ago.

apparel, home furnishings, and medical supplies, whereas cottonseeds are widely used for food oil, animal feeds, and industrial materials (such as soap). Cottonseed oil is ranked fifth in production and consumption volume among all vegetable oils in the past decades, accounting for 8% of the world's vegetable oil consumption. The business stimulated by cotton is hundreds of billion dollars in the world. In the USA alone, for instance, the annual cotton business revenue exceeds $120 billion (Agricultural Statistics Board 1999; National Cotton Council of America, http://www.cotton.org/news/releases/2003/cotton-trade.cfm). Moreover, nearly a billion barrels of petroleum worldwide are used in every year to synthesize artificial "synthetic" fibers. Further improvement of cotton fibers in yield and quality will replace or significantly reduce the consumption of fossil oil for synthetic fiber production, thus being saved for energy production. Finally, cottonseed oil, the main by-product of cotton fiber production, could be potentially used as biofuel.

In addition to their economic importance, cottons are an excellent model system for several important biological studies, including plant genome size evolution, plant polyploidization and single-celled biological processes. The genomes of angiosperm plants vary over 1000 folds in size, ranging from 100 to >100,000 Mb/1C (haploid) [6]. It has long been recognized that polyploidy is a common, prominent, ongoing, and dynamic process of genome organization, function diversification, and evolution in angiosperms [7]. The genomes of most angiosperms are thought to have incurred one or more polyploidization events during evolution [8]. Studies have demonstrated that genome doubling has also been significant in the evolutionary history of all vertebrates and in many other eukaryotes [9–12]. It is estimated that about 70% of the flowering plant species are polyploids. For instance, of the world-leading field, forage, horticultural, and environmental crops, many are contributed by polyploid species, such as cotton, wheat, soybean, potatoes, canola, sugarcane, *Brassica*, oats, peanut, tobacco, rose, coffee, and banana. Therefore, studies of both genome size evolution and polyploidization have long attracted the interests of scientists in different disciplines. Nevertheless, much remains to be learned. Examples include impacts of polyploidization on genome size, genome organization, gene duplication and function, and gene family evolution; the role of transposable elements in structural and regulatory gene evolution and gene functions; and mechanisms and functional significance of rapid genome changes.

Cottons have several advantages over other polyploid complexes for plant genome size and polyploidization studies. First, the genome sizes of 37 of the 45∼50 *Gossypium* species, including all eight genomes and polyploidy species, have been determined and shown to vary extremely significantly ([3]; Figure 1). At the diploid level, the genome sizes vary by three folds, ranging from 885 Mb/1C in the D-genome species to 2,572 Mb/1C in the K-genome species. Within each lineage, the genome sizes vary most in the A+F+B+E+C+G+K lineage, ranging from 1,311 to 2,778 Mb/1C with a difference of 1,467 Mb (110.2%); second in the D-genome lineage, ranging from 841 to 934 Mb/1C with a difference of 93 Mb (10.5%); and least in the polyploidy lineage, ranging from 2,347 to 2,489 Mb/1C with a difference of 142 Mb (5.9%). Variations were also observed within a species. For instance, within G. *hirsutum*, the variation ($n = 5$) was from 2,347 to 2,489 Mb/1C, differing by 142 Mb (5.9%) while within *G. arboreum*, the variation ($n = 5$) was from 1,677 to 1,746 Mb/1C, differing by 69 Mb (4.0%).

Second, the evolutionary history of the allotetraploid species of *Gossypium* has been established (Figure 1), especially for the two cultivated AD-genome cottons, *G. hirsutum* and *G. barbadense,* and their closely related diploid progenitors, *G. herbaceum* (A_1), *G. arboreum* (A_2), *G. raimondii* (D_5), and *G. gossypioides* (D_6). The A-genome species are African-Asian in origin, whereas the D-genome species are endemic to the New World subtropics, primarily Mexico. Following the transoceanic dispersal of an A-genome taxon to the New World, hybridization between the immigrant A-genome taxon and a local D-genome taxon led to the origin and evolution of the New World allopolyploids (AD-genome) [13, 14]. Subsequent to the polyploidization event, the allopolyploids radiated into three sublineages [15], among which included are the world's commercially most important species, *G. hirsutum* and *G. barbadense*. Studies showed that the A subgenome of the AD-genome-cultivated cottons is the most closely related to the genome of the extant diploid *G. herbaceum* (A_1) [16]; the D subgenome of the AD-genome-cultivated cottons is the most closely related to the genome of the extant diploid, *G. raimondii* (D_5) or *G. gossypioides* (D_6) [13]; and the cytoplasm of the AD-genome-cultivated cottons is the most closely related to that of the extant diploids *G. herbaceum* (A_1) and *G. arboreum* (A_2) [14, 17]. Sequence analysis and paleontological record suggest that the A-genome and the D-genome groups diverged from a common ancestor 5–10 million years ago, and that the two diverged diploid genomes became reunited in a common nucleus to form the polyploid cottons, via allopolyploidization, in the mid-Pleistocene, or 1-2 million years ago [14, 15, 18, 19].

Finally, as in the wheat polyploid complex, cottons have a long history of research at the cytological level. A wealth of cytogenetic stocks has been developed, including artificially synthesized AD-genome polyploids between the A-genome and D-genome diploid species [20] as well as individual chromosome addition and substitution lines [21]. These cytogenetic stocks are unique and valuable not only for cotton genetics research, but also for deciphering the ramifications of polyploidization on genome organization, function, and evolution.

Cotton fiber is an excellent single-celled model system for studies of many single-celled biological processes, particularly cell expansion and cellulose biosynthesis. Cotton fibers are unicellular, unbranched, simple trichomes that differentiate from the protoderm of developing seeds. There are probably over one-half million quasi-synchronously elongating fibers in each boll or ovary. Although all plant cells extend to some degree during development and differentiation, cotton fibers can reach up to 5.0 cm in length in some genotypes, being among the longest cells. Therefore, they offer a unique opportunity to study cell expansion at the single cell level. Cellulose is a major component of the cell walls of all higher plants, constituting perhaps the largest component of plant biomass, with an estimated annual world production of 100 million metric tons. The fiber cell wall of cottons consists of >90% cellulose. Therefore, cotton fiber cells have long been used as a model system to study cellulose biosynthesis [22] that is the basis for biomass-based bioenergy production.

2. ADVANCES IN COTTON GENOMICS RESEARCH

Genome research has been demonstrated to be promising for continued and enhanced crop plant genetic improvement. Therefore, efforts have been made in cotton genome research, especially development of genomic resources and tools for basic and applied genetics, genomics, and breeding research. These resources and tools include different types of DNA markers such as restriction fragment length polymorphism (RFLP), randomly amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), resistance gene analogs (RGA), sequence-related amplified polymorphism (SRAP), simple sequence repeat (SSR) or microsatellites, DNA marker-based genetic linkage maps, QTLs and genes for the traits important to agriculture, expressed sequence tags (ESTs), arrayed large-insert bacterial artificial chromosome (BAC) and plant-transformation-competent binary BAC (BIBAC) libraries, and genome-wide, cDNA-, or unigene EST-based microarrays. Efforts are also being made to develop the genome-wide, BAC/BIBAC-based integrated physical and genetic maps, and sequence the genomes of the key cotton species. However, compared with other major crops, such as rice, maize, and soybean, the genome research of cottons is far behind, mainly due to the limited funds allocated to the species. Summarized below are the major advances achieved recently in cotton genomics research.

2.1. DNA markers and molecular linkage maps

Genetic maps constructed in the *Gossypium* species and the types of markers used are listed in Table 1. As in most plant species, the early application of DNA markers in cotton genomic research has been in the form of RFLPs. It is, therefore, not surprising that the first molecular linkage map of the *Gossypium* species was constructed from an interspecific

G. hirsutum × *G. barbadense* F_2 population based on RFLPs [23]. The map contained 705 loci that were assembled into 41 linkage groups and spanned 4,675 cM. This map later was further advanced by Rong et al. [24] that comprised 2,584 loci at 1.74-cM intervals and covered all 13 homeologous chromosomes of the allotetraploid cottons, representing the most complete genetic map of the *Gossypium* to date. Many of the DNA probes of the map were also mapped in crosses of the D-genome diploid species *G. trilobum* × *G. raimondii* [24] and the A-genome diploid species *G. arboreum* × *G. herbaceum* [16]. Detailed comparative analysis of the relationship of gene orders between the tetraploid AD-subgenomes with the maps of the A and D diploid genomes has revealed intriguing insights on the organization, transmission and evolution of the *Gossypium* genomes.

Because RFLPs are labor-intensive and require large amounts of DNA, tedious blot hybridization and autoradiographic methods, polymerase chain reaction (PCR)-based DNA marker methods have come into vogue. Several types of PCR-based DNA markers have been utilized in cotton genome research. Methods, such as RAPD, AFLP, RGA, and SRAP, offer an excellent opportunity to scan enormous numbers of DNA loci rapidly, often targeting the DNA elements that are rapidly-evolving and therefore, are more likely to contain loci differing among genotypes. Kohel et al. [25] constructed a genetic map based on a population derived from an interspecific cross between Texas Marker-1 (TM-1) (*G. hirsutum*) and 3–79 (*G. barbadense*) in which a total of 355 DNA markers (216 RFLPs and 139 RAPDs) were assembled into 50 linkage groups, covering 4,766 cM. Brubaker and Brown [26] presented the first AFLP genetic linkage map for the *Gossypium* G-genome that was constructed from an interspecific *G. nelsonii* × *G. australe* population. The AFLP genetic linkage maps were used to identify G-genome chromosome-specific molecular markers, which, in turn, were used to track the fidelity and frequency of *G. australe* chromosome transmission in a *G. hirsutum* × *G. australe* hexaploid bridging family.

Advent of SSR or microsatellite markers has brought a new, user-friendly, and highly polymorphic class of genetic markers for cotton. The latter feature is especially useful to the cultivated Upland cotton due to its low intraspecific polymorphism. SSRs are PCR-based markers, usually codominant, well dispersed throughout the genome, easily shared between labs via flanking primer sequences, and well portable from one population to another [84]. Reddy et al. [85] suggested that the total pool of SSRs present in the cotton genome is sufficiently abundant to satisfy the requirements of extensive genome mapping and marker-assisted selection (MAS). Liu et al. [86] reported the assignment of SSRs to cotton chromosomes by making use of aneuploid stocks. SSRs have been widely employed in genetic diversity analyses of cotton [87–90] and several genetic linkage maps based mostly on SSRs have now been developed [37–41].

Several methods have been pursued to develop SSR markers in cottons, including analysis of SSR-enriched small-insert genomic DNA libraries [29, 85, 86, 91], SSR mining from ESTs (see below; [35, 38, 39, 92], and large-insert BAC derivation by end sequence analysis [36] or SSR-containing BAC subcloning as described by Lichtenzveig et al. [93]. Currently, a total of approximately 5,484 SSRs have been developed in cotton ([94]; http://www.cottonmarker.org).

The development of a large number of ESTs (see below) provides a good source of PCR-based primers for targeting SSRs [92, 95, 96]. Taliercio et al. [97] sequenced ESTs representing a variety of tissues and treatments with SSRs identified among the ESTs. Their results indicated that these SSRs could potentially map the genes represented by the ESTs. Guo et al. [98] examined the transferability of 207 *G. arboreum*-derived EST-SSR primer pairs among 25 different diploid accessions from 23 species representing 7 *Gossypium* genomes. Their results demonstrated that the transferability of EST-SSR markers among these diploid species could assist the introgression of genes into cultivated cotton species especially by molecular tagging of the important genes existing in these diploid species. Guo et al. [40] also developed 2,218 EST-SSRs, with 1,554 from *G. raimondii*-derived ESTs and 754 from *G. hirsutum*-derived ESTs. By integrating these new EST-SSRs to enhance the genetic map constructed by Han et al. [39], the present SSR-based genetic map consists of 1,790 loci in 26 linkage groups and covers 3,425.8 cM with an average distance between markers of 1.91 cM. This SSR-based high-density map contains 71.96% functional marker loci, of which 87.11% are EST-SSR loci.

DNA sequences derived from clone end sequencing of BAC libraries provide yet another resource for SSR marker development. In addition to the uses as genetic markers, SSRs developed from BAC-end sequences provide the possibility to efficiently integrate the genetic and physical maps of cotton. Frelichowski et al. [36] developed 1,316 PCR primer pairs to flank SSR motif sequences from 2,603 new BAC-end genomic sequences developed from *G. hirsutum* Acala "Maxxa." An interspecific recombinant inbred population was used to map 433 marker loci in 46 linkage groups with a total genetic distance of 2,126.3 cM and an average distance between loci of 4.9 cM which covered approximately 45% of the cotton genome.

To overcome the paucity of a particular type of DNA markers, genetic maps were developed by incorporating different classes of markers. For example, Lacape et al. [28] constructed a combined RFLP-SSR-AFLP map based on an interspecific *G. hirsutum* × *G. barbadense* backcross population of 75 BC_1 plants. The map consists of 888 loci that ordered into 37 linkage groups and spanning 4,400 cM. This map was updated, mostly with new SSR markers, to contain 1,160 loci that spanned 5,519 cM with an average distance between loci of 4.8 cM [29]. Mei et al. [27] developed a genetic map using an interspecific *G. hirsutum* and *G. barbadense* F_2 population that contained 392 genetic loci, including AFLPs, SSRs, and RFLPs, and mapped into 42 linkage groups that spanned 3,287 cM, thus covering approximately 70% of the cotton genome. Lin et al. [33] constructed a linkage map of tetraploid cotton using SRAPs, SSRs, and RAPDs to screen an interspecific *G. hirsutum* × *G. barbadense* F_2 population. A total of 566 loci were assembled into 41 linkages that covered 5,141.8 cM with a mean interlocus space of 9.08 cM. He et al. [34] constructed a more detailed cotton map with this same F_2 population [33] using SSRs, SRAP,

Table 1: Genetic maps constructed for *Gossypium* species.

Marker type	Total loci	Map distance	Population[(a)]	Cross type[(b)]	References
AFLP	176	773 cM	F2	GN × GAU	[26]
AFLP	213	931 cM	F2	GN × GAU	[26]
AFLP, SSR, and RFLP	392	3,287 cM	F2	GH × GB	[27]
AFLP, SSR, and RFLP	888	4,400 cM	BC1	GH × GB	[28]
AFLP, SSR, and RFLP	1,160	5,519 cM	BC1	GH × GB	[29]
RFLP	275	1,147 cM	F2	GAR × GHE	[16]
RFLP	284	1,503 cM	F2 and F3	GH × GH	[30]
RFLP	589	4,259 cM	F2	GH × GTO	[31]
RFLP	705	4,675 cM	F2	GH × GB	[23]
RFLP	763	1,493 cM	F2	GT × GR	[24]
RFLP	2,584	4,448 cM	F2	GH × GB	[24]
RFLP and RAPD	355	4,766 cM	F2	GH × GB	[25]
SRAP	237	3,031 cM	F2	GH × GB	[32]
SRAP, SSR, and RAPD	566	5,142 cM	F2	GH × GB	[33]
SRAP, SSR, RAPD and REMAP	1,029	5,472 cM	F2	GH × GB	[34]
SSR	193	1,277 cM	RIL	GH × GB	[35]
SSR	433	2,126 cM	RIL	GH × GB	[36]
SSR	442	4,331 cM	BC1	GH × GB	[37]
SSR	444	3,263 cM	DH	GH × GB	[37]
SSR	624	5,644 cM	BC1	GH × GB	[38]
SSR	907	5,060 cM	BC1	GH × GB	[39]
SSR	1,790	3,426 cM	BC1	GH × GB	[40]
SSR and RAPD	489	3,315 cM	DH	GH × GB	[41]

[(a)] RIL = recombinant inbred line, and DH = doubled haploid.
[(b)] GH = *G. hirsutum*, GB = *G. barbadense*, GTO = *G. tomentosum*, GR = *G. raimondii*, GAR = *G. arboretum*, GHE = *G. herbaceum*, GN = *G. nelsonii*, and GAU = *G. australe*.

RAPD, and retrotransposon-microsatellite amplified polymorphisms (REMAPs). One thousand twenty nine loci were mapped to 26 linkage groups that extended for 5,472.3 cM with an average distance between loci of 5.32 cM. The linkage groups of the genetic maps have been assigned to their corresponding chromosomes by using the available cotton aneuploid stocks [21, 23] and fluorescent in situ hybridization using mapped genetic marker-containing BACs as probes [99].

2.2. Gene and QTL mapping

Although molecular linkage maps have contributed greatly to our understanding of the evolution and organization of the cotton genomes, a primary purpose of the map construction is to provide a common point of reference for locating the genes affecting qualitative and quantitative traits. DNA markers that are associated with genes conferring important agronomic traits that are costly or laborious to measure will provide a less costly and yet more dependable means of selection for identifying desirable progenies in breeding programs.

Mapping qualitative traits

Qualitative or simple Mendelian inherited traits are traits of individuals that differ as to kind and not of degree, typically controlled by single genes and the phenotypic variation falls into discrete classes in the segregating progenies. Over 200 qualitative traits have been identified in either the diploid (*G. arboreum* and *G. herbaceum*) or tetraploid (mostly in *G. hirsutum* and *G. barbadense*) species [1]. Examples of such traits include leaf shape, pollen color, leaf color, lint color, pubescent, bract shape, and so on. Because many qualitative traits are either morphological mutants that have arisen through spontaneous mutation, irradiation, or from natural variation between species in interspecific hybrids, they have little utility in crop improvement. Consequently, there have been little efforts in mapping qualitative traits onto the molecular genetic map. Qualitative traits that have been mapped using molecular markers were recently summarized in [105]. Many of these traits were mapped not as the main objective but as a tool for aligning the various linkage groups to chromosomes assigned by the classical map. Noteworthy exceptions include those that are related to agricultural productivity and quality of cotton and can be broadly grouped into four categories: genes for leaf shape, fiber development, resistant to disease and insect pests, and fertility restoration [105].

Mapping quantitative traits

Quantitative traits are traits of individuals that differ as to degree and not of kind, typically considered as interactions of multiple loci, tend to exhibit continuous variation in a

Table 2: QTLs or genes identified for various traits in cottons.

Traits/genes	Parental materials	Reference
Resistance to the bacterial blight pathogen	Empire B2/B3/B2b6, S295 and Pima S-7*	[42]
Resistance to the bacterial blight pathogen	CS50 and Pima S-7*	[43]
Density of leaf and stem trichomes	Pima S-7 and Empire B2b6*	[44]
Fiber quality and yield	CAMD-E and Sea Island Seaberry*	[45]
Agronomic and fiber traits	MARCABUCAG8US-1-88 and HS46	[46]
Cotton leaf morphology and other traits	Seaberry and Deltapine 61 with morphological mutants*	[47]
Productivity and quality	Siv'on and F-177*	[48]
Physiological variables and crop productivity	Siv'on and F-177*	[49]
Fiber quality	TM-1 and 3-79*	[25]
Yield components, fiber, flowering date, et al.	TM-1 and 3-79*	[50]
Rf1 fertility-restoring gene	CMS and the restoring lines	[51]
Fiber quality	Siv'on and F-177*	[52]
Fiber strength	7235 and TM-1	[53]
Rf1 fertility-restoring gene	XiangyuanA, ZMS12A, Sumian16A and 0-613-2R	[54]
Fiber-related traits	Acala-44 and Pima S-7*	[27]
Agronomic and fiber quality traits	MD5678ne and Prema	[55]
Fiber and yield traits	MARCABUCAG8US-1-88, HS46, MD5678ne et al.	[56]
Resistance to Verticillium wilt	Pima S-7 and Acala 44*	[57]
Fiber elongation	Tamcot 2111 and Pima S6*	[58]
Fiber length, length uniformity, and short fiber content	Tamcot 2111 and Pima S6*	[59]
Fiber fineness and micronaire (MIC)	Tamcot 2111 and Pima S6*	[60]
Li1, Li2, N1, Fbl, n2, sma-4(ha), and sma-4(fz)	Pima S-7, Li1, Li2, N1, Fbl,n2, SMA4, A1-97*	[61]
Leaf morphology	TMS-22 and WT936*	[31]
Leaf morphological traits and chlorophyll content	TM-1 and Hai 7124	[62]
Fiber quality traits	TM1 and Pima 3-79*	[35]
Leaf and stem pubescence	Guazuncho 2 and VH8-4602*	[63]
Fiber quality	Guazuncho 2 and VH8*	[64]
Lint percentage and fiber quality traits	Yumian 1 and T586	[65]
Fiber traits	Handan208 and Pima90*	[33]
Fiber yield and yield components	Handan 208 and Pima 90*	[66]
Fiber quality and yield component	Handan 208 and Pima 90*	[34]
Fiber traits	7235, TM-1, HS427-10, PD6992 and SM3	[67]
Root-knot nematode resistance gene	M-120 RNR and Pima S-6*	[68]
Fiber and yield component traits	7235 and TM-1	[69]
Fiber quality and yield components	7235 and TM-1	[70]
Root-knot nematode resistance gene *(rkn1)*	Acala SJ-2, Acala NemX, and Pima S-7*	[71]
Root-knot nematode resistance gene *(rkn1)*	Acala SJ-2 and Acala NemX	[72]
Root-knot nematode resistance gene	Resistant near isoline and susceptible near isoline	[73]
Lint percentage and morphological marker genes	TM-1 and T586	[74]
Fiber-related traits	TM-1 and 3-79*	[36]
Yield, yield component and fiber quality	Near-isogenicBC5S1 chromosome substitution lines, TM-1	[75]
Plant architecture traits	Zhongmiansuo12 and 8891	[76]
Fiber quality traits	Zhongmiansuo12 and 8891	[77]
Yield and yield-component traits	Zhongmiansuo12 and 8891	[78]

* Interspecific cross.

segregating population, and are readily subjected to variation of environments. With the increased availability of DNA markers for use in cotton genetic map construction in the last ten years, activities in identifying and locating quantitative trait loci (QTLs) have blossomed. QTLs that have been identified in cotton include yield and yield components, fiber quality, plant architecture, resistance to diseases such as bacterial blight and *Verticillium* wilt, resistance to pests like root-knot nematode, and flowering date. A list of QTLs mapped in cotton is presented in Table 2.

Table 3: Upland cotton BAC and BIBAC libraries that have been published or are accessible to the public (as of May 2007).

Genotype	Mean insert size (kb)	No. of clones	Genome equivalents	Vector[(a)]	Cloning site	References/locations where libraries are available
Tamcot HQ95	93	51,353	2.3x	pBeloBAC11	*Hin*dIII	http://hbz7.tamu.edu
Auburn 623	140	44,160	2.7x	pBeloBAC11	*Bam*HI	http://hbz7.tamu.edu
Texas Marker-1	130	76,800	4.4x	pCLD04541	*Bam*HI	http://hbz7.tamu.edu
Texas Marker-1	175	76,800	6.0x	pECBAC1	*Eco*RI	http://hbz7.tamu.edu
Maxxa	137	129,024	8.3x	pCUGI-1	*Hin*dIII	[79]
0-613-2R	130	97,825	5.7x	pIndigoBAC-5	*Hin*dIII	[80]

[(a)] The vectors, pBeloBAC11 (Kim et al. [81]), pECBAC1 (Frijters et al. [82]), pCUGI-1 [79], and pIndigoBAC-5 (http://www.epibio.com/item.asp?ID=328), are BAC vectors whereas pCLD04541 is plant-transformation-competent BIBAC vector (http://www.jic.bbsrc.ac.uk/staff/ian-bancroft/vectorspage.htm; [83]) that can be directly transformed into cotton plants via *Agrobacterium*.

Table 4: Summary of ESTs of major crops available in GenBank (as of April 27, 2007).

Species	Genome[(a)]	No. of ESTs
Cotton and related species (*Gossypium* species):		
G. hirsutum (Upland cotton)	$(AADD)_1$	177,154
G. raimondii	D_5D_5	63,577
G. arboreum	A_2A_2	39,232
G. barbadense (Sea Island)	$(AADD)_2$	1,023
G. herbaceum var. africanum	A_1A_1	247
Total:		**281,233**
Rice and Related Species (*Oryza* species):		
O. sativa (rice)	AA	1,211,447
O. minuta	BBCC	5,760
O. grandiglumis	CCDD	128
Total:		**1,217,335**
Maize and Related Species (*Zea* species):		
Z. mays (maize)		1,161,241
Total:		**1,161,241**
Wheat and Related Species (*Triticum* and *Aegilops* species):		
T. aestivum (wheat)	AABBDD	1,050,131
T. monococcum	AA	10,139
T. turgidum ssp. *durum*	AABB	8,924
T. turgidum	AABB	1,938
Ae. speltoides	BB	4,315
Ae. tauschii	DD	116
Total:		**1,075,563**
Soybean and Related Species (*Glycine* species):		
G. max (soybean)	GG	371,817
G. soja	GG	18,511
G. clandestina	A_1A_1	931
Total:		**391,259**

[(a)] There is no relationship in the genome letter designation between genera, but there is a relationship in the genome letter designation between species within a genus, the species with the same genome letter being closely related.

Several noteworthy findings have come out of QTL mapping in cotton. First, in tetraploid cottons, although the D-subgenome was derived from an ancestor that does not produce spinnable fibers, many QTLs influencing fiber quality traits were detected on the D-subgenome [106]. For example, Jiang et al. [45] pointed out that D-subgenome QTLs may partly explain the fact that domestication and breeding of tetraploid cottons has resulted in fiber with a higher quality than those achieved by parallel improvement of the A-genome diploid cottons which produce spinnable fibers. The merger of the A- and D-genomes in tetraploid cottons, where each genome has a different evolutionary history, may have offered unique avenues for phenotypic response to selection. Second, numerous studies have shown

Table 5: Summary of cotton ESTs (as of May 2007).

Genotypes	Library name	Tissues used	No. of ESTs	No. of unigenes	Authors	References
Gossypium arboreum (A_2):						
AKA8401	GA-Ea	7- to 10-dpa fibers (normalized)	46,603		Wing et al. Arpat et al.	
	Subtotal		**46,603**	**13,947**		[100]
G. Raimondii (D_5):						
GN34	GR_Ea	Whole seedlings with 1st true leaves	33,671		Udall et al.	
GN34	GR-Eb	−3-dpa buds to +3-dpa bolls	33,061		Udall et al.	
	Subtotal		**68,732**			
G. hirsutum (AD_1):						
Coker 312	GH_MD	8- to 10-dpa boll (irrigated)	1,144		Allen	
Coker 312	GH_MDDS	8- to 10-dpa boll (drought stressed)	1,238		Allen	
Coker 312	GH_LDI	15- to 20-dpa boll (irrigated)	1,799		Allen & Payton	
Coker 312	GH_LDDI	15- to 20-dpa boll (drought stressed)	1,409		Allen & Payton	
Acala Maxxa	GH_BNL	5-dpa fibers	8,022		Blewitt & Burr	[101]
Xu-142	GH_FOX	0- to 5-dpa ovule and 1- to 22-dpa fibers	7,997		Gou & Chen	
Deltapine 90	GH_SCW	2nd versus 1st primary fibers	7,385		Haigler & Wilkerson	
Zhongmian 12	GH_SUO	0-dpa ovules	1,240		Suo & Xue	
Deltapine 16	GH-CHX	−3- to 0-dpa ovules	7,631		Wu & Dennis	
Deltapine 16	GH_OCF	0-dpa ovules	867		Wu and Dennis	
Deltapine 16	GH_ON	0-dpa ovules (normalized)	5,903		Wu & Dennis	
Stv 7A gl	GH_ECT	18 h etiolated seedlings	2,880		Chapman	
Delta emerald	GH_CRH	Root and hypocotyls	1,464		Dowd & McFadden	
Delta emerald	GH_CFUS	RH tissues infected with *Fusarium oxysporum*	820		Dowd & McFadden	
Sicot	GH_LSL	S9i leaves, late season	1,810		Faivre-Nitschke & Dennis	
Coker 312	GH_SDL	Seedlings (control)	1,918		Klueva et al.	
Coker 312	GH_SDLD	Seedlings (drought stressed)	1,142		Klueva & Nguyen	
Coker 312	GH_SDCH	Seedling (chilling stressed)	576		Klueva & Nguyen	
Deltapine 16	GH_IME	Immature embryo	1,536		Liu & Dennis	
Im216	GH_IMX	Leaf 8, 14, 20, 30, 45, 60 hpi *Xanthomonas*	1,134		Patil et al.	
AcB4Blnb7	GH_ACXE	Leaf 8 + 14 hpi *Xanthomonas*	647		Phillips et al.	
AcB4Blnb7	GH_ACXM	Leaf 20 + 30 hpi *Xanthomonas*	1,328		Phillips et al.	
AcB4Blnb7	GH_ACXL	Leaf 45 + 60 hpi *Xanthomonas*	862		Phillips et al.	

Table 5: Continued.

Genotypes	Library name	Tissues used	No. of ESTs	No. of unigenes	Authors	References
T25	GH_pAR	Leaves	1,230		Trolinder	
DES119	GH_STEM	Mature stem	8,643		Taliercio	
DP62	GH_ECOT	Etiolated cotyledon	2,772		Ni & Trelease	
91-D-92	GH_CBAZ	Ball abscission zone	1,306		Wan & Wing	
			185,198[(a)]	51,107		[102]
Texas Marker-1 −1 (TM-1)	GH_TMO	−3- to 3-dpa ovules	32,789	8,540	Chen	[103]
Xuzhou 142 (Xu-142)	Not available	5- to 10-dpa fibers	29,992	12,992	Zhu	[104]
	Subtotal		**132,644**			
	Total		**247,979**			

[(a)] The number was the sum of numbers of all ESTs above the line including those of *G. arboreum*, *G. raimondii* and *G. hirsutum* [102]. Of the 247,979 cotton ESTs, 187,014 (75.4%) were from developing fibers or ovules whereas 160,965 (24.6%) from nonfiber or nonovule organs.

that QTLs occur in clusters genetically in the cotton genome [27, 46, 55, 56, 76, 106]. Ulloa et al. [56] suggested the possible existence of highly recombined regions in the cotton genome with abundant putative genes. QTL clusters might exert their multiple functions to compensate for a numerical deficiency, expanding their roles in cotton growth and development [76]. Finally, the position and effect of QTLs for fiber quality are not comparable in different populations and environments evaluated [60, 106]. This suggests that QTL studies conducted thus far have detected only a small number of loci for fiber growth and development and that additional QTLs remain to be discovered [58, 59]. Furthermore, because quantitative traits are readily subjected to variation of environments, mapping efforts of these traits need to be pursued in multiple environments including years and locations.

2.3. BAC and BIBAC resources

Large-insert BAC and BIBAC libraries have been demonstrated essential and desirable for advanced genomics and genetics research [107–111]. Because of their low-level chimerism, readily amenability to high-throughput purification of cloned insert DNA, and high stability in the host cell [83, 112, 113], BACs and BIBACs have quickly assumed a central position in genome research. BAC and BIBAC libraries have widely been used in many research areas of genomics and molecular biology, including whole-genome or chromosome physical mapping [110, 114–128], large-scale genome sequencing [129–133], positional cloning of genes and QTLs (for review, see [134]), isolation and characterization of structural and regulatory genes [135, 136], long-range genome analysis [135, 136], organization and evolution of multigene families [136], and cytologically physical mapping [137].

BAC libraries have been developed for a number of species, including plants, animals, insects, and microbes and made available to the public (http://hbz7.tamu.edu; http://bacpac.chori.org; http://www.genome.clemson.edu). To facilitate cotton genome research, BAC and BIBAC libraries have been developed for several genotypes of Upland cotton, *G. hirsutum* (Table 3). As of May 1, 2007, at least six BAC and BIBAC libraries have been developed and made available to the public. These libraries were constructed from five genotypes of Upland cotton, including Tamcot HQ95, Auburn 623, TM-1, Maxxa, and 0-613-2R, in four BAC vectors and one *Agrobacterum*-mediated, plant-transformation-competent BIBAC vector with three restriction enzymes. The libraries have average insert sizes ranging from 93 to 175 kb and each have a genome coverage ranging from 2.3 to 8.3x genome equivalents, collectively covering >21x haploid genomes of the polyploid cotton (Figure 2). Moreover, BAC libraries have also been constructed for several other *Gossypium* species, including *G. barbadense* (Pima S6), *G. arboreum* (AKA8401), *G. raimindii*, and *G. longicalyx* (A.H. Paterson, pers. communication). These BAC and BIBAC libraries provide resources essential for advanced genomics and genetics research of cottons.

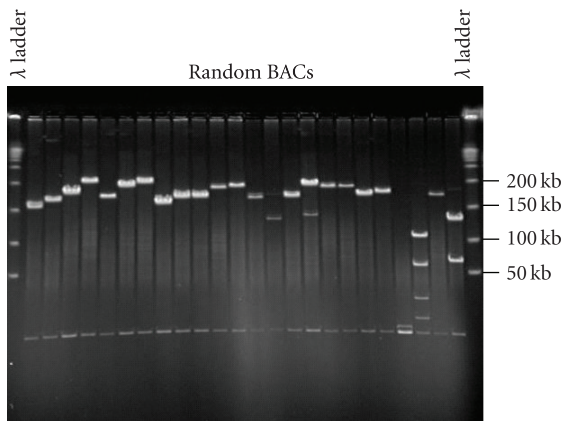


Figure 2: BACs randomly selected from the TM-1/*Eco* RI BAC library (see Table 3; C. Scheuring and H.-B. Zhang, unpublished). BAC DNA was isolated, digested with *Not*I, and run on a pulsed-field gel.

2.4. ESTs, microarrays, and gene expression profiling

ESTs

Cloning and sequencing of expressed gene sequence tags (ESTs) by single sequencing pass from one or both ends of cDNA clones have been widely used to rapidly discover and characterize genes in a large-scale and high-throughput manner. As have been done in many other plant and animal species of biological and/or economical importance, significant efforts have been made to generate ESTs in cottons. As of April 27, 2007, 281,233 ESTs have been available for the *Gossypium* species in GenBank (Table 4; http://www.ncbi.nlm.nih.gov/dbEST). Of these ESTs, 178,177 were from the polyploid cultivated cottons with 177,154 (63.0%) from *G. hirsutum* and 1,023 (<1.0%) from *G. barbadnese* while 103,056 were from the related diploid species with 39,232 (13.9%) from *G. arboreum* (A_2), 63,577 (22.6%) from *G. raimondii* (D_5), and 247 (<1.0%) from *G. herbaceum* (A_1). This number of cotton ESTs, compared with that of five years ago, has been significantly increased, due to several large EST projects funded [100–104]. Nevertheless, when compared with those of other major crop species such as rice, maize, wheat, and soybean, the number of the cotton ESTs is very low, only being about one-forth of those of rice, maize, or wheat (Table 4).

Table 5 summaries 247,979 ESTs of cottons published [100, 102–104]. These ESTs were collectively generated from 32 cDNA libraries constructed from mRNA isolated from 18 genotypes of three species, *G. hirsutum*, *G. arboreum,* and *G raimondii*, by one-pass sequencing of cDNA clones from one (3′ or 5′ end) or both ends. They were generated from 12 different organs, including developing fibers, seedlings, buds, bolls, ovules, roots, hypocotyls, immature embryos, leaves, stems, and cotyledons. Some of the ESTs were generated from plants growing under biotic or abiotic stress conditions such as drought, chilling, and pathogens. By analyzing approximately 185,000 ESTs from both fibers/ovules (124,299 ESTs) and nonfiber/ovule tissues (60,899 ESTs) of *G. hirsutum*, *G. arboreum* and *G. raimondii*, Udall et al. [102] obtained 51,107 unigenes. A few months later, Yang et al. [103] analyzed their 32,789 ESTs generated from −3- to +3-dpa fibers of Upland cotton cv. TM-1, along with 211,397 cotton ESTs downloaded from GenBank (as of April 2006), resulting in 55,673 unigenes and updating The Institute of Genomic Research Cotton Gene Index version 6 (CGI6) into CGI7 (http://www.tigr.org). The unigene EST number may provide a reasonable estimation about the number of expressed genes in the cotton genomes. Of the unigene set, those derived from fibers or fiber-bearing ovules suggest the number of genes potentially involved in fiber development and genetic complexity of fiber traits.

A predominant feature of the cotton EST set is the significant preference of their tissue sources for fiber or fiber-bearing ovules than other organs. Of the 247,979 ESTs listed in Table 5, 187,080 (75.4%) were from developing fibers or fiber-bearing ovules while only 60,899 (24.6%) were from nonfiber and nonovule organs. Within each of the two categories, fiber/fiber-bearing ovules and nonfiber/ovule organs, there is also a significant bias in the number of ESTs. Cotton fiber development is classified into four clearly characterized, but overlapping stages, including fiber initiation (−3- to 5 dpa), elongation (5–25 dpa), secondary cell wall deposition (15–45 dpa), and maturation/dehydration (45–70 dpa) (see Figure 3). All of the 187,080 fiber ESTs were generated from the fibers or fiber-bearing ovules collected from the first three stages with 43.6% from the initiation stage, 46.5% from the elongation stage, and 5.7% from the secondary cell wall deposition stage. It is apparent that the number of fiber ESTs from the secondary cell wall deposition stage is much smaller than that of either initiation or elongation stage. Although the initiation and elongation stages are of significance for the number of fibers per seed and fiber length, the secondary cell wall deposition stage is crucial to fiber strength. Of the 60,899 nonfiber/nonovule ESTs, 66.0% were from seedlings, 14.2% from stems, and 2.4% from roots.

The cotton ESTs have been used in several aspects, including development of genome-wide cotton microarrays (see below), mining of SSRs (see above) and study of polyploidization. The development of the significant numbers of ESTs from the cultivated tetraploid cotton, *G. hirsutum* $[(AADD)_1]$, and its closely related diploid species, *G. arboreum* (A_1A_1) and *G. raimondii* (D_5D_5) (see Table 5) made it possible to compare the transcriptomes among the three species. Udall et al. [102] comparatively analyzed 31,424, 68,732, and 69,853 ESTs derived from *G. arboreum*, *G. raimondii*, and *G. hirsutum*, respectively. Although the comparison was significantly affected by the tissue sources and developmental status, they identified the putative homoeologs among the four genomes, A, D, A_1, and D_5. This information is useful for our understanding of how the cotton genomes function and evolve during the courses of speciation, domestication, plant breeding, and polyploidization.

Microarray

Microarray has been a technology that is widely used in many aspects of genomics research, including gene discovery, gene expression profiling, mutation assay, high-throughput genetic mapping, gene expression mapping (eQTL mapping), and comparative genome analysis. It involves robotically printing tens of thousands of cDNA amplicons or gene-specific long (70 mers) oligonucleotides as array elements on a chemically-coated glass slide, followed by hybridizing the array with one or more fluorescent-labeled cDNA or cRNA targets derived from mRNA isolated from particular tissues, organs, or cells. Therefore, it allows the simultaneous monitoring of the expression/activities of all genes arrayed on the array in a single hybridization experiment. To facilitate cotton genomics research, microarrays have been developed from the cotton ESTs (Table 5) in several laboratories worldwide.

The first batch of cotton microarrays was fabricated from 70-mers oligos designed from the 7–10 dpa fiber nonredundant (NR) or unigene ESTs of *G. arboreum* (Table 5) ([100]; http://cottongenomecenter.ucdavis.edu/microarrays.asp). Each microarray consists of 12,227 elements corresponding to 12,227 NR fiber ESTs, with a duplicate of each element.

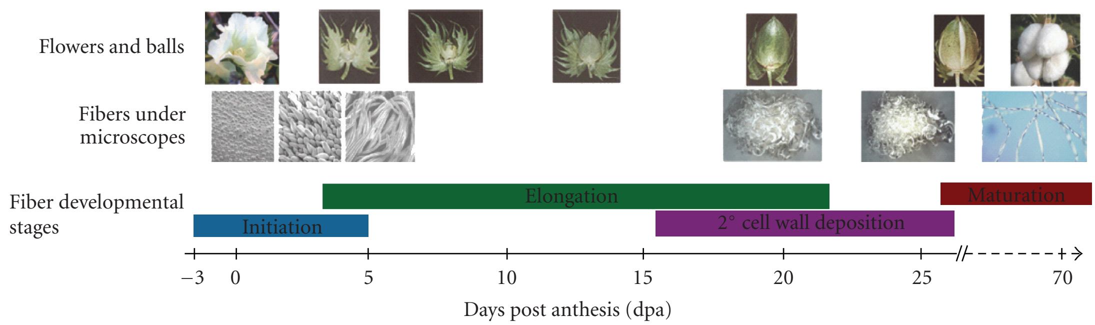


Figure 3: Cotton fiber development and corresponding morphogenesis stages (according to [138, 139]). The initiation stage is characterized by the enlargement and protrusion of epidermal cells from the ovular surface; during the elongation stage the cells expend in polar directions with a rate of >2 mm/day; during the secondary cell wall deposition stage celluloses are synthesized rapidly until the fibers contain ∼90% of cellulose; and at the maturation stages minerals accumulate in the fibers and the fibers dehydrate.

Using the microarrays, Arpat et al. [100] compared the expression of the genes between 10-dpa fibers at elongation or primary cell wall synthesis stage and 24-dpa fibers at secondary cell wall disposition stage (see Figure 3). The expression of fiber genes was found to change dynamically from elongation or primary cell wall to secondary cell wall biogenesis, with 2,553 of the fiber genes being significantly downregulated and 81 being significantly upregulated. This result suggests that the expression of fiber genes is stage-specific or cell expansion-associated. Annotation of the genes upregulated in the secondary cell wall synthesis relative to the primary cell wall biogenesis showed that most of the genes felt in three major functional categories, energy/metabolism, cell structure, organization and biogenesis, and cytoskeleton. This finding is consistent with the fact of massive cellulose synthesis and cell wall biogenesis during this stage. The fiber gene microarrays have been updated recently by incorporating nearly 10,000 gene elements designed from the fiber and ovary ESTs of the tetraploid cultivated cotton, *G. hirsutum* (Table 5; T.A. Wilkins, pers. communication). The current fiber microarrays each slide consist of four duplicated arrays with 22,406 60-mers oligo elements per array and a duplicate of each element (see, e.g., Figure 4). The new version of fiber gene arrays covers 100% of the fiber ESTs of diploid cotton and 65% of the fiber ESTs of the tetraploid cultivated cotton, *G. hirsutum* that are available in GenBank, thus representing the most comprehensive coverage of the cotton fiber genes. The elements are printed on a slide in a randomized manner instead of the conventional ordered manner. The fabrication of four duplicated arrays per slide and randomized printing design have significantly minimized the systematic problems that are frequently encountered in the conventional array design (one array per slide and ordered printing), thus further enhancing the reproducibility and accuracy of the microarray analysis results.

Recently, several additional batches of EST- or cDNA-based microarrays with different formats and elements have been reported in cotton [102, 104, 140]. Shi et al. [104] reported the fabrication of microarrays from unigene ESTs derived from 5–10 dpa ovules of the Upland cotton cv. Xuzhou 142 and using the amplicons of the EST clones as the array elements. The microarrays each consist of 11,962 uniEST elements. Using the microarrays, Shi et al. [104] comparatively studied the wild-type Xuzhou 142 versus its *fuzzless-lintless* (*fl*) mutation using the RNAs isolated from the ovules at stages of 0-, 3-, 5-, 10-, 15-, and 20-dpa. It was found that ethylene biosynthesis is one of the most significantly upregulated biochemical pathways during fiber elongation. Similarly, Wu et al. [140] also fabricated a set of microarrays from amplicons of 10,410 cDNA clones derived from −3- to 0-dpa ovules of the Upland cotton cv. DP16 (see Table 5, Wu & Dennis). The arrays were analyzed with RNAs isolated from 0-dpa whole ovules, outer integument, and inner integument/nucellus of five lintless mutation lines against the wild-type DP16. Of the 10,410 gene elements on the array, 60 to 243 were found to significantly differentially express between each pair of the wild type and mutant when the array was hybridized with the RNAs isolated from the 0-dpa whole ovules. Of these differentially expressed genes, 70.6% were upregulated and 29.4% downregulated in the fiber mutant, suggesting that the mutation caused not only gene downregulation, but also gene upregulation. However, when the whole ovule was dissected into three layers, outer integument, inner integument, and nucellus, of which cotton fibers develop from the epidermal cells of the outer integument, and analyzed with the outer integument against the inner integument and the nucellus, the number of the genes downregulated in the mutants was reduced to 13. These include an *Myb* transcription factor, a putative *homeodomain* protein, a *cyclin D* gene, and some fiber-expressed structural and metabolic genes, suggesting that these genes may be involved in the process of fiber initiation.

In summary, three batches of EST- or cDNA-based cotton microarrays were fabricated from fiber genes of either cultivated tetraploid cotton, *G. hirsutum* [104, 140], or cultivated diploid cotton, *G. arboreum* [100]. Using the microarrays, the expression of the fiber genes was profiled and comparatively analyzed at fiber initiation stage [140], elongation stage [100, 104], and secondary cell wall deposition stage [100]. However, the expression of other cotton genes such as those from nonfiber and nonovary tissues remains to profile. To fill this gap, another two batches of long oligo-based microarrays have been developed. The first batch contains approximately 21,000 gene elements per array (http://cotton.agtec.uga.edu/CottonFiber/pages/mcriarray/Array.aspx). These genes were from 52 cDNA libraries constructed from a variety of tissues and organs in a range of conditions, including drought stress and pathogen challenges, and represents tetraploid (*G. hirsutum*) and its diploid relatives (*G. arboreum* and *G. raimondii*). Of the 21,000 genes, approximately one-forth were from fiber genes and three-forth were from nonfiber and nonovary tissues (J. A. Udall, pers. communication). The second batch contains 38,716 gene elements per array. Of the gene elements, 22,409 are designed from fiber ESTs and 16,307 from nonfiber ESTs (T.A. Wilkins, pers. communication). There is no doubt that these versions of cotton microarrays will provide new tools for comprehensive functional and comparative genomics research of cottons.

2.5. Physical mapping

Whole-genome, BAC- and/or BIBAC-based, integrated physical/genetic maps have played a central role in genomics research of humans, plants, animals, and microbes [110, 123, 127]. This is because they provide central platforms for many areas, if not all, of modern genomics research, including large-scale transcript or gene mapping, region-targeted marker development for fine mapping and MAS of genes and QTLs, map-based gene/QTL cloning, local- and whole-genome comparative analysis, genome sequencing, and functional analysis of DNA sequences and component network. Therefore, whole-genome, BAC/BIBAC-based, integrated physical/genetic maps have been developed for a number of plant and animal species. In plants, whole-genome BAC physical maps have been developed for several species, including *Arabidopsis* [114, 118], indica rice [121], japonica rice [117], soybean [124], and maize [141]. However, whole-genome physical maps of cottons have only been initiated in several laboratories. One is the laboratory of H.-B. Zhang, Texas A&M University, College Station (Texas, USA). This laboratory is developing a whole-genome BAC/BIBAC physical map of the Upland cotton cv. TM-1 by using the latest physical mapping technology [123, 126]. The project was a collaborative effort among the laboratories of H.-B. Zhang, R. J. Kohel, USDA/ARS, College Station (Texas, USA) (who provided a part of the fund for the project), and D. M. Stelly, Texas A&M University (Texas, USA). Nearly 120,000 (~7.3x) BIBACs and BACs selected from the TM-1 BIBAC and BAC libraries (see Table 3) have been fingerprinted and a draft BAC/BIBAC contig map has been constructed. The draft physical map consists of 5,088 contigs collectively spanning approximately 2,300 Mb of the 2,400 Mb Upland genome (unpublished). Currently, additional clones (to reach about 10x genome coverage clones) are being analyzed. Furthermore, because the Upland cotton is an allotetraploid which makes the physical map construction more complicated, several approaches are being used to sort the map contigs according to their origin of subgenomes. The laboratory of A. H. Paterson, University of Georgia (Athens, Georgia) is also working toward development of a whole-genome BAC-based physical map of the diploid species, *G. raimondii* (A. H. Paterson, pers. communication). Given the importance of physical maps for modern genome research, there is no doubt that development of a robust integrated physical/genetic map will greatly promote advanced genomics research of cottons and related species (also see below).

2.6. Genome sequencing

Sequence maps represent the most-fine physical maps of genomes [108]. They provide not only physical positions of and distances between genes and other components constituting a genome [142], but also their sequences and putative functions inferred from the sequences. Therefore, development of a complete genome sequence map of a species will significantly promote genomics research of the species in a variety of aspects. Because of this reason, the whole genomes of several plant and animal species have been sequenced. In plants, the genomes of two model species, *Arabidopsis* [130] and rice [132], have been completely sequenced and the genomes of several other species, including *Medicago truncatula* (http://www.medicago.org), *Lotus japonicus* (http://www.kazusa.or.jp/lotus), tomato (http://www.sgn.cornell.edu/about/tomato_sequencing.pl), maize (http://www.maizegenome.org), and soybean (http://genome.purdue.edu/isgc/Tsukuba07/ISGC_report_Apr2007.htm), are currently being sequenced.

However, there is only a limited amount of genomic sequences available for cotton and related species in GenBank. A major source of the genomic sequences of *Gossypium* species was from Hawkins et al. [143]. To understand the underlying genome size variation and evolution of *Gossypium* species, Hawkins et al. [143] constructed whole-genome shotgun libraries for *G. raimondii* (D_5D_5), *G. herbaceum* (A_1A_1), *G. exiguum* (KK), and the species that was used as the outgroup species for phylogenetic analysis of the *Gossypium* species, *Gossypioides kirkii*, with each species library containing 1920–10,368 clones. From each of the four shotgun libraries, 1,464–6,747 clones were sequenced, together covering a total length of 11.4 Mb. Annotation of these clone sequences and estimation of the copy number of each type of the sequences suggested that differential lineage-specific amplification of transposable elements is responsible for genome size variation in the *Gossypium* species. Moreover, *G. raimondii* has been selected recently by the DOE Joint Genome Institute, U.S. Department of Energy to be sequenced for genomic study of cotton and related species (http://www.jgi.doe.gov/sequencing/cspseqplans2007.html). At the first phase of the sequencing project, a whole-genome shotgun library covering about 1x of the *G. raimondii* genome will be sequenced. While this number is far from the genome coverage of clones (>6x) that is needed to assemble the sequence map of the genome, it will provide the first glimpse into the cotton genome and useful information for sequencing the entire genomes of this and other cotton key species efficiently.

3. APPLICATIONS OF GENOMIC TOOLS IN COTTON GENETIC IMPROVEMENT

One of the major goals of genome research is to use the genomic tools developed to promote or assist continued crop genetic improvement. In cottons, the development of the genomic resources and tools has allowed addressing many significantly scientific questions that are impossible to do so before. These include, but not limited to, construction of genome-wide genetic maps (Table 1), identification and mapping of genes and loci controlling traits underlying qualitative and quantitative inheritance (Table 2), determination of mechanisms of cotton genome evolution, and identification and determination of genes that are involved in cotton fiber initiation, elongation, and secondary cell wall biogenesis. The genomic resources and tools could be used to promote or facilitate cotton genetic improvement in numerous ways. Marker-assisted selection (MAS) is likely one of the most important and practical applications at present time and in near future. The MAS technology could offer many potential benefits to a breeding program. For instance, DNA linked to a gene of interest could be utilized in early generation of breeding cycle to improve the efficiency of selection. This approach has a particular advantage when screening for phenotypes in which the selection is expensive or difficult to perform, as is the case involving recessive or multiple genes, seasonal or geographical considerations, and late expression of the phenotype [144]. However, application of MAS in cotton breeding programs is still in its infancy as the major effort of cotton genome research in the past has been on the development of genomic resources and tools for the eventual goal of enhanced cotton genetic improvement.

Fiber quality

Zhang et al. [53] used a *G. anomalum* introgression line 7235 with good fiber quality properties to identify molecular markers linked to fiber-strength QTLs. A major QTL, *QTLFS1*, was detected at the Nanjing and Hainan field locations (China) and College Station, Texas, (USA). This QTL was associated with eight markers and explained more than 30% of the phenotypic variation. *QTLFS1* was first thought to be mapped to chromosome 10, however, further study showed that this QTL was located on LGD03 [67]. Guo et al. [145] showed that the specific SCAR4311920 marker could be applied to large-scale screening for the presence or absence of this major fiber strength QTL in breeding populations. The DNA markers tightly linked to this QTL could be useful for developing commercial cultivars with enhanced fiber length properties [67].

Wang et al. [76] identified a stable fiber length QTL, *qFL-D2-1*, simultaneously in four environments in Xiangzamian 2. The high degree of stability suggests this QTL might be particularly valuable for use in MAS programs. Chee et al. [59] dissected the molecular basis of genetic variation governing 15 parameters that reflect fiber length by applying a detailed RFLP map to 3,662 BC_3F_2 plants from 24 independently derived BC_3 families utilizing *G. barbadense* as the donor parent. The discovery of many QTLs unique to each trait indicates that maximum genetic gain will require breeding efforts that target each trait. Lacape et al. [64] performed QTL analysis of 11 fiber properties in BC_1, BC_2, and BC_2S_1 backcross generations derived from the cross between *G. hirsutum* "Guazuncho 2" and *G. barbadense* "VH8." They detected 15, 12, 21, and 16 QTLs for length, strength, fineness, and color, respectively, in one or more populations. The results showed that favorable alleles came from the *G. barbadense* parent for the majority of QTLs, and cases of colocalization of QTLs for different traits were more frequent than isolated positioning. Taking these QTL-rich chromosomal regions into consideration, they identified 19 regions on 15 different chromosomes as target regions for the marker-assisted introgression strategy. The availability of DNA markers linked to *G. barbadense* QTLs promises to assist breeders in transferring and maintaining valuable traits from exotic sources during cultivar development.

Cytoplasmic male sterility

In cotton, cytoplasmic male sterility conditioned by the D8 alloplasm (CMS-D8) is independently restored to fertility by its specific D8 restorer (D8R) and by the D2 restorer (D2R) that was developed for the D2 cytoplasmic male sterile alloplasm (CMS-D2). Zhang and Stewart [146] concluded that the two restorer loci are nonallelic, but are tightly linked with an average genetic distance of 0.93 cM. The D2 restorer gene is redesignated as *Rf1*, and *Rf2* is assigned to the D8 restorer gene. The identification of molecular markers closely linked to restorer genes of the cytoplasmic male sterile could facilitate the development of parental lines for hybrid cotton. Guo et al. [147] found that one RAPD marker fragment, designated OPV-15(300), was closely linked with the fertility-restoring gene *Rf1*. Zhang and Stewart [148] identified RAPD markers linked to the restorer gene and, furthermore, converted the three RAPD markers into reliable and genome-specific sequence tagged site (STS) markers. Liu et al. [51] determined that the *Rf1* locus is located on the long arm of chromosome 4. Two RAPD and three SSR markers were identified to be closely linked to the *Rf1* gene. These markers are restorer-specific and should be useful in MAS for developing restorer parental lines. Yin et al. [54] further constructed a high-resolution genetic map of *Rf1* containing 13 markers in a genetic distance of 0.9 cM. They constructed a physical map for the *Rf1* locus and enclosed the possible location of the *Rf1* gene to a minimum of two BAC clones spanning an interval of approximately 100 kb between two clones, designated as 081-05K and 052-01N. Work to isolate the *Rf1* gene in cotton is now in progress.

Resistance to diseases and insect pests

Breeding for disease resistance is of great importance in cotton breeding program. To facilitate analysis, cloning, and manipulation of the genes conferring resistance to different pathogens, including bacteria, fungi, viruses, and nematodes, He et al. [149] isolated and characterized the family of nucleotide-banding site-leucine-rich repeat (NBS-LRR)-encoding genes or resistance gene analogues (RGAs) in the

Upland cotton cv. Auburn 634 genome. Genetic mapping of a sample (21 genes) of the RGAs indicated that the gene family resides on a limited number of the cotton AD-genome chromosomes with those from a single subfamily tending to cluster on the cotton genetic map and more RGAs in the A subgenome than in the D subgenome. Of the 16 RGAs mapped, two happened to be comapped with the cotton bacterial blight resistance QTLs previously mapped by Wright et al. [42]. Since nearly 80% of the genes (>40 genes) cloned to date that confer resistance to bacteria, fungi, viruses, and nematodes are contributed by the NBS-LRR gene family, the cotton RGAs of the NBS-LRR family have provided valuable tools for cloning, characterization, and manipulation of the resistant genes to different pathogens and pests in cottons.

Root-knot nematodes (RKN), *Meloidogyne incognita*, can cause severe yield loss in cotton. Wang et al. [71] identified one SSR marker CIR316 on the linkage group A03 tightly linked to a major RKN resistant gene (*rkn1*) in the resistant cultivar *G. hirsutum* "Aacla NemX." In a companion study, a bulked segregant analysis (BSA) combined with AFLP was used to identify additional molecular markers linked to *rkn1* [72]. An AFLP marker linked to *rkn1* designated as GHACC1 was converted to a cleaved amplified polymorphic sequence (CAPS) marker. These two markers have potential for utilization in MAS. Shen et al. [68] identified RFLP markers on chromosome 7 and chromosome 11 showing significant association with RKN resistance from the Auburn 634 source, a different source of resistant germplasm than Acala NemX. The association was further confirmed by detection of a minor and a major dominant QTL on chromosomes 7 and 11, respectively, using SSR markers. Ynturi et al. [73] identified two SSR markers which together accounted for 31% of the variation in galling index. The marker BNL 3661 is mapped to the short arm of chromosome 14 while BNL 1231 to the long arm of chromosome 11. The association of two different chromosomes with RKN resistance suggests at least two genes are involved in resistance to RKN.

Bacterial blight caused by the pathogen *Xanthomonas campestris* pv. *malvacearum* (*Xcm*) is another economically important disease in cotton. Wright et al. [42] and Rungi et al. [43] both used mapped RFLP markers to investigate the chromosomal location of genes conferring resistance to the bacterial blight pathogen. The mapping data suggest that the resistance locus segregates with a marker on chromosome 14 known to be linked to the broad-spectrum *B12* resistance gene originally from African cotton cultivars. AFLPs and SSRs were also used to search for novel markers linked to the *Xcm* resistance locus to facilitate introgression of this trait into *G. barbadense* through MAS.

4. CONCLUDING REMARKS

A significant amount of genomic resources and tools has been available in cottons though cotton genomics research is far behind those of other major crops such as rice, maize, wheat, and soybean. These resources and tools have allowed identifying and mapping many genes and QTLs of importance to cotton fiber quality, fiber yield, and biotic and abiotic stresses and addressing several significant questions to plant biology in general and to cotton in particular. Nevertheless, many efforts are needed to further develop the resources and tools and to make the tools readily usable in applications in order to fully and effectively use them in cotton genetic improvement and biology research. In particular, the following areas of cotton genomics research should be emphasized.

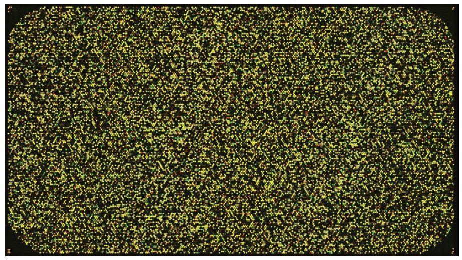


FIGURE 4: Comparative analysis of the expression of fiber genes in 10-dpa fibers between *G. hirsutum* and *G. barbadense* (M. Goebel, M. Alabady, C.W. Smith, T. A. Wilkins and H.-B. Zhang, unpublished). The cotton fiber microarrays are available in the laboratory of T. A. Wilkins, Texas Tech University (Texas, USA). One of the four arrays printed on a single slide is shown.

(i) *Development of whole-genome BAC/BIBAC-based, integrated physical maps of cottons*. There is no whole-genome, robust BAC/BIBAC-based, integrated physical/genetic map that has been developed for cottons. The maps should be developed for at least two species of *Gossypium*. One is the Upland cotton that produces >90% of the world's cotton whereas the other is *G. raimondii*, the species having the smallest genome among all *Gossypium* species, thus likely having highest density of genes. This research is emphasized because it has been proven in model and other species, including *Arabidopsis*, rice, *Drosophila*, human, mouse, and chicken, that whole-genome integrated physical/genetic maps provide powerful platforms and "freeways" for many, if not all, modern genetics and genomics research ([110, 123, 127]; see above). These include not only genome sequencing (see below), but also development of closely linked, user-friendly DNA markers for any region or loci of the genome, fine QTL and gene mapping (see below), map-based gene/QTL cloning, and high-throughput and high-resolution transcript (unigene EST) mapping [150]. Development of the integrated physical maps will allow rapidly and efficiently integrating all existing genetic maps, mapped genes and QTLs, and BAC and BIBAC resources and cotton unigene ESTs, and accelerate the efficiency and reduce the cost of research in all areas by manifold.

(ii) *QTL fine mapping*. Many genes and QTLs that are important to cotton fiber yield fiber quality, and biotic and abiotic stresses have been genetically mapped, but two problems are apparent. The first one is that almost all of the QTLs were mapped using F_2, BC_1, or early segregating generations in a single or a limited number of environments (Table 2). Since quantitative traits are readily subjected to environmental variation, the mapping results using the early generations in a single or a limited number of environments would vary from experiments to experiments [59, 60, 106]. The other problem is that the genetic distances between DNA markers and most of the QTLs are too far to be used for MAS. Therefore, it is of significance to fine map the QTLs using large and advanced generation or homozygous populations, such as RILs and DHs, in multiple environments, and closely linked DNA markers, for which advantage of integrated physical maps could be taken. In addition to accurate mapping of the QTLs and development of DNA markers that are well-suited (closely linked and user-friendly) for MAS, fine mapping is also an essential step toward the final isolation of the QTL genes by map-based cloning [134]. The isolated genes are not only the sources for molecular breeding via genetic transformation, but also the most desirable for marker development for MAS because there is no recombination between the gene and its derived marker.

(iii) *Sequencing of one or more key cotton genomes*. While it is costly using the current sequencing technology, whole-genome sequencing is a most-efficient method to discover and decode all cotton genes and provides a most-desired and most-fine integrated physical and genetic map of the cotton genome. Comparative genomics studies demonstrated that the gene contents and orders are highly conserved among the genomes of *Gossypium* species even they are significantly different in genome size [16, 24]. Based on this result, *G. raimondii* is an excellent choice to be sequenced because it has the smallest genome among all *Gossypium* species though it is not cultivated. If an integrated physical map is available for the major cultivated cotton, *G. hirsutum*, that has a three-fold larger genome than *G. raimondii*, the sequence information of *G. raimondii* could be readily transferred to the cultivated cotton by using the BAC end sequences of its integrated physical map as anchors.

(iv) *ESTs from nonfiber and nonovary tissues and fibers at the secondary cell wall deposition stage*. As shown above, the number of cotton ESTs available in GenBank has been increased significantly in the past few years; however, the distribution of the ESTs among tissue sources are extremely biased. The numbers of ESTs from both nonfiber/nonovary tissues and fibers at the secondary cell wall deposition stage (15–45 dpa), particularly after 20 dpa, are especially small. The former set of expressed genes, despite of not directly contributing to fiber yield and quality, is of significance to fiber yield and quality, whereas there is no doubt that the later set of expressed genes directly contribute to the fiber strength.

(v) *Profiling and identification of genes involved in individual biological processes and conditions with emphasis on those involved in fiber development*. Development and availability of cDNA- or unigene EST-based microarrays have provided unprecedented opportunities for research of molecular biology, functional genomics, and evolutionary genomics, however, cotton research in these regards are very limited. Identifying and characterizing genes that are involved in the processes of fiber development, plant growth and development, and biotic and abiotic stresses will greatly facilitate our understanding of underlying molecular basis of these processes in cottons, and thus, enhance breeders' ability to cotton genetic improvement.

(vi) *Translating the gene activities or expressions at different tissues and stages into fiber yield and fiber quality, thus assisting in cotton breeding*. The genes that are involved in fiber initiation [104, 140], elongation [100, 104], and secondary cell wall deposition [100] have been identified from several genotypes of cottons, but it is unknown about what the up- or downregulation, or active expression of fiber genes at a developmental stage and organ means to final fiber yield and/or quality. For instance, does the active expression of a gene at fiber elongation stage in fiber suggest longer fibers? Studies in this regard are essential to use the gene expression data in cotton germplasm analysis and breeding.

REFERENCES

[1] J. E. Endrizzi, E. L. Turcotte, and R. J. Kohel, "Qualitative genetics, cytology, and cytogenetics," in *Cotton*, R. J. Kohel and C. F. Lewis, Eds., pp. 81–129, American Society of Agronomy, Madison, Wis, USA, 1984.

[2] J. F. Wendel and R. C. Cronn, "Polyploidy and the evolutionary history of cotton," *Advances in Agronomy*, vol. 78, pp. 139–186, 2003.

[3] B. Hendrix and J. McD. Stewart, "Estimation of the nuclear DNA content of *Gossypium species*," *Annals of Botany*, vol. 95, no. 5, pp. 789–797, 2005.

[4] J. E. Endrizzi, E. L. Turcotte, and R. J. Kohel, "Genetics, cytology, and evolution of *Gossypium*," *Advances in Genetics*, vol. 23, pp. 271–375, 1985.

[5] A. E. Percival, "The national collection of *Gossypium germplasm*," Department of Agriculture Comm., So. Crops, College Station, USA, Ser. Bull. No. 321, 1987.

[6] M. D. Bennett and I. J. Leitch, "Nuclear DNA amounts in angiosperms," *Annals of Botany*, vol. 76, pp. 113–176, 1995.

[7] I. J. Leitch and M. D. Bennett, "Polyploidy in angiosperms," *Trends in Plant Science*, vol. 2, pp. 470–476, 1997.

[8] J. Masterson, "Stomatal size in fossil plants: evidence for polyploid in majority of angiosperms," *Science*, vol. 264, pp. 421–424, 1994.

[9] L. G. Lundin, "Evolution of the vertebrate genome as reflected in paralogous chromosomal regions in man and the house mouse," *Genomics*, vol. 16, no. 1, pp. 1–19, 1993.

[10] J. H. Postlethwait, Y.-L. Yan, M. A. Gates, et al., "Vertebrate genome evolution and the zebrafish gene map," *Nature Genetics*, vol. 18, no. 4, pp. 345–349, 1998.

[11] A. Sidow, "Gen(om)e duplications in the evolution of early vertebrates," *Current Opinion in Genetics & Development*, vol. 6, no. 6, pp. 715–722, 1996.

[12] J. Spring, "Vertebrate evolution by interspecific hybridisation—are we polyploid?" *FEBS Letters*, vol. 400, no. 1, pp. 2–8, 1997.

[13] J. F. Wendel, A. Schnabel, and T. Seelanan, "Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*)," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 92, no. 1, pp. 280–284, 1995.

[14] J. F. Wendel, "New world tetraploid cottons contain old world cytoplasm," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 86, pp. 4132–4136, 1989.

[15] R. L. Small, J. A. Ryburn, R. C. Cronn, T. Seelanan, and J. F. Wendel, "The tortoise and the hare: choosing between noncoding plastome and nuclear A*dh* sequences for phylogeny reconstruction in a recently diverged plant group," *American Journal of Botany*, vol. 85, no. 9, pp. 1301–1315, 1998.

[16] A. Desai, P. W. Chee, J. Rong, O. L. May, and A. H. Paterson, "Chromosome structural changes in diploid and tetraploid a genomes of *Gossypium*," *Genome*, vol. 49, no. 4, pp. 336–345, 2006.

[17] R. L. Small, J. A. Ryburn, and J. F. Wendel, "Low levels of nucleotide diversity at homoeologous A*dh* loci in allotetraploid cotton (*Gossypium* L.)," *Molecular Biology and Evolution*, vol. 16, no. 4, pp. 491–501, 1999.

[18] T. Seelanan, A. Schnabel, and J. F. Wendel, "Congruence and consensus in the cotton tribe (Malvaceae)," *Systematic Botany*, vol. 22, no. 2, pp. 259–290, 1997.

[19] J. F. Wendel and V. A. Albert, "Phylogenetics of the cotton genus (*Gossypium* L.): character-state weighted parsimony analysis of chloroplast DNA restriction site data and its systematic and biogeographic implications," *Systematic Botany*, vol. 17, pp. 115–143, 1992.

[20] J. O. Beasley, "Meiotic chromosome behavior in species hybrids, haploids, and induced polyploids of *Gossypium*," *Genetics*, vol. 27, pp. 25–54, 1942.

[21] D. M. Stelly, "Interfacing cytogenetics with the cotton genome mapping effort," in *Proceedings of the Beltwide Cotton Improvement Conference*, pp. 1545–1550, National Cotton Council, New Orleans, La, USA, January 1993.

[22] H. J. Kim and B. A. Triplett, "Cotton fiber growth in planta and in vitro. Models for plant cell elongation and cell wall biogenesis," *Plant Physiology*, vol. 127, no. 4, pp. 1361–1366, 2001.

[23] A. J. Reinisch, J.-M. Dong, C. L. Brubaker, D. M. Stelly, J. F. Wendel, and A. H. Paterson, "A detailed RFLP map of cotton, *Gossypium hirsutum x Gossypium barbadense*: chromosome organization and evolution in a disomic polyploid genome," *Genetics*, vol. 138, no. 3, pp. 829–847, 1994.

[24] J. K. Rong, C. Abbey, J. E. Bowers, et al., "A 3347-locus genetic recombination map of sequence-tagged sites reveals types of genome organization, transmission and evolution of cotton (*Gossypium*)," *Genetics*, vol. 166, pp. 389–417, 2004.

[25] R. J. Kohel, J. Yu, Y.-H. Park, and G. R. Lazo, "Molecular mapping and characterization of traits controlling fiber quality in cotton," *Euphytica*, vol. 121, no. 2, pp. 163–172, 2001.

[26] C. L. Brubaker and A. H. D. Brown, "The use of multiple alien chromosome addition aneuploids facilitates genetic linkage mapping of the *Gossypium G* genome," *Genome*, vol. 46, pp. 774–791, 2003.

[27] M. Mei, N. H. Syed, W. Gao, et al., "Genetic mapping and QTL analysis of fiber-related traits in cotton (*Gossypium*)," *Theoretical and Applied Genetics*, vol. 108, no. 2, pp. 280–291, 2004.

[28] J.-M. Lacape, T.-B. Nguyen, S. Thibivilliers, et al., "A combined RFLP-SSR-AFLP map of tetraploid cotton based on a *Gossypium hirsutum x Gossypium barbadense* backcross population," *Genome*, vol. 46, no. 4, pp. 612–626, 2003.

[29] T.-B. Nguyen, M. Giband, P. Brottier, A.-M. Risterucci, and J.-M. Lacape, "Wide coverage of the tetraploid cotton genome using newly developed microsatellite markers," *Theoretical and Applied Genetics*, vol. 109, no. 1, pp. 167–175, 2004.

[30] M. Ulloa, W. R. Meredith Jr., Z. W. Shappley, and A. L. Kahler, "RFLP genetic linkage maps from four F2.3 populations and a joinmap of *Gossypium hirsutum* L.," *Theoretical and Applied Genetics*, vol. 104, no. 2-3, pp. 200–208, 2002.

[31] V. N. Waghmare, J. Rong, C. J. Rogers, G. J. Pierce, J. F. Wendel, and A. H. Paterson, "Genetic mapping of a cross between *Gossypium hirsutum* (cotton) and the Hawaiian endemic, *Gossypium tomentosum*," *Theoretical and Applied Genetics*, vol. 111, no. 4, pp. 665–676, 2005.

[32] Z. Lin, X. Zhang, Y. Nie, D. He, and M. Wu, "Construction of a genetic linkage map for cotton based on SRAP," *Chinese Science Bulletin*, vol. 48, no. 19, pp. 2063–2067, 2003.

[33] Z. Lin, D. He, X. Zhang, et al., "Linkage map construction and mapping QTL for cotton fibre quality using SRAP, SSR and RAPD," *Plant Breeding*, vol. 124, no. 2, pp. 180–187, 2005.

[34] D.-H. He, Z.-X. Lin, X.-L. Zhang, et al., "QTL mapping for economic traits based on a dense genetic map of cotton with PCR-based markers using the interspecific cross of *Gossypium hirsutum* x *Gossypium barbadense*," *Euphytica*, vol. 153, no. 1-2, pp. 181–197, 2007.

[35] Y.-H. Park, M. S. Alabady, M. Ulloa, et al., "Genetic mapping of new cotton fiber loci using EST-derived microsatellites in an interspecific recombinant inbred line cotton population," *Molecular Genetics and Genomics*, vol. 274, no. 4, pp. 428–441, 2005.

[36] J. E. Frelichowski Jr, M. B. Palmer, D. Main, et al., "Cotton genome mapping with new microsatellites from Acala 'Maxxa' BAC-ends," *Molecular Genetics and Genomics*, vol. 275, no. 5, pp. 479–491, 2006.

[37] X. Song, K. Wang, W. Guo, J. Zhang, and T. Zhang, "A comparison of genetic maps constructed from haploid and BC1 mapping populations from the same crossing between *Gossypium hirsutum* L. and *Gossypium barbadense* L.," *Genome*, vol. 48, no. 3, pp. 378–390, 2005.

[38] Z.-G. Han, W.-Z. Guo, X.-L. Song, and T.-Z. Zhang, "Genetic mapping of EST-derived microsatellites from the diploid *Gossypium arboreum* in allotetraploid cotton," *Molecular Genetics and Genomics*, vol. 272, no. 3, pp. 308–327, 2004.

[39] Z. Han, C. Wang, X. Song, et al., "Characteristics, development and mapping of *Gossypium hirsutum* derived EST-SSRs in allotetraploid cotton," *Theoretical and Applied Genetics*, vol. 112, no. 3, pp. 430–439, 2006.

[40] W. Guo, P. Cai, C. Wang, et al., "A microsatellite-based, gene-rich linkage map reveals genome structure, function and evolution in *Gossypium*," *Genetics*, vol. 176, no. 1, pp. 527–541, 2007.

[41] J. Zhang, W. Guo, and T. Zhang, "Molecular linkage map of allotetraploid cotton (*Gossypium hirsutum* L. x *Gossypium barbadense* L.) with a haploid population," *Theoretical and Applied Genetics*, vol. 105, no. 8, pp. 1166–1174, 2002.

[42] R. J. Wright, P. M. Thaxton, K. M. El-Zik, and A. H. Paterson, "D-subgenome bias of *Xcm* resistance genes in tetraploid *Gossypium* (cotton) suggests that polyploid formation has created novel avenues for evolution," *Genetics*, vol. 149, no. 4, pp. 1987–1996, 1998.

[43] D. Rungis, D. Llewellyn, E. S. Dennis, and B. R. Lyon, "Investigation of the chromosomal location of the bacterial blight resistance gene present in an Australian cotton (*Gossypium hirsutum* L.) cultivar," *Australian Journal of Agricultural Research*, vol. 53, no. 5, pp. 551–560, 2002.

[44] R. J. Wright, P. M. Thaxton, K. M. El-Zik, and A. H. Paterson, "Molecular mapping of genes affecting pubescence of cotton," *Journal of Heredity*, vol. 90, no. 1, pp. 215–219, 1999.

[45] C.-X. Jiang, R. J. Wright, K. M. El-Zik, and A. H. Paterson, "Polyploid formation created unique avenues for response to selection in *Gossypium* (cotton)," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, no. 8, pp. 4419–4424, 1998.

[46] Z. W. Shappley, J. N. Jenkins, J. Zhu, and J. C. McCarty Jr., "Quantitative trait loci associated with agronomic and fiber traits of upland cotton," *Journal of Cotton Science*, vol. 2, no. 4, pp. 153–163, 1998.

[47] C. Jiang, R. J. Wright, S. S. Woo, T. A. DelMonte, and A. H. Paterson, "QTL analysis of leaf morphology in tetraploid *Gossypium* (cotton)," *Theoretical and Applied Genetics*, vol. 100, no. 3-4, pp. 409–418, 2000.

[48] Y. Saranga, M. Menz, C.-X. Jiang, R. J. Wright, D. Yakir, and A. H. Paterson, "Genomic dissection of genotype x environment interactions conferring adaptation of cotton to arid conditions," *Genome Research*, vol. 11, no. 12, pp. 1988–1995, 2001.

[49] Y. Saranga, C.-X. Jiang, R. J. Wright, D. Yakir, and A. H. Paterson, "Genetic dissection of cotton physiological responses to arid conditions and their inter-relationships with productivity," *Plant, Cell and Environment*, vol. 27, no. 3, pp. 263–277, 2004.

[50] L.-H. Ren, W.-Z. Guo, and T.-Z. Zhang, "Identification of quantitative trait loci (QTLs) affecting yield and fiber properties in chromosome 16 in cotton using substitution line," *Acta Botanica Sinica*, vol. 44, no. 7, pp. 815–820, 2002.

[51] L. Liu, W. Guo, X. Zhu, and T. Zhang, "Inheritance and fine mapping of fertility restoration for cytoplasmic male sterility in *Gossypium hirsutum* L.," *Theoretical and Applied Genetics*, vol. 106, no. 3, pp. 461–469, 2003.

[52] A. H. Paterson, Y. Saranga, M. Menz, C.-X. Jiang, and R. J. Wright, "QTL analysis of genotype x environment interactions affecting cotton fiber quality," *Theoretical and Applied Genetics*, vol. 106, no. 3, pp. 384–396, 2003.

[53] T. Zhang, Y. Yuan, J. Yu, W. Guo, and R. J. Kohel, "Molecular tagging of a major QTL for fiber strength in Upland cotton and its marker-assisted selection," *Theoretical and Applied Genetics*, vol. 106, no. 2, pp. 262–268, 2003.

[54] J. Yin, W. Guo, L. Yang, L. Liu, and T. Zhang, "Physical mapping of the Rf1 fertility-restoring gene to a 100 kb region in cotton," *Theoretical and Applied Genetics*, vol. 112, no. 7, pp. 1318–1325, 2006.

[55] M. Ulloa and W. R. Meredith Jr, "Genetic linkage map and QTL analysis of agronomic and fiber traits in a intraspecific population," *Journal of Cotton Science*, vol. 4, no. 3, pp. 161–170, 2000.

[56] M. Ulloa, S. Saha, J. N. Jenkins, W. R. Meredith Jr., J. C. McCarty Jr., and D. M. Stelly, "Chromosomal assignment of RFLP linkage groups harboring important QTLs on an intraspecific cotton (*Gossypium hirsutum* L.) joinmap," *Journal of Heredity*, vol. 96, no. 2, pp. 132–144, 2005.

[57] Y. Bolek, K. M. El-Zik, and A. E. Pepper, "Mapping of verticillium wilt resistance genes in cotton," *Plant Science*, vol. 168, pp. 1581–1590, 2005.

[58] P. Chee, X. Draye, C. X. Jiang, et al., "Molecular dissection of interspecific variation between *Gossypium hirsutum* and *Gossypium barbadense* (cotton) by a backcross-self approach: I. Fiber elongation," *Theoretical and Applied Genetics*, vol. 111, pp. 757–763, 2005.

[59] P. Chee, X. Draye, C. X. Jiang, et al., "Molecular dissection of interspecific variation between *Gossypium hirsutum* and *Gossypium barbadense* (cotton) by a backcross-self approach: III. Fiber length," *Theoretical and Applied Genetics*, vol. 111, pp. 772–781, 2005.

[60] X. Draye, P. Chee, C.-X. Jiang, et al., "Molecular dissection of interspecific variation between *Gossypium hirsutum* and *G. barbadense* (cotton) by a backcross-self approach: II. Fiber fineness," *Theoretical and Applied Genetics*, vol. 111, no. 4, pp. 764–771, 2005.

[61] J. Rong, G. J. Pierce, V. N. Waghmare, et al., "Genetic mapping and comparative analysis of seven mutants related to seed fiber development in cotton," *Theoretical and Applied Genetics*, vol. 111, no. 6, pp. 1137–1146, 2005.

[62] X.-L. Song, W.-Z. Guo, Z.-G. Han, and T.-Z. Zhang, "Quantitative trait loci mapping of leaf morphological traits and chlorophyll content in cultivated tetraploid cotton," *Journal of Integrative Plant Biology*, vol. 47, no. 11, pp. 1382–1390, 2005.

[63] J.-M. Lacape and T. B. Nguyen, "Mapping quantitative trait loci associated with leaf and stem pubescence in cotton," *Journal of Heredity*, vol. 96, no. 4, pp. 441–444, 2005.

[64] J.-M. Lacape, T.-B. Nguyen, B. Courtois, et al., "QTL analysis of cotton fiber quality using multiple *Gossypium hirsutum x Gossypium barbadense* backcross generations," *Crop Science*, vol. 45, no. 1, pp. 123–140, 2005.

[65] Z.-S. Zhang, Y.-H. Xiao, M. Luo, et al., "Construction of a genetic linkage map and QTL analysis of fiber-related traits in upland cotton (*Gossypium hirsutum* L.)," *Euphytica*, vol. 144, no. 1-2, pp. 91–99, 2005.

[66] D.-H. He, Z.-X. Lin, X.-L. Zhang, et al., "Mapping QTLs of traits contributing to yield and analysis of genetic effects in tetraploid cotton," *Euphytica*, vol. 144, no. 1-2, pp. 141–149, 2005.

[67] X. Shen, W. Guo, X. Zhu, et al., "Molecular mapping of QTLs for fiber qualities in three diverse lines in Upland cotton using SSR markers," *Molecular Breeding*, vol. 15, no. 2, pp. 169–181, 2005.

[68] X. Shen, G. V. Becelaere, P. Kumar, R. F. Davis, O. L. May, and P. Chee, "QTL mapping for resistance to root-knot nematodes in the M-120 RNR Upland cotton line (*Gossypium hirsutum* L.) of the Auburn 623 RNR source," *Theoretical and Applied Genetics*, vol. 113, no. 8, pp. 1539–1549, 2006.

[69] X. Shen, T. Zhang, W. Guo, X. Zhu, and X. Zhang, "Mapping fiber and yield QTLs with main, epistatic, and QTL x environment interaction effects in recombinant inbred lines of Upland cotton," *Crop Science*, vol. 46, no. 1, pp. 61–66, 2006.

[70] X. Shen, W. Guo, Q. Lu, X. Zhu, Y. Yuan, and T. Zhang, "Genetic mapping of quantitative trait loci for fiber quality and yield trait by RIL approach in Upland cotton," *Euphytica*, vol. 155, no. 3, pp. 371–380, 2007.

[71] C. Wang, M. Ulloa, and P. A. Roberts, "Identification and mapping of microsatellite markers linked to a root-knot nematode resistance gene (*rkn1*) in Acala NemX cotton (*Gossypium hirsutum* L.)," *Theoretical and Applied Genetics*, vol. 112, no. 4, pp. 770–777, 2006.

[72] C. Wang and P. A. Roberts, "Development of AFLP and derived CAPS markers for root-knot nematode resistance in cotton," *Euphytica*, vol. 152, no. 2, pp. 185–196, 2006.

[73] P. Ynturi, J. N. Jenkins, J. C. McCarty Jr., O. A. Gutierrez, and S. Saha, "Association of root-knot nematode resistance genes with simple sequence repeat markers on two chromosomes in cotton," *Crop Science*, vol. 46, no. 6, pp. 2670–2674, 2006.

[74] W.-Z. Guo, G.-J. Ma, Y.-C. Zhu, C.-X. Yi, and T.-Z. Zhang, "Molecular tagging and mapping of quantitative trait loci for lint percentage and morphological marker genes in upland cotton," *Journal of Integrative Plant Biology*, vol. 48, no. 3, pp. 320–326, 2006.

[75] S. Saha, J. N. Jenkins, J. Wu, et al., "Effects of chromosome-specific introgression in upland cotton on fiber and agronomic traits," *Genetics*, vol. 172, no. 3, pp. 1927–1938, 2006.

[76] B.-H. Wang, Y.-T. Wu, N.-T. Huang, X.-F. Zhu, W.-Z. Guo, and T.-Z. Zhang, "QTL mapping for plant architecture traits in upland cotton using RILs and SSR markers," *Acta Genetica Sinica*, vol. 33, no. 2, pp. 161–170, 2006.

[77] B. Wang, W. Guo, X. Zhu, Y. Wu, N. Huang, and T. Zhang, "QTL mapping of fiber quality in an elite hybrid derived-RIL population of upland cotton," *Euphytica*, vol. 152, no. 3, pp. 367–378, 2006.

[78] B. Wang, W. Guo, X. Zhu, Y. Wu, N. Huang, and T. Zhang, "QTL mapping of yield and yield components for elite hybrid derived-RILs in Upland cotton," *Journal of Genetics and Genomics*, vol. 34, no. 1, pp. 35–45, 2007.

[79] J. P. Tomkins, D. G. Peterson, T. J. Yang, et al., "Development of genomic resources for cotton (*Gossypium hirsutum* L.): BAC library construction, preliminary STC analysis, and identification of clones associated with fiber development," *Molecular Breeding*, vol. 8, no. 3, pp. 255–261, 2001.

[80] J.-M. Yin, W.-Z. Guo, and T.-Z. Zhang, "Construction and identification of bacterial artificial chromosome library for 0-613-2R in upland cotton," *Journal of Integrative Plant Biology*, vol. 48, no. 2, pp. 219–222, 2006.

[81] U-J Kim, B. W. Birren, T. Slepak, et al., "Construction and Characterization of a Human Bacterial Artificial Chromosome Library," *Genomics*, vol. 34, no. 2, pp. 213–218, 1996.

[82] A. C. G. Frijters, Z. Zhang, M. van Damme, et al., "Construction of a bacterial artificial chromosome library containing large Eco RI and Hin dIII genomic fragments of lettuce," *Theoretical and Applied Genetics*, vol. 94, no. 3-4, 1997.

[83] Q. Tao and H.-B. Zhang, "Cloning and stable maintenance of DNA fragments over 300 kb in Escherichia coli with conventional plasmid-based vectors," *Nucleic Acids Research*, vol. 26, no. 21, pp. 4901–4909, 1998.

[84] M. A. Saghai Maroof, R. M. Biyashev, G. P. Yang, Q. Zhang, and R. W. Allard, "Extraordinarily polymorphic microsatellite DNA in barley: species diversity, chromosomal locations, and population dynamics," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 91, no. 12, pp. 5466–5470, 1994.

[85] O. U. K. Reddy, A. E. Pepper, I. Abdurakhmonov, et al., "New dinucleotide and trinucleotide microsatellite marker resources for cotton genome research," *Journal of Cotton Science*, vol. 5, no. 2, pp. 103–113, 2001.

[86] S. Liu, S. Saha, D. Stelly, B. Burr, and R. G. Cantrell, "Chromosomal assignment of microsatellite loci in cotton," *Journal of Heredity*, vol. 91, no. 4, pp. 326–332, 2000.

[87] J.-M. Lacape, D. Dessauw, M. Rajab, J.-L. Noyer, and B. Hau, "Microsatellite diversity in tetraploid *Gossypium germplasm*: assembling a highly informative genotyping set of cotton SSRs," *Molecular Breeding*, vol. 19, no. 1, pp. 45–58, 2007.

[88] D. Q. Liu, X. P. Guo, Z. X. Lin, Y. C. Nie, and X. L. Zhang, "Genetic diversity of Asian cotton (*Gossypium arboreum* L.) in China evaluated by microsatellite analysis," *Genetic Resources and Crop Evolution*, vol. 53, pp. 1145–1152, 2006.

[89] D. Rungis, D. Llewellyn, E. S. Dennis, and B. R. Lyon, "Simple sequence repeat (SSR) markers reveal low levels of polymorphism between cotton (*Gossypium hirsutum* L.) cultivars," *Australian Journal of Agricultural Research*, vol. 56, no. 3, pp. 301–307, 2005.

[90] J. Zhang, Y. Lu, R. G. Cantrell, and E. Hughs, "Molecular marker diversity and field performance in commercial cotton cultivars evaluated in the southwestern USA," *Crop Science*, vol. 45, no. 4, pp. 1483–1490, 2005.

[91] J. P. Connell, S. Pammi, M. J. Iqbal, T. Huizinga, and A. S. Reddy, "A high through-put procedure for capturing microsatellites from complex plant genomes," *Plant Molecular Biology Reporter*, vol. 16, no. 4, pp. 341–349, 1998.

[92] S. N. Qureshi, S. Saha, R. V. Kantety, and J. N. Jenkins, "Molecular biology and physiology: EST-SSR: a new class of genetic markers in cotton," *Journal of Cotton Science*, vol. 8, no. 2, pp. 112–123, 2004.

[93] J. Lichtenzveig, C. Scheuring, J. Dodge, S. Abbo, and H.-B. Zhang, "Construction of BAC and BIBAC libraries and their applications for generation of SSR markers for genome analysis of chickpea, *Cicer arietinum* L," *Theoretical and Applied Genetics*, vol. 110, no. 3, pp. 492–510, 2005.

[94] A. Blenda, J. Scheffler, B. Scheffler, et al., "CMD: a cotton microsatellite database resource for *Gossypium* genomics," *BMC Genomics*, vol. 7, p. 132, 2006.

[95] P. W. Chee, J. Rong, D. Williams-Coplin, S. R. Schulze, and A. H. Paterson, "EST derived PCR-based markers for functional gene homologues in cotton," *Genome*, vol. 47, no. 3, pp. 449–462, 2004.

[96] S. Saha, M. Karaca, J. N. Jenkins, A. E. Zipf, O. U.K. Reddy, and R. V. Kantety, "Simple sequence repeats as useful resources to study transcribed genes of cotton," *Euphytica*, vol. 130, no. 3, pp. 355–364, 2003.

[97] E. Taliercio, R. D. Allen, M. Essenberg, et al., "Analysis of ESTs from multiple *Gossypium hirsutum* tissues and identification of SSRs," *Genome*, vol. 49, no. 4, pp. 306–319, 2006.

[98] W. Guo, W. Wang, B. Zhou, and T. Zhang, "Cross-species transferability of *G. arboreum*-derived EST-SSRs in the diploid species of *Gossypium*," *Theoretical and Applied Genetics*, vol. 112, no. 8, pp. 1573–1581, 2006.

[99] K. Wang, X. Song, Z. Han, et al., "Complete assignment of the chromosomes of *Gossypium hirsutum* L. by translocation and fluorescence in situ hybridization mapping," *Theoretical and Applied Genetics*, vol. 113, no. 1, pp. 73–80, 2006.

[100] A. Arpat, M. Waugh, J. P. Sullivan, et al., "Functional genomics of cell elongation in developing cotton fibers," *Plant Molecular Biology*, vol. 54, pp. 911–929, 2004.

[101] C. H. Haigler, D. Zhang, and C. G. Wilkerson, "Biotechnological improvement of cotton fibre maturity," *Physiologia Plantarum*, vol. 124, no. 3, pp. 285–294, 2005.

[102] J. A. Udall, J. M. Swanson, K. Haller, et al., "A global assembly of cotton ESTs," *Genome Research*, vol. 16, no. 3, pp. 441–450, 2006.

[103] S. S. Yang, F. Cheung, J. J. Lee, et al., "Accumulation of genome-specific transcripts, transcription factors and phytohormonal regulators during early stages of fiber cell development in allotetraploid cotton," *The Plant Journal*, vol. 47, no. 5, pp. 761–775, 2006.

[104] Y.-H. Shi, S.-W. Zhu, X.-Z. Mao, et al., "Transcriptome profiling, molecular biological, and physiological studies reveal a major role for ethylene in cotton fiber cell elongation," *The Plant Cell*, vol. 18, no. 3, pp. 651–664, 2006.

[105] M. Ulloa, C. Brubaker, and P. Chee, "Cotton," in *Genome Mapping & Molecular Breeding*, C. Kole, Ed., vol. 7 of *Technical Crops*, Springer, New York, NY, USA, 2006.

[106] J. Rong, F. A. Feltus, V. N. Waghmare, et al., "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*, vol. 176, no. 4, pp. 2577–2588, 2007.

[107] L. He, C. Du, Y. Li, C. Scheuring, and H.-B. Zhang, "Large-insert bacterial clone libraries and their applications," in *Aquaculture Genome Technologies*, Z. Liu, Ed., pp. 215–244, Blackwell, Ames, Iowa, USA, 2007.

[108] C. Ren, Z. Y. Xu, S. Sun, et al., "Genomic DNA libraries and physical mapping," in *The Handbook of Plant Genome Mapping: Genetic and Physical Mapping*, K. Meksem and G. Kahl, Eds., pp. 173–213, Wiley-VCH Verlag GmbH, Weinheim, Germany, 2005.

[109] C. Wu, Z. Xu, and H.-B. Zhang, "DNA libraries," in *Encyclopedia of Molecular Cell Biology and Molecular Medicine*, R. A. Meyers, Ed., pp. 385–425, Wiley-VCH Verlag GmbH, Weinheim, Germany, 2nd edition, 2004.

[110] H.-B. Zhang and C. Wu, "BAC as tools for genome sequencing," *Plant Physiology and Biochemistry*, vol. 39, no. 3-4, pp. 195–209, 2001.

[111] H.-B. Zhang, S.-S. Woo, and R. A. Wing, "BAC, YAC and cosmid library construction," in *Plant Gene Isolation: Principles and Practice*, G. Foster and D. Twell, Eds., pp. 75–99, John Wiley & Sons, Chichester, UK, 1996.

[112] P. A. Ioannou, C. T. Amemiya, J. Garnes, et al., "A new bacteriophage P1-derived vector for the propagation of large human DNA fragments," *Nature Genetics*, vol. 6, no. 1, pp. 84–89, 1994.

[113] H. Shizuya, B. Birren, U.-J. Kim, et al., "Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 89, no. 18, pp. 8794–8797, 1992.

[114] Y.-L. Chang, Q. Tao, C. Scheuring, K. Meksem, and H.-B. Zhang, "An integrated map of *Arabidopsis thaliana* for functional analysis of its genome sequence," *Genetics*, vol. 159, pp. 1231–1242, 2001.

[115] R. A. Hoskins, C. R. Nelson, B. P. Berman, et al., "A BAC-based physical map of the major autosomes of *Drosophila melanogaster*," *Science*, vol. 287, no. 5461, pp. 2271–2274, 2000.

[116] International Human Genome Mapping Consortium, "A physical map of the human genome," *Nature*, vol. 409, pp. 934–941, 2001.

[117] Y. Li, T. Uhm, C. Ren, et al., "A plant-transformation-competent BIBAC/BAC-based map of rice for functional analysis and genetic engineering of its genomic sequence," *Genome*, vol. 50, no. 3, pp. 278–288, 2007.

[118] M. Marra, T. Kucaba, M. Sekhon, et al., "A map for sequence analysis of the *Arabidopsis thaliana* genome," *Nature Genetics*, vol. 22, pp. 265–270, 1999.

[119] T. Mozo, K. Dewar, P. Dunn, et al., "A complete BAC-based physical map of the *Arabidopsis thaliana* genome," *Nature Genetics*, vol. 22, no. 3, pp. 271–275, 1999.

[120] C. Ren, M.-K. Lee, B. Yan, et al., "A BAC-based physical map of the chicken genome," *Genome Research*, vol. 13, no. 12, pp. 2754–2758, 2003.

[121] Q. Tao, Y.-L. Chang, J. Wang, et al., "Bacterial artificial chromosome-based physical map of the rice genome constructed by restriction fingerprint analysis," *Genetics*, vol. 158, no. 4, pp. 1711–1724, 2001.

[122] J. W. Wallis, J. Aerts, M. A. M. Groenen, et al., "A physical map of the chicken genome," *Nature*, vol. 432, no. 7018, pp. 761–764, 2004.

[123] C. Wu, S. Sun, M.-K. Lee, Z. Y. Xu, C. Ren, and H.-B. Zhang, "Whole genome physical mapping: an overview on methods for DNA fingerprinting," in *The Handbook of Plant Genome Mapping: Genetic and Physical Mapping*, K. Meksem and G. Kahl, Eds., pp. 257–284, Wiley-VCH Verlag GmbH, Weinheim, Germany, 2005.

[124] C. Wu, S. Sun, P. Nimmakayala, et al., "A BAC- and BIBAC-based physical map of the soybean genome," *Genome Research*, vol. 14, no. 2, pp. 319–326, 2004.

[125] Z. Xu, M. V. D. Berg, C. Scheuring, et al., "Genome-wide physical mapping from large-insert clones by fingerprint analysis with capillary electrophoresis: a robust physical map of Penicillium chrysogenum," *Nucleic Acids Research*, vol. 33, p. e50, 2005.

[126] Z. Xu, S. Sun, L. Covaleda, et al., "Genome physical mapping with large-insert bacterial clones by fingerprint analysis: methodologies, source clone genome coverage, and contig map quality," *Genomics*, vol. 84, no. 6, pp. 941–951, 2004.

[127] H.-B. Zhang and R. A. Wing, "Physical mapping of the rice genome with BACs," *Plant Molecular Biology*, vol. 35, no. 1-2, pp. 115–127, 1997.

[128] X. Zhang, C. Scheuring, S. Tripathy, et al., "An integrated BAC and genome sequence physical map of *Phytophthora sojae*," *Molecular Plant-Microbe Interactions*, vol. 19, no. 12, pp. 1302–1310, 2006.

[129] M. D. Adams, S. E. Celniker, R. A. Holt, et al., "The genome sequence of *Drosophila melanogaster*," *Science*, vol. 287, pp. 2185–2195, 2000.

[130] The Arabidopsis Genome Initiative, "Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*," *Nature*, vol. 408, pp. 796–815, 2000.

[131] International Human Genome Sequencing Consortium, "Initial sequencing and analysis of the human genome," *Nature*, vol. 409, pp. 860–921, 2001.

[132] T. Sasaki, "The map-based sequence of the rice genome," *Nature*, vol. 436, no. 7052, pp. 793–800, 2005.

[133] B. M. Tyler, S. Tripathy, X. Zhang, et al., "*Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis," *Science*, vol. 313, no. 5791, pp. 1261–1266, 2006.

[134] H.-B. Zhang, "Map-based cloning of genes and QTLs," in *Plant Molecular Mapping and Breeding*, C. Kole and A. Abbott, Eds., Springer, New York, NY, USA, 2007.

[135] M. Chen, P. Sanmiguel, A. C. de Oliveira, et al., "Microcolinearity in *sh*2-homologous regions of the maize, rice, and sorghum genomes," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 94, no. 7, pp. 3431–3435, 1997.

[136] A. Patocchi, B. A. Vinatzer, L. Gianfranceschi, et al., "Construction of a 550 kb BAC contig spanning the genomic region containing the apple scab resistance gene Vf," *Molecular and General Genetics*, vol. 262, no. 4-5, pp. 884–891, 1999.

[137] M. S. Zwick, M. N. Islam-Faridi, D. G. Czeschin Jr., et al., "Physical mapping of the liguleless linkage group in sorghum bicolor using rice RFLP-selected sorghum BACs," *Genetics*, vol. 148, no. 4, pp. 1983–1992, 1998.

[138] A. S. Basra and C. P. Malik, "Development of the cotton fiber," *International Review of Cytology*, vol. 89, pp. 65–113, 1984.

[139] T. A. Wilkins and A. B. Arpat, "The cotton fiber transcriptome," *Physiologia Plantarum*, vol. 124, no. 3, pp. 295–300, 2005.

[140] Y. Wu, A. C. Machado, R. G. White, D. J. Llewellyn, and E. S. Dennis, "Expression profiling identifies gene expressed early during lint fiber initiation in cotton," *Plant Cell Physiology*, vol. 18, pp. 651–664, 2006.

[141] W. M. Nelson, A. K. Bharti, E. Butler, et al., "Whole-genome validation of high-information-content fingerprinting," *Plant Physiology*, vol. 139, no. 1, pp. 27–38, 2005.

[142] C. Wu, S. Wang, and H.-B. Zhang, "Interactions among genomic structure, function, and evolution revealed by comprehensive analysis of the *Arabidopsis* genome," *Genomics*, vol. 88, no. 4, pp. 394–406, 2006.

[143] J. S. Hawkins, H. Kim, J. D. Nason, R. A. Wing, and J. F. Wendel, "Differential lineage-specific amplification of transposable elements is responsible for genome size variation in *Gossypium*," *Genome Research*, vol. 16, no. 10, pp. 1252–1261, 2006.

[144] K. Dreher, M. Morris, M. Khairallah, J. M. Ribaut, S. Pandey, and G. Srinivasan, "Is marker-assisted selection cost-effective compared to conventional plant breeding methods? The case of quality protein maize," in *Proceedings of the 4th Annual Conference of the International Consortium on Agricultural Biotechnology Research (ICABR '00)*, pp. 203–236, Ravello, Italy, August 2000.

[145] W. Guo, T. Zhang, X. Shen, J. Z. Yu, and R. J. Kohel, "Development of SCAR marker linked to a major QTL for high fiber strength and its usage in molecular-marker assisted selection in upland cotton," *Crop Science*, vol. 43, no. 6, pp. 2252–2256, 2003.

[146] J. F. Zhang and J. McD. Stewart, "Inheritance and genetic relationships of the D8 and D2-2 restorer genes for cotton cytoplasmic male sterility," *Crop Science*, vol. 41, no. 2, pp. 289–294, 2001.

[147] W. Guo, T. Zhang, J. Pan, and R. J. Kohel, "Identification of RAPD marker linked with fertility-restoring gene of cytoplasmic male sterile lines in upland cotton," *Chinese Science Bulletin*, vol. 43, no. 1, pp. 52–54, 1998.

[148] J. Zhang and J. McD. Stewart, "Identification of molecular markers linked to the fertility restorer genes for CMS-D8 in cotton," *Crop Science*, vol. 44, no. 4, pp. 1209–1217, 2004.

[149] L. He, C. G. Du, L. Covaleda, et al., "Cloning, characterization, and evolution of the NBS-encoding resistance gene analogue family in polyploid cotton (*Gossypium hirsutm* L.)," *Molecular Plant-Microbe Interaction*, vol. 17, pp. 1234–1241, 2004.

[150] J. Gardiner, S. Schroeder, M. L. Polacco, et al., "Anchoring 9,371 maize expressed sequence tagged unigenes to the bacterial artificial chromosome contig map by two-dimensional overgo hybridization," *Plant Physiology*, vol. 134, no. 4, pp. 1317–1326, 2004.

Hindawi Publishing Corporation
International Journal of Plant Genomics
Volume 2008, Article ID 820274, 12 pages
doi:10.1155/2008/820274

Review Article

Structural and Functional Genomics of Tomato

Amalia Barone,[1] Maria Luisa Chiusano,[1] Maria Raffaella Ercolano,[1] Giovanni Giuliano,[2] Silvana Grandillo,[3] and Luigi Frusciante[1]

[1] *Department of Soil, Plant, Environmental and Animal Production Sciences, University of Naples "Federico II", Via Università 100, 80055 Portici, Italy*
[2] *Ente per le Nuove Tecnologie, l'Energia e l'Ambiente, Casaccia Research Center, Via Anguillarese 301, S.M. di Galeria, 00123 Roma, Italy*
[3] *CNR-Institute of Plant Genetics, Via Università 133, 80055 Portici, Italy*

Correspondence should be addressed to Luigi Frusciante, fruscian@unina.it

Received 27 July 2007; Accepted 22 November 2007

Recommended by P. K. Gupta

Tomato (*Solanum lycopersicum* L.) is the most intensively investigated Solanaceous species both in genetic and genomics studies. It is a diploid species with a haploid set of 12 chromosomes and a small genome (950 Mb). Based on the detailed knowledge on tomato structural genomics, the sequencing of the euchromatic regions started in the year 2005 as a common effort of different countries. The manuscript focuses on markers used for tomato, on mapping efforts mainly based on exploitation of natural biodiversity, and it gives an updated report on the international sequencing activities. The principal tools developed to explore the function of tomato genes are also summarized, including mutagenesis, genetic transformation, and transcriptome analysis. The current progress in bioinformatic strategies available to manage the overwhelming amount of data generated from different tomato "omics" approaches is reported, and emphasis is given to the effort of producing a computational workbench for the analysis of the organization, as well as the functionality and evolution of the Solanaceae family.

1. INTRODUCTION

Tomato (*Solanum lycopersicum* L., formerly *Lycopersicon esculentum* Miller) is an economically important crop worldwide, and a preeminent model system for genetic studies in plants. It is also the most intensively investigated Solanaceous species, with simple diploid genetics, a short generation time, routine transformation technology, and availability of rich genetic and genomic resources. It has a diploid genome with 12 chromosome pairs and a genome size of 950 Mb [1] encoding approximately 35,000 genes that are largely sequestered in contiguous euchromatic regions [2]. Several resources are available for genetic/genomic research in tomato including the following: (i) tomato wild species and mutant collections; (ii) marker collections; (iii) F_2 synteny mapping population and permanent recombinant inbred (RI) mapping populations; (iv) BAC libraries and an advanced physical map; (v) TILLING populations; and (vi) tomato microarrays, gene silenced tomato lines, and VIGS libraries (for transient silencing).

Till recently, tomato genomics has largely relied on molecular marker analysis and functional analysis of genesets. However, for a better understanding of structural and functional aspects of its genome, following latest high-throughput technologies are also being utilized: (i) RNA transcription and protein analysis, (ii) screening of post-translational modifications and protein-protein interactions, and (iii) discovery of metabolic networks. The information generated by large-scale genome sequencing can lead a major revolution in the understanding of tomato biology.

The International Solanaceae Genome Project (SOL) was established to develop a network of knowledge on the Solanaceae family and to coordinate the research efforts of different groups from around the world [3]. The Solanaceae Genomics Network website (SGN; http://www.sgn.cornell.edu) was created to facilitate distribution of genomic information for tomato in particular and for Solanaceous species in general in a comparative genomic context [4]. The challenge facing SOL in the coming years is to develop methodologies that will enable genomic information to be associated with phenotypes of interest for crop improvement. The framework for organizing these data is the highly conserved genetic map of the Solanaceae that will allow the information basis to be extended beyond individual species.

Progress in tomato research will depend on our ability to tie together the independent components into higher-order complexity with multiple dimensions. Multidisciplinary research efforts, involving the increased input of chemistry, physics, statistics, mathematics, and computing sciences, are becoming increasingly crucial for the success of such approach.

2. STRUCTURAL GENOMICS

2.1. Molecular markers

Beginning in the 1980s, different types of molecular markers have been developed in tomato. Among crop species, tomato is one of the richest in the number and type of these genetic markers, including restriction fragment length polymorphisms (RFLPs), simple sequence repeats (SSRs), cleaved amplified polymorphic sequence (CAPS), amplified fragment length polymorphisms (AFLPs), and single nucleotide polymorphism (SNP). Chronologically, RFLPs were the first markers developed. Currently, more than 1000 RFLPs have been mapped on the 12 tomato chromosomes. A subset of RFLP markers has been converted into PCR-based markers through sequencing of their ends. These sequences are available from the SGN Database, thus allowing specific primers for PCR reaction to be designed. Other PCR-based markers were developed both as random markers, such as random amplified polymorphic DNA (RAPD), AFLPs, and locus-specific markers, such as SSRs, CAPS, and conserved ortholog sets (COSs); and many of them have been mapped onto the high-density tomato genetic map [5].

Given the huge number of markers that have been set up for tomato using different methods, a database collecting the different datasets is available at the SGN website. Indeed, all information for more than 15,000 markers is collected in the SGN [6], where a specific tool for “marker search” allows markers to be located on the map. Markers can be searched by name, chromosome position, mapping population, and BAC associations (if they have been associated with BAC from the tomato sequencing project by hybridization with overgo probes or computationally by BLAST comparisons). Some of them have also been grouped into collections for organizational purposes or because they are part of a particular project. So, it is possible to select either COS (markers that have been mapped in both tomato and Arabidopsis) or COSII markers (markers that have been mapped in several Asterid species, including several Solanaceous species) [7, 8]. Other groups comprise known function genes (KFG), or EST-derived (TM) markers.

Recently, large-scale sequencing work in tomato has been generating sequences of whole BAC and cloned genes, ESTs collected from different cDNA libraries, and the sequences of full-length cDNAs. The cataloguing of these sequences in public databases is providing useful information to develop markers with high resolving power, such as SNPs and InDels, thus initiating an era of in silico tomato marker discovery. The tomato SSRs are an example of genetic markers that can be mined from existing sequence data. Smulders et al. [9], by screening the EMBL and GenBank Databases, identified 36 primer pairs, which detected polymorphisms at or close to coding regions. In recent studies, as many as 2627 SSRs were mined from an EST dataset by screening the 26,363 tomato EST unigene dataset and 57,222 full-length cDNA sequences that were available in MiBASE (http://www.accelrys.com/products/gcg). Most of these SSRs (around 80%) were novel SSRs, since they did not match any of the the mapped markers, thus being candidates for novel microsatellite markers [10]. In addition, more than 250,000 ESTs derived from cDNA libraries are currently catalogued on the SGN website. All these sequences are potentially source of new markers, such as SSRs and SNPs, useful for tomato genome analysis. In fact, besides SSRs, SNPs can also be mined from sequence data [11]; and an efficient in silico SNPs discovery is feasible for tomato due to the availability of EST in public databases [12]. Moreover, in the framework of the tomato sequencing project, around 400,000 BAC ends are being sequenced that could also be mined to search SNPs among *Solanum* genotypes (Ercolano et al., unpublished results). These will allow useful PCR marker to be derived that also fall in intron regions, thus complementing the detection of polymorphism in the coding regions represented by the ESTs.

Recently, oligonucleotide-based arrays have been used to identify DNA sequence polymorphisms in different species, since they allow high-throughput development of markers. Total genomic DNA hybridization methods are also being exploited in tomato with the aim of identifying markers such as single feature polymorphisms (SFPs). For instance, a 15.27 K gene NimbleGen tomato array was used by Sim et al. [13] for a study of polymorphism between *S. lycopersicum* and its closely related wild species.

2.2. Genetic and physical maps

Genetic mapping of morphological traits in tomato started at the beginning of last century, and by 1973 a total of 257 morphological and disease resistance markers had been mapped [14]. By the 1990s, tomato had become one of the first plants for which RFLPs were used to generate a high-density linkage map [15]. Later several genetic maps using PCR-based markers were developed and integrated with the RFLP maps, as reviewed by Labate et al. [16]. The first PCR-based reference genetic map covering the entire tomato genome was reported by Frary et al. [5] for a population derived from the cross *S. lycopersicum* × *S. pennellii* .

The Solanaceae is the first family of flowering plants for which comparative mapping was conducted [17, 18]. As a result, several genetic maps not only for tomato genome, but also for the genomes of other Solanaceous crops are now available at the SGN site [4]. Comparative genome analysis showed that tomato and potato genomes differ in only five paracentric inversions [15], whereas the tomato and pepper genomes differ in numerous rearrangements including several translocations as well as both pericentric and paracentric inversions [19, 20]. More recently, Doganlar et al. [21] have shown that eggplant and tomato genomes are differentiated by 28 rearrangements, which could be explained by 23 paracentric inversions and five translocations. These data suggest that paracentric inversions have been the primary mechanism for chromosome evolution in the Solanaceae.

Comparative genomics research is presently gaining momentum in Solanaceae due to availability of sequencing data for several species. This will greatly enhance the resolution of comparative mapping in this family. This research activity received further support due to the availability of whole genome sequence of *A. thaliana*, which facilitated the development of PCR-based COS markers using genes shared between distantly related plant taxa [7, 8]. For instance, in an effort to determine the level of synteny between *Arabidopsis* genome and the genomes of tomato and other Solanceous species, COSII markers are being mapped not only on tomato genome, but also on the genomes of other major Solanaceous species including eggplant, pepper, and *Nicotiana* (http://www.sgn.cornell.edu/markers/cosii_markers.pl). Also, in order to test the efficacy of COSII markers for comparative mapping across large phylogenetic distances, a subset of COSII markers is being mapped on the genomes of both tomato and diploid coffee (*Coffea canephora*) [8].

Besides genetic linkage maps, cytological and cytogenetic maps are also available for tomato. For example, Sherman and Stack [22] developed a physical map that was used to quantify the distribution of crossovers along each chromosome. Physical maps were also developed by in situ hybridisation techniques and allowed a comparison of linear order and distances between markers on genetic linkage maps and physical maps [23–25]. The results obtained by Peterson et al. [23] have shown that the linear order of markers is not always conserved between genetic and cytological positions.

The availability of mapped markers and of FISHed BAC allowed the construction of a high-density integrated genetic and physical map, whose definition is still in progress, and which is the foundation for the tomato genome sequencing project. Overgo analysis has been used to match BAC to probes based on markers from the *S. lycopersicon* × *S. pennellii* map. This analysis found 600 markers that unambiguously anchor over 5000 BACs to the genetic map [26]. Actually, at SGN site there are more than 10 000 BACs, which are anchored to the genetic maps.

2.3. QTL mapping and exploitation of natural biodiversity

The high-density RFLP linkage map developed for tomato facilitated extensive mapping of qualitative traits such as various disease resistance genes, for example, [27, 28]. This allowed tomato breeders to use marker-assisted selection (MAS) for variety improvement. Furthermore, tomato was the first species for which a whole genome molecular linkage map was used to identify quantitative trait loci (QTL) [29], leading to an understanding of the genetic basis of numerous quantitative traits including morphology, yield, fruit quality, fruit primary and secondary metabolites, as well as resistance to a variety of abiotic and biotic stresses [26]. The QTL mapping studies conducted by de Vicente and Tanksley [30] and by Eshed and Zamir [31] using mapping populations derived from interspecific tomato crosses provided stronger evidence that despite the inferior phenotype, unadapted germplasm could also be used as a source of complementary positive alleles that can result in favorable transgressive phenotypes once incorporated in the cultivated background.

The above results indicated that new molecular breeding strategies need to be devised in order to allow more efficient use of the genetic potential stored in seed banks and exotic germplasm. One such approach, the "advanced backcross QTL mapping method" was proposed by Tanksley and Nelson [32] with the purpose of combining the process of QTL analysis with variety development, by simultaneously identifying and transferring favorable QTL alleles from unadapted to cultivated germplasm. The AB-QTL strategy was initially developed and tested in tomato [33], and since then, it has been adapted for use in other crops including rice, maize, wheat, pepper, barley, and bean [34]. So far five AB-QTL studies have been conducted in tomato involving crosses with five wild *Solanum* species and in all cases favorable wild QTL alleles have been detected for more than 45% of the evaluated traits [34]. These data suggest that continued sampling of exotic germplasm should guarantee the discovery of new and useful QTL alleles. Another advantage of the AB-QTL method is that once favorable QTL alleles are detected in segregating populations (i.e., BC_2 or BC_3), few additional marker-assisted generations are required to develop near isogenic lines (NILs) or introgression lines (ILs) that can be phenotyped in replicated trials in order to confirm the QTL effect and subsequently be used for variety development. Numerous QTL-NILs or ILs have been developed starting from the tomato AB-QTL mapping populations, and several of them have been characterised for numerous quantitative traits, for example, [35, 36].

Since exotic germplasm is an important source of favorable alleles for the improvement of quantitative traits, introgression lines (ILs) developed in tomato have a special significance. This also supports the proposal by Zamir [37] for investment in the development of "exotic libraries." Each such library consists of a set of ILs, each IL carrying a single marker-defined chromosome segment derived from a donor exotic parent in an otherwise homogeneous elite genetic background. The alien segment in each IL generally also carries a specific gene, preferably in homozygous condition. A set of overlapping ILs would together represent the entire genome of the donor parent and several such sets of ILs constitute a permanent genetic resource, since they can be maintained by self-pollination.

In tomato, the first exotic library ensuring whole genome coverage was developed by Eshed and Zamir [31] from the cross between the wild green-fruited species *S. pennellii* (acc. LA716) and the cultivated tomato *S. lycopersicum* (cv. M82). Presently, this library consists of 76 RFLP-defined ILs which partition the entire genetic map into 107 bins defined by single or overlapping segments [38]. Over the past 15 years, the *S. pennellii* ILs and their hybrids have been phenotyped for more than one hundred traits. For 20 different characters, such as yield, fruit morphology, and biochemical traits, repeated measurements are available, and the resulting data are presented, in silico, in a search engine called "Real Time QTL" [39] (http://zamir.sgn.cornell.edu/Qtl/Html/home.htm).

The studies conducted on the *S. pennellii* IL library have highlighted the higher QTL mapping power of "exotic libraries" compared with conventional segregating populations. Moreover, ILs have shown to be a powerful genetic tool to study the epistatic interactions among QTLs [40], to obtain more precise estimates of the magnitude of QTL x genetic background interaction [31, 36, 41], and of QTL x environment interaction [34, 36, 38, 41]. The high-resolution mapping approach applied to *S. pennellii* ILs has led to the map-based cloning of the first two QTLs ever cloned: the fruit weight QTL, *fw2.2* [42], and the sugar yield QTL, *Brix9-2-5* [43].

More recently, the *S. pennellii* "exotic library" is now being used to identify the genes that influence heterosis [44, 45]. Furthermore, MAS pyramiding of valuable wild QTLs in the genetic background of cultivated tomato has demonstrated to be a successful approach for developing breeding lines that can significantly outperform leading commercial hybrids under both wet and dry field conditions [41]. The outcome of the application of the IL breeding concept has been the development of a new processing tomato hybrid that is currently the leading variety in California (http://www.ptab.org/ranking9.htm) (D. Zamir, personal communication).

The advantages of the IL approach have motivated the public and private sectors to invest in the development of new library resources starting from interspecific crosses with different wild species of tomato including *S. habrochaites*, *S. pimpinellifolium*, *S. lycopersicoides*, and *S. chmielewskii* [35, 46–49].

To enhance the rate of progress of introgression breeding in tomato, within the framework of a currently running EU project (EU-SOL) (http://www.eu-sol.net), a genetic infrastructure of "exotic libraries" is being further refined from a diverse selection of accessions. Moreover, the IL populations are being anchored to a common PCR marker-based framework, mostly consisting of COSII markers, which will facilitate QTL identification, mapping, cloning of the underlying genes, and the use of the novel variation in marker-assisted breeding.

3. STRATEGIES FOR TOMATO GENOME SEQUENCING

The tomato genome is being sequenced as the cornerstone of an International Solanaceae Genome Initiative, a project that aims at developing the family Solanaceae as a model for systems biology for understanding plant adaptation and diversification (see International Solanaceae Genome Initiative white paper (http://sgn.cornell.edu/solanaceae-project)). A sequencing strategy on a BAC by BAC basis of approximately 220 Mb euchromatin was proposed. The tomato genome comprises approximately 950 Mb of DNA—more than 75% of which is heterochromatin and largely devoid of genes [2]. Most genes are found in long contiguous stretches of gene dense euchromatin located on the distal portions of each chromosome arm. The starting points for sequencing the genome are approximately 1500 "seed" BAC clones individually anchored to the tomato map by means of overgo markers. Since most of the genetic markers anchoring the BAC correspond to genes (or ESTs), the BAC are likely to be biased towards the euchromatic portion of the genome.

The division of sequencing activities between countries was effected on a chromosome basis (see Figure 1). Funding agencies of each country supported the sequencing of corresponding chromosomes. Additional funds to complete this task were provided for European countries by the EU-SOL project.

In order to facilitate the sequencing task, marker analysis strategies, cytogenetic protocols, and a number of bionformatics and molecular tools have been developed in recent years. Most of the genome sequencing resources, such as BAC libraries (LE_HBa *Hin*dIII library, SL_*MboI*, and SL_*Eco*RI, based on the Heinz 1706 genotype) and web repositories, were provided by different partners. Seed BAC and contigs were mapped to each chromosome at Cornell University by means of overgo probes. A fingerprint map of approximately 10X genome equivalents from the LE_HBa library has been constructed at the University of Arizona through funding from the National Science Foundation (http://www.genome.arizona.edu/fpc/tomato). Recently, a Sanger Initiative was focussed on the generation of additional fingerprint data from the *Mbo*I library in order to allow comparison and integration of the two datasets. Fluorescent in situ hybridization (FISH) was provided to help guide the extension of the tiling path through the euchromatic arms of each chromosome and to determine the location of heterochromatin regions [22, 50]. Validation of single BAC assigned to each individual chromosome arm was also performed in different participant countries. Moreover, a 3D pooling library to perform new BAC screening was developed in Japan [10].

Starting in 2005, during the last two years of tomato sequencing activity, the participant countries set up their own sequencing pipelines and started to construct the sequence scaffold of assigned chromosomes. Before starting the sequencing work according to mapping information available at SGN, seed BAC were selected using different strategies (IL mapping, internal sequencing strategy, FISH localization). After a low-coverage sequencing of each seed BAC, the construction of a minimal tiling path of BAC clones was performed by BLASTing the sequence of each "seed" BAC against the BAC-end STC Database to identify BAC with minimal overlap in either directions. The BAC-end Database consisting of 200,000 clones (from *Eco*RI, *Hind*III, and *Mbo*I libraries) was used both to confirm and extend the euchromatin minimal tiling path (e-MTP). Each BAC-end sequence was subjected to automated annotation to determine the proportion of ends that are likely to correspond to genic regions. To improve this process, different strategies have been developed. In the Netherlands, BAC walking was supported using a sequence-tagged connector approach based on AFLP fingerprinting as outlined in Peters et al. [51]. In Japan, selected BAC Mixture (SBM) shotgun sequencing has been set up. In

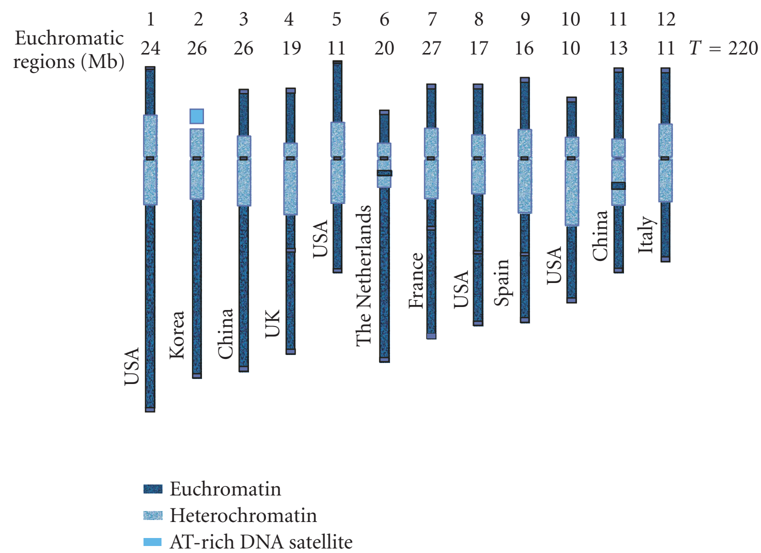


Figure 1: Countries partecipating to the genome sequencing project. For each chromosome the distribution of euchromatic portions is also indicated (modified from the International Solanaceae Genome Initiative white paper http://sgn.cornell.edu/solanaceae-project).

this method, BAC clones whose end sequences do not contain repeat sequences are selected: then the selected BACs are mixed and shotgun sequencing is performed [10]. In our own laboratory, the use of combined bioinformatics tools and molecular data to select a minimum tiling path has been proven to reduce the overlap between adjacent clones. Good extension candidate BAC have been selected using the software "BacEnds Extension v 0.1" [52], which is complementary to the SGN Online BLAST Interface. Following selection based on bioinformatics analysis, and using the IL mapping strategy, chromosome location of the selected extending BAC was experimentally confirmed. Also, the detection of SNPs between *S. lycopersicon* and *S. pennellii* in both the resequenced anchor marker region and the BAC-ends allowed positioning of each extending BAC on chromosome 12 (see Figure 2). Despite a nonuniform distribution of seed BAC on chromosome 12, small contigs consisting of overlapping BAC started to emerge. Currently, sixty-five BACs are in different sequencing phases, and 20 of them will be available on public databases by the end of 2007. For 15 seed BAC, at least one round of extension was performed; in some cases, where two or three rounds of extension were performed, overlapping BACs were merged in sequence islands of >300 kb. The contig of approximately 500 kb between markers T1045 and T1211 was also filled up. Interestingly, the ratio of physical and genetic distance in this region is 250 kb/cM.

In all, a draft sequence has been constructed for approximately 24% of the tomato euchromatic genome space, including all the twelve chromosomes. Sequence islands spanning the genome are being joined and edited in large regions. Progress can be viewed through the development of the TPF and AGP files, available from the SGN repository. The TPF indicates the expected relative positions of the BAC and the AGP provides assembly information. The best coverage is on chromosome 2, where the sequencing of 141 BAC is finished. Large portions of chromosomes 4 (with 77 BAC already sequenced) and chromosome 8 (with 91 BAC already sequenced) have also been sequenced. In all, more than 800 BAC are in different phases of sequencing, and sequences belonging to more than 500 BAC accounting for approximately 21% of total BAC have already been submitted to the SGN website.

Assuming that work on the tomato genome project will continue at the current pace, high-quality sequencing of the euchromatic space should be completed within the next one or two years (by 2008 or 2009). Since the euchromatic portion of the genome is estimated to be approximately 220 Mb, the average physical distance between two adjacent seed BAC should be as little as 200 kb. However, the available map has insufficient density and resolution to provide a template for complete sequencing, since there are large chromosome regions, which are not yet targeted with markers. Therefore, in order to complement the ongoing sequencing project, several new strategies have been undertaken. For instance, selection of additional seed BAC with different verification methods (e.g., IL mapping, FISH, etc.) has been proposed. The recent release of markers from Syngenta to the SGN repository also allowed the identification of new candidate seed BAC, which are distributed throughout the full genome. This may prove useful for filling in gene spaces at a later stage of the project. Whole genome shotgun sequencing and the availability of new generation sequencing technologies, including 454/Roche's sequencer FLX, Solexa's sequencing system, and ABI's SOLiD, may also prove useful in completing the whole genome sequencing of the tomato genome.

4. FUNCTIONAL GENOMICS

In order to understand the function of specific genes and their role in metabolic pathways, as also to identify the key steps in their coregulation mechanisms, several approaches

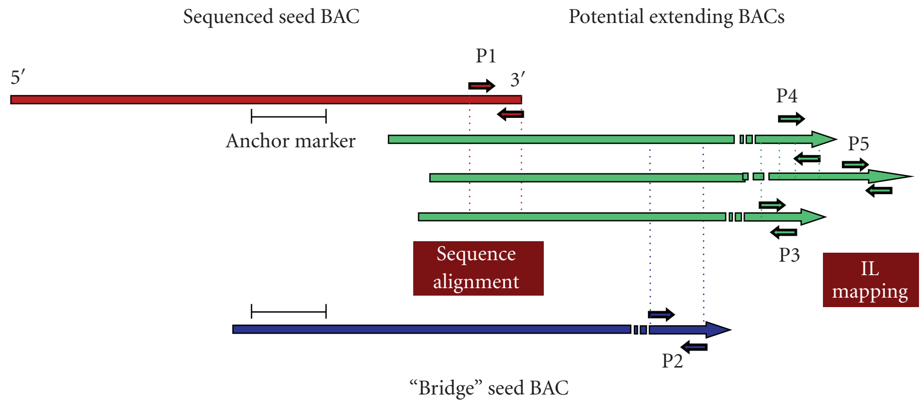


Figure 2: Strategy for selection of candidate extending BACs based on combined use of bioinformatics analysis and experimental molecular data. Specific primer pairs were designed on sequenced seed BAC-end (P1 pair), on "bridge" seed BAC (P2 pair), and on potential extending BACs (P3, P4, P5 pairs), selected from the BAC-end database. All primer pairs were tested to amplify fragments from candidate extending BACs. Sequences of P1 and P2 fragments were aligned to the sequenced seed BAC and to the "bridge" seed BAC, respectively, to confirm both overlapping and direction of candidate extending BACs. Amplification of P3, P4 and P5 fragments allowed the selection of the longest extending BAC. IL mapping through specific IL-12 lines confirmed the position of the selected extending BAC on chromosome 12.

have been exploited, including mutagenesis, genetic transformation, and transcriptome analysis.

4.1. Insertional mutagenesis

Both classical and insertional mutageneses have been used in tomato. Indeed, together with barley, *Arabidopsis*, and maize, tomato was the focus of early, extensive mutagenesis programs. In a paper published in 1964, Hans Stubbe reviewed over 250 tomato mutants arising from the seminal work of the Gatersleben group [53]. To date, over 600 characterized monogenic mutations are available in a variety of genetic backgrounds at the Tomato Genetics Resource Center (http://tgrc.ucdavis.edu). More recently, an extensive mutant population consisting of 6000 EMS-induced and 7000 fast neutron-induced mutant lines has been obtained (54). This population is probably saturating. For instance, extensive allelic tests confirmed that all the *wiry* mutants with 3 to 7 alleles present in TGRC are represented in the population. Two new *wiry* loci have also been described in the collection, each with 10 alleles. A detailed phenotypic description of the mutants is available online (http://zamir.sgn.cornell.edu/mutants).

Insertional mutagenesis systems exploiting exogenous transposon systems have also been described in tomato [55–57]. Nevertheless, these systems, some of which utilize the Micro-Tom cultivar, have not yielded saturating mutant collections and have thus not been utilized extensively. Highly efficient protocols for transformation of Micro-Tom have been described [58], which may serve as a tool for extensive T-DNA mutagenesis programs also.

4.1.1. Targeting induced local lesions IN genomes

In addition to the above, a more recent strategy called targeting induced local lesions IN genomes (TILLING) was described by McCallum et al. [59] for targeting local mutations in the genome. This is, a PCR-based strategy that provides an allelic series of induced point mutations in genes of interest. As such, it can be applied to most organisms, even to those for which an efficient transformation system is not available. TILLING has been used for high-throughput isolation of mutants in Arabidopsis [60] as well as in several crop plants [61]. TILLING platforms for tomato are under development in several countries, including the US, France, Italy, and India. The Franco-Italian effort is coordinated by the EU-SOL project (http://www.eu-sol.net).

4.2. Gene silencing (RNAi and VIGS)

Strategies for gene silencing have also been widely used as a tool for functional genomics research in tomato. Indeed, tomato fruit ripening was one of the early systems in which both sense and antisense silencing were found to be effective [62, 63]. More recently, RNA interference (RNAi) and virus-induced gene silencing (VIGS) have also been successfully used as functional genomics tools in tomato. Interestingly, the use of RNAi remains confined in the fruit, thus making the fruit-specific silencing of genes possible [64]. Similarly, VIGS has been described in tomato roots [65] and fruits [66] although the extent to which silencing remains confined to these organs has not been extensively investigated. Several viral vectors have been used, including Tobacco rattle virus (TRV) [67], Tomato yellow leaf curl China virus isolate [68], and potato virus X [69]. Of these, TRV displays the widest host range, allowing silencing in several *Solanum* species [70], as well as in non-Solanaceous species like opium poppy [71] and *Arabidopsis* [72].

4.3. Transient expression of exogenous genes

Transient expression of exogenous genes has also been achieved through several transient transformation techniques, such as particle bombardment or agroinfiltration.

Recently, an agroinjection technique was developed for tomato fruits [73], which allow the functional analysis of several genes in fruits in a short time. This technique has been used both for expression of exogenous genes and for TRV-induced gene silencing.

4.4. Transcriptional profiling

Finally, transcriptional profiling is being widely explored since the extensive EST collection available in tomato [4] has allowed designing of several microarray platforms: the most widely used to date has been Tom1, a cDNA-based microarray containing probes for approximately 8000 independent genes; and Tom2, a long oligonucleotide-based microarray containing probes for approximately 11000 independent genes. Both these microarrays are already available from BTI (http://bti.cornell.edu/CGEP/CGEP.html) and soon Tom2 will also be available from the EU-SOL project (http://www.eu-sol.net). The third array is an Affymetrix Genechip, which contains probe sets for approximately 9000 independent genes (http://www.affymetrix.com/products/arrays/specific/tomato.affxspecific/tomato.affx). As the tomato genome project progresses, a comprehensive, public tomato microarray platform will become indispensable.

5. BIOINFORMATICS

In order to address key questions arising from the SOL initiative, an overwhelming amount of data from different "*omics*" approaches is being generated and can be utilized for genomics research. Therefore, bioinformatics approaches assume major importance in order to convert raw data into biologically meaningful information. The SOL network is planning a bioinformatics infrastructure that should support integration of information from Solanaceae research into a "one-stop shop" on the web. This will ultimately allow Solanaceae biology to be approached from a systems biology perspective. The bioinformatics centers in the SOL network are all involved in building this infrastructure. It will rely on web service approaches [74] to implement a virtual online center of information dedicated to Solanaceae.

The preliminary effort of the bioinformatician in the SOL network is mainly focussed on setting up a distributed annotation pipeline to provide a high-quality, information-enriched tomato genome. For this purpose, the International Tomato Annotation Group (ITAG) has been constituted, which is a collaborative effort in annotating the tomato genome. It involves several groups from Europe, Asia, and the US. These groups of scientists are organizing data and sharing methodologies to provide a reliable tomato genome annotation. The ITAG annotation pipeline (http://www.sgn.cornell.edu/sequencing/ITAG/status_html.pl) is currently being developed through work on batches of sequences, which are generated at the SGN by grouping BAC which are being submitted by each sequencing center. Analyses, such as repeat masking, EST alignment and gene predictions, are performed on the BAC. These data are fed into the EuGene combiner software [75] for homology searches using protein or genomic sequences from other species. The resulting predicted genes are then functionally annotated based on homology searches, protein domain identifications, and Gene Ontology assignments. Each consortium member takes on different tasks according to a predefined job distribution and in accordance with its specific expertise.

The preliminary effort of a genome annotation exercise requires a definition of reliable gene models from tomato to support the training of gene predictors. The definition of a tomato-specific set of reference gene models is a necessary step towards reliable predictions and a preliminary task of the ITAG. EST/cDNA sequences can be fully exploited if they are first clustered and assembled into high-quality consensus and are then properly aligned against genomic sequences. The organization of tomato and other Solanaceae EST collections is a prerequisite to provide a preliminary annotation of the tomato genome, which is supported by experimental evidence.

Several specific EST repositories from *S. lycopersicum* are available worldwide (Table 1). The TIGR Tomato Gene Index (LeGI) is a collection of virtual tentative consensus (TC) sequences constructed by clustering and assembling 213,974 ESTs and 2,043 ETs (release 11) generated in several laboratories, including the TIGR Institute, Cornell University, and the Boyce Thompson Institute.The SOL Genomics Network (SGN) [4], a website dedicated to the biology of the Solanaceae, organizes and distributes ESTs (∼239,593), sequenced from 35 different cDNA libraries from *S. lycopersicum* (32), *S. pennellii* (2), *S. habrochaites* (1), and the corresponding "combined" consensus sequences. Other EST resources are as follows: (i) the Tomato Stress EST Database (TSED), which contains ESTs from more than ten stress-treated substractive cDNA libraries from *S. lycopersicum*; (ii) the Micro-Tom Database (MiBASE) [76], which distributes unigenes obtained by assembling 35,824 Micro-Tom (a miniature and dwarf tomato cultivar) ESTs from full-length cDNA libraries and 150,581 ESTs from other tomato lines; (iii) the PlantGDB [77], which collects PlantGDB-assembled Unique Transcripts (PUT) from many different species including those of *S. lycopersicum* generated from EST sequences available at the NCBI dbEST Database [78].

CAB-developed TomatEST [79], a secondary database of EST/cDNA sequences, today contains 112 libraries from all the tomato species available at dbEST. TomatEST has been designed to provide a workbench for mining the complexity of EST sequences information content from multiple tomato species. This will then allow expression pattern analysis and gene discovery in the framework of the *S. lycopersicum* genome project.

The CAB group within the EU-SOL project (http://www.eu-sol.net) is committed to collect all EST data from Solanaceae species available in dbEST (Table 1) and to provide EST alignments to the tomato genome draft sequences under production. Also ESTs from two species of the *Coffea* genus (Rubiaceae) were considered in the CAB collection, since coffee and tomato share common gene repertoires [80]. A specific tool has been designed to remove the over-represented EST sequences from each of the 16 species collections in order to clip the original datasets and produce nonredundant sets of sequences. These EST collections are

Table 1: Solanaceae EST resources in the world

Resource	Web address	Species included
Solanaceae Genomics Network (SGN)	www.sgn.cornell.edu	Tomato, potato, pepper, eggplant, and petunia
TIGR Plant Gene Indices	compbio.dfci.harvard.edu/tgi/plant.html	Tomato, potato, pepper, petunia, tobacco, *and N. benthamiana*
PlantGDB—Plant Gene Indices	www.plantgdb.org/prj/ESTCluster/index.php	Tomato, potato, petunia, and different species of *Nicotiana* genus
Tomato Stress EST Database (TSED)	ibs.sinica.edu.tw/ibsdb/app_all/index.php	Tomato ESTs from stress-treated substractive cDNA libraries
MIBASE—Micro-Tom Database	www.kazusa.or.jp/jsol/nicrotom/indexj.html	Micro-Tom EST libraries
Italian Solanaceae EST Database	biosrv.cab.unina.it	Tomato, potato, and other Solanaceae species EST collections

independently processed by the ParPEST pipeline [81] to generate *tentative consensus sequences* (TCs).

The definition of gene models based on EST/cDNA data is a complex effort. D'Agostino et al. [82] proposed a methodology based on the analysis of spliced alignments of EST and tentative consensus sequences to automatically define a reliable dataset of gene models in tomato. Suitable methodologies are used for EST clustering and assembling, and for EST/TC to genome mapping; and resulting conflicts and ambiguities are independently classified.

In the current update of the tomato genome, 582 TCs of *S. lycopersicum* have been selected as they are consistently supported by EST evidence [81]. Among these TCs, 70 cover at least 95% of the length of the most similar protein sequence; 168 cover at least 50%; 257 cover less than 50% of the matching protein, and the remaining 87 show no significant similarity to known proteins. Considering TCs from other tomato species among those mapping to the *S. lycopersicum* BAC, additional 21 loci are located. The number accordingly increases to 251 loci if the potato TC sequences are also considered. Only the TCs covering 95% of the length of the matching protein are selected as *complete* reliable gene models for training gene predictors. To date they account for a total of 108 gene models.

In order to contribute to the tomato genome annotation and to accomplish the requirements of an efficient data integration, the CAB group developed ISOL@ (Chiusano et al., unpublished results), an Italian bioinformatics resource for Solanaceae genomics. This effort is conceived to support the analysis of the genome organization and its functionality in the light of evolutionary approaches over the entire Solanaceae family.

ISOL@ is currently organized into two main levels: the *genome* level and the *expression* level. The cornerstone of the genome level is represented by the tomato genome draft sequences. The founding elements of the expression level are the Solanaceae EST collections and the oligonucleotide probe sets, which have been generated for the production of the tomato expression microarrays (http://www.affymetrix.com/products/arrays/specific/tomato.affx). A nonstop crosstalk between the levels is based on data source sharing and on integration of the information, which belongs to the respective under parts. Each level can be independently accessed through specific web interfaces which allow user-driven data investigation (http://biosrv.cab.unina.it/isola/isola.html).

In order to provide a preliminary annotation of the BAC sequences while waiting for the whole genome annotation that will be provided by the ITAG, the CAB group has set up an automated annotation pipeline in order to ensure daily retrieval of new *S. lycopersicum* BAC sequences from GenBank, which are used to feed the genome annotation process. The BAC annotation process aims at identifying genes and other genetic elements on the draft genome sequence. The protein coding "gene finding" process is exclusively based on spliced-alignments of expressed sequence tags (ESTs) to the *S. lycopersicum* genomic sequences that are also provided to the distributed pipeline, which was set up within the ITAG effort. To accomplish this task, ESTs from the different plant source collection at CAB (Solanaceae and Rubiaceae species), and the corresponding TCs, created by assembling ESTs in a cluster, are both used. Noncoding RNAs (ncRNAs) from the Rfam collection [83] are also aligned to the genomic sequences. The TIGR Solanaceae Repeats Database [84] is the resource selected for the identification of repetitive sequences in the *S. lycopersicum* genome. The Italian platform also includes alignments of all the RNA sequences from Arabidopsis to the tomato genomic sequences in order to identify genes that are conserved between the two species.

The collection, as of July 2007, comprises 186 BAC sequences which have been uploaded from all the sequencing centers to the GenBank Database. The BAC sequences collected and annotated are released to the scientific community through the Gbrowse [85] web application at CAB (http://biosrv.cab.unina.it/GBrowse). Tracks showing annotations and other features are displayed and cross-linked with other local or external databases which can be explored through web interfaces.

This above effort aims at producing a computational workbench for the analysis of the organization, the functionality, and the evolution of genomes in the Solanaceae family. In addition, it provides experimental biologists with a preliminary annotation of tomato genome data, and represents a reference point while sequencing the tomato genome. Indeed, the crosstalk between the sequencing data and the computationally defined TCs may support BAC extension [86] and a preliminary evaluation of the gene content of each BAC under sequencing.

6. PERSPECTIVES

In the "-omics" age, strategies for integrated genomics that include DNA sequence mining, expression profiling data, and functional and molecular diversity analyses of candidate genes, combined with the use of introgression lines, can increase the efficiency in discovery, candidate gene identification, and cloning of target QTLs [87]. Given the large amount of data that is being generated and will be generated in future, a current bioinformatics challenge is to develop user-friendly bioinformatics management systems that will allow description of the genetic components of subtle QTLs and their integration with genome information including gene content, expression, and function [41]. The development of tomato into a model Solanaceous plant, with a large collection of genetic and genomic tools and a high-quality reference genome sequence, and the high throughput sequencing of 100 additional Solanaceae genomes (SOL-100 project) will facilitate the understanding of the incredible ecological and morphological adaptation exhibited by the Solanaceae family.

ACKNOWLEDGMENTS

Tomato chromosome 12 sequencing is supported by the Agronanotech Project (MIPAF, Italy), by the FIRB project (MUR, Italy), and by the EU-SOL project (European Union) and is a coordinated effort between ENEA, the University of Naples, the CNR-IGV of Portici, the University of Padua, and BIOGEN SrL. Contribution no. 120 from CNR-IGV.

REFERENCES

[1] M. J. Michaelson, H. J. Price, J. R. Ellison, and J. S. Johnston, "Comparison of plant DNA contents determined by feulgen microspectrophotometry and laser flow cytometry," *American Journal of Botany*, vol. 78, no. 2, pp. 183–188, 1991.

[2] R. van der Hoeven, C. Ronning, J. Giovannoni, G. Martin, and S. D. Tanksley, "Deductions about the number, organization, and evolution of genes in the tomato genome based on analysis of a large expressed sequence tag collection and selective genomic sequencing," *The Plant Cell*, vol. 14, no. 7, pp. 1441–1456, 2002.

[3] L. A. Mueller, S. D. Tanskley, J. J. Giovannoni, et al., "The tomato sequencing project, the first cornerstone of the International Solanaceae Project (SOL)," *Comparative and Functional Genomics*, vol. 6, no. 3, pp. 153–158, 2005.

[4] L. A. Mueller, T. H. Solow, N. Taylor, et al., "The SOL genomics network. A comparative resource for Solanaceae biology and beyond," *Plant Physiology*, vol. 138, no. 3, pp. 1310–1317, 2005.

[5] A. Frary, Y. Xu, J. Liu, S. Mitchell, E. Tedeschi, and S. D. Tanksley, "Development of a set of PCR-based anchor markers encompassing the tomato genome and evaluation of their usefulness for genetics and breeding experiments," *Theoretical and Applied Genetics*, vol. 111, no. 2, pp. 291–312, 2005.

[6] D. Shibata, "Genome sequencing and functional genomics approaches in tomato," *Journal of General Plant Pathology*, vol. 71, no. 1, pp. 1–7, 2005.

[7] T. M. Fulton, R. van der Hoeven, N. T. Eannetta, and S. D. Tanksley, "Identification, analysis, and utilization of conserved ortholog set markers for comparative genomics in higher plants," *The Plant Cell*, vol. 14, no. 7, pp. 1457–1467, 2002.

[8] F. Wu, L. A. Mueller, D. Crouzillat, V. Pétiard, and S. D. Tanksley, "Combining bioinformatics and phylogenetics to identify large sets of single-copy orthologous genes (COSII) for comparative, evolutionary and systematic studies: a test case in the euasterid plant clade," *Genetics*, vol. 174, no. 3, pp. 1407–1420, 2006.

[9] M. J. M. Smulders, G. Bredemeijer, W. Rus-Kortekaas, P. Arens, and B. Vosman, "Use of short microsatellites from database sequences to generate polymorphisms among *Lycopersicon esculentum* cultivars and accessions of other Lycopersicon species," *Theoretical and Applied Genetics*, vol. 94, no. 2, pp. 264–272, 1997.

[10] E. Asamizu, "Tomato genome sequencing: deciphering the euchromatin region of the chromosome 8," *Plant Biotechnology*, vol. 24, no. 1, pp. 5–9, 2007.

[11] W. Yang, X. Bai, E. Kabelka, et al., "Discovery of single nucleotide polymorphisms in *Lycopersicon esculentum* by computer aided analysis of expressed sequence tags," *Molecular Breeding*, vol. 14, no. 1, pp. 21–34, 2004.

[12] J. A. Labate and A. M. Baldo, "Tomato SNP discovery by EST mining and resequencing," *Molecular Breeding*, vol. 16, no. 4, pp. 343–349, 2005.

[13] S.-C. Sim, W. Yang, E. van der Knaap, S. Hogenhout, H. Xiao, and D. M. Francis, "Microarray-based SNP discovery for tomato genetics and breeding," in *Plant & Animal Genome XV Conference*, p. 173, San Diego, Calif, USA, January 2007.

[14] Linkage Committee, "Linkage summary," *Tomato Genetics Cooperative*, vol. 23, pp. 9–11, 1973.

[15] S. D. Tanksley, M. W. Ganal, J. P. Prince, et al., "High density molecular linkage maps of the tomato and potato genomes," *Genetics*, vol. 132, no. 4, pp. 1141–1160, 1992.

[16] J. A. Labate, S. Grandillo, T. Fulton, et al., "Tomato," in *Genome Mapping and Molecular Breeding in Plants*, C. Kole, Ed., pp. 1–125, Springer, New York, NY, USA, 2007.

[17] S. D. Tanksley, R. Bernatzky, N. L. V. Lapitan, and J. P. Prince, "Conservation of gene repertoire but not gene order in pepper and tomato," *Proceedings of the National Academy of Science of the United States of America*, vol. 85, no. 17, pp. 6419–6423, 1988.

[18] M. W. Bonierbale, R. L. Plaisted, and S. D. Tanksley, "RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato," *Genetics*, vol. 120, no. 4, pp. 1095–1103, 1988.

[19] J. P. Prince, E. Pochard, and S. D. Tanksley, "Construction of a molecular linkage map of pepper and a comparison of synteny with tomato," *Genome*, vol. 36, no. 3, pp. 404–417, 1993.

[20] K. D. Livingstone, V. K. Lackney, J. Blauth, R. van Wijk, and M. K. Jahn, "Genome mapping in Capsicum and the evolution of genome structure in the Solanaceae," *Genetics*, vol. 152, no. 3, pp. 1183–1202, 1999.

[21] S. Doganlar, A. Frary, M.-C. Daunay, R. N. Lester, and S. D. Tanksley, "A comparative genetic linkage map of eggplant (*Solanum melongena*) and its implications for genome evolution in the Solanaceae," *Genetics*, vol. 161, no. 4, pp. 1697–1711, 2002.

[22] J. D. Sherman and S. M. Stack, "Two-dimensional spreads of synaptonemal complexes from solanaceous plants—VI: high-resolution recombination nodule map for tomato (*Lycopersicon esculentum*)," *Genetics*, vol. 141, no. 2, pp. 683–708, 1995.

[23] D. G. Peterson, N. L. V. Lapitan, and S. M. Stack, "Localization of single- and low-copy sequences on tomato synaptonemal complex spreads using fluorescence in situ hybridization (FISH)," *Genetics*, vol. 152, no. 1, pp. 427–439, 1999.

[24] L. C. Harper and W. Z. Cande, "Mapping a new frontier; development of integrated cytogenetic maps in plants," *Functional & Integrative Genomics*, vol. 1, no. 2, pp. 89–98, 2000.

[25] Y. Wang, X. Tang, Z. Cheng, L. A. Mueller, J. Giovannoni, and S. D. Tanksley, "Euchromatin and pericentromeric heterochromatin: comparative composition in the tomato genome," *Genetics*, vol. 172, no. 4, pp. 2529–2540, 2006.

[26] B. Skwarecki, N. Taylor, T. Solow, et al., "The solanaceae genomics network: data, methods, tools, and the tomato genome," 2005, http://sgn.cornell.edu.

[27] N. D. Young, D. Zamir, M. W. Ganal, and S. D. Tanksley, "Use of isogenic lines and simultaneous probing to identify DNA markers tightly linked to the *Tm-2^a^* gene in tomato," *Genetics*, vol. 120, no. 2, pp. 579–585, 1988.

[28] G. B. Martin, J. G. K. Williams, and S. D. Tanksley, "Rapid identification of markers linked to a *Pseudomonas* resistance gene in tomato by using random primers and near-isogenic lines," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 88, no. 6, pp. 2336–2340, 1991.

[29] A. H. Paterson, E. S. Lander, J. D. Hewitt, S. Peterson, S. E. Lincoln, and S. D. Tanksley, "Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms," *Nature*, vol. 335, no. 6192, pp. 721–726, 1988.

[30] M. C. deVicente and S. D. Tanksley, "QTL analysis of transgressive segregation in an interspecific tomato cross," *Genetics*, vol. 134, no. 2, pp. 585–596, 1993.

[31] Y. Eshed and D. Zamir, "An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield- associated QTL," *Genetics*, vol. 141, no. 3, pp. 1147–1162, 1995.

[32] S. D. Tanksley and J. C. Nelson, "Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines," *Theoretical and Applied Genetics*, vol. 92, no. 2, pp. 191–203, 1996.

[33] S. D. Tanksley, S. Grandillo, T. M. Fulton, et al., "Advanced backcross QTL analysis in a cross between an elite processing line of tomato and its wild relative *L. pimpinellifolium*," *Theoretical and Applied Genetics*, vol. 92, no. 2, pp. 213–224, 1996.

[34] S. Grandillo, S. D. Tanksley, and D. Zamir, "Exploitation of natural biodiversity through genomics," in *Genomics-Assisted Crop Improvement: Genomics Approaches and Platforms*, R. K. Varshney and R. Tuberosa, Eds., pp. 121–150, Springer, New York, NY, USA, 2007.

[35] A. J. Monforte and S. D. Tanksley, "Development of a set of near isogenic and backcross recombinant inbred lines containing most of the *Lycopersicon hirsutum* genome in a *L. esculentum* genetic background: a tool for gene mapping and gene discovery," *Genome*, vol. 43, no. 5, pp. 803–813, 2000.

[36] A. J. Monforte, E. Friedman, D. Zamir, and S. D. Tanksley, "Comparison of a set of allelic QTL-NILs for chromosome 4 of tomato: deductions about natural variation and implications for germplasm utilization," *Theoretical and Applied Genetics*, vol. 102, no. 4, pp. 572–590, 2001.

[37] D. Zamir, "Improving plant breeding with exotic genetic libraries," *Nature Reviews Genetics*, vol. 2, no. 12, pp. 983–989, 2001.

[38] Y.-S. Liu, A. Gur, G. Ronen, et al., "There is more to tomato fruit colour than candidate carotenoid genes," *Plant Biotechnology Journal*, vol. 1, no. 3, pp. 195–207, 2003.

[39] A. Gur, Y. Semel, A. Cahaner, and D. Zamir, "Real time QTL of complex phenotypes in tomato interspecific introgression lines," *Trends in Plant Science*, vol. 9, no. 3, pp. 107–109, 2004.

[40] Y. Eshed and D. Zamir, "Less-than-additive epistatic interactions of quantitative trait loci in tomato," *Genetics*, vol. 143, no. 4, pp. 1807–1817, 1996.

[41] A. Gur and D. Zamir, "Unused natural variation can lift yield barriers in plant breeding," *PLoS Biology*, vol. 2, no. 10, p. e245, 2004.

[42] A. Frary, T. C. Nesbitt, A. Frary, et al., "*fw*2.2: a quantitative trait locus key to the evolution of tomato fruit size," *Science*, vol. 289, no. 5476, pp. 85–88, 2000.

[43] E. Fridman, T. Pleban, and D. Zamir, "A recombination hotspot delimits a wild-species quantitative trait locus for tomato sugar content to 484 bp within an invertase gene," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 9, pp. 4718–4723, 2000.

[44] Y. Semel, J. Nissenbaum, N. Menda, et al., "Overdominant quantitative trait loci for yield and fitness in tomato," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 35, pp. 12981–12986, 2006.

[45] Z. B. Lippman and D. Zamir, "Heterosis: revisiting the magic," *Trends in Genetics*, vol. 23, no. 2, pp. 60–66, 2007.

[46] R. Finkers, A. W. van Heusden, F. Meijer-Dekens, J. A. L. van Kan, P. Maris, and P. Lindhout, "The construction of a *Solanum habrochaites* LYC4 introgression line population and the identification of QTLs for resistance to *Botrytis cinerea*," *Theoretical and Applied Genetics*, vol. 114, no. 6, pp. 1071–1080, 2007.

[47] S. Doganlar, A. Frary, H.-M. Ku, and S. D. Tanksley, "Mapping quantitative trait loci in inbred backcross lines of *Lycopersicon pimpinellifolium* (LA1589)," *Genome*, vol. 45, no. 6, pp. 1189–1202, 2002.

[48] M. A. Canady, V. Meglic, and R. T. Chetelat, "A library of *Solanum lycopersicoides* introgression lines in cultivated tomato," *Genome*, vol. 48, no. 4, pp. 685–697, 2005.

[49] J. D. Peleman and J. R. van der Voort, "Breeding by design," *Trends in Plant Science*, vol. 8, no. 7, pp. 330–334, 2003.

[50] M. A. Budiman, S.-B. Chang, S. Lee, et al., "Localization of *jointless*-2 gene in the centromeric region of tomato chromosome 12 based on high resolution genetic and physical mapping," *Theoretical and Applied Genetics*, vol. 108, no. 2, pp. 190–196, 2004.

[51] S. A. Peters, J. C. van Haarst, T. P. Jesse, et al., "TOPAAS, a tomato and potato assembly assistance system for selection and finishing of bacterial artificial chromosomes," *Plant Physiology*, vol. 140, no. 3, pp. 805–817, 2006.

[52] D. Campagna, C. Romualdi, N. Vitulo, et al., "RAP: a new computer program for de novo identification of repeated sequences in whole genomes," *Bioinformatics*, vol. 21, no. 5, pp. 582–588, 2005.

[53] H. Stubbe, "Mutanten der Kulturtomate *Lycopersicon esculentum* Miller V," *Genetic Resources and Crop Evolution*, vol. 12, no. 1, pp. 121–152, 1964.

[54] N. Menda, Y. Semel, D. Peled, Y. Eshed, and D. Zamir, "*In silico* screening of a saturated mutation library of tomato," *The Plant Journal*, vol. 38, no. 5, pp. 861–872, 2004.

[55] M. B. Cooley, A. P. Goldsbrough, D. W. Still, and J. I. Yoder, "Site selected insertional mutagenesis of tomato with maize Ac and Ds elements," *Molecular and General Genetics*, vol. 252, no. 1-2, pp. 184–194, 1996.

[56] R. Meissner, V. Chague, Q. Zhu, E. Emmanuel, Y. Elkind, and A. A. Levy, "Technical advance: a high throughput system for transposon tagging and promoter trapping in tomato," *The Plant Journal*, vol. 22, no. 3, pp. 265–274, 2000.

[57] E. Emmanuel and A. A. Levy, "Tomato mutants as tools for functional genomics," *Current Opinion in Plant Biology*, vol. 5, no. 2, pp. 112–117, 2002.

[58] H.-J. Sun, S. Uchii, S. Watanabe, and H. Ezura, "A highly efficient transformation protocol for Micro-Tom, a model cultivar for tomato functional genomics," *Plant & Cell Physiology*, vol. 47, no. 3, pp. 426–431, 2006.

[59] C. M. McCallum, L. Comai, E. A. Greene, and S. Henikoff, "Targeting induced local lesions IN genomes (TILLING) for plant functional genomics," *Plant Physiology*, vol. 123, no. 2, pp. 439–442, 2000.

[60] B. J. Till, T. Colbert, C. Codomo, et al., "High-throughput TILLING for *Arabidopsis*," *Methods in Molecular Biology*, vol. 323, pp. 127–135, 2006.

[61] A. J. Slade and V. C. Knauf, "TILLING moves beyond functional genomics into crop improvement," *Transgenic Research*, vol. 14, no. 2, pp. 109–115, 2005.

[62] C. J. S. Smith, C. F. Watson, C. R. Bird, J. Ray, W. Schuch, and D. Grierson, "Expression of a truncated tomato polygalacturonase gene inhibits expression of the endogenous gene in transgenic plants," *Molecular and General Genetics*, vol. 224, no. 3, pp. 477–481, 1990.

[63] J. Gray, S. Picton, J. Shabbeer, W. Schuch, and D. Grierson, "Molecular biology of fruit ripening and its manipulation with antisense genes," *Plant Molecular Biology*, vol. 19, no. 1, pp. 69–87, 1992.

[64] G. R. Davuluri, A. van Tuinen, P. D. Fraser, et al., "Fruit-specific RNAi-mediated suppression of *DET1* enhances carotenoid and flavonoid content in tomatoes," *Nature Biotechnology*, vol. 23, no. 7, pp. 890–895, 2005.

[65] T. Valentine, J. Shaw, V. C. Blok, M. S. Phillips, K. J. Oparka, and C. Lacomme, "Efficient virus-induced gene silencing in roots using a modified tobacco rattle virus vector," *Plant Physiology*, vol. 136, no. 4, pp. 3999–4009, 2004.

[66] D.-Q. Fu, B.-Z. Zhu, H.-L. Zhu, W.-B. Jiang, and Y.-B. Luo, "Virus-induced gene silencing in tomato fruit," *The Plant Journal*, vol. 43, no. 2, pp. 299–308, 2005.

[67] Y. Liu, M. Schiff, and S. P. Dinesh-Kumar, "Virus-induced gene silencing in tomato," *The Plant Journal*, vol. 31, no. 6, pp. 777–786, 2002.

[68] X. Tao and X. Zhou, "A modified viral satellite DNA that suppresses gene expression in plants," *The Plant Journal*, vol. 38, no. 5, pp. 850–860, 2004.

[69] L. Giliberto, G. Perrotta, P. Pallara, et al., "Manipulation of the blue light photoreceptor cryptochrome 2 in tomato affects vegetative development, flowering time, and fruit antioxidant content," *Plant Physiology*, vol. 137, no. 1, pp. 199–208, 2005.

[70] G. Brigneti, A. M. Martín-Hernández, H. Jin, et al., "Virus-induced gene silencing in *Solanum* species," *The Plant Journal*, vol. 39, no. 2, pp. 264–272, 2004.

[71] L. C. Hileman, S. Drea, G. de Martino, A. Litt, and V. F. Irish, "Virus-induced gene silencing is an effective tool for assaying gene function in the basal eudicot species *Papaver somniferum* (opium poppy)," *The Plant Journal*, vol. 44, no. 2, pp. 334–341, 2005.

[72] T. M. Burch-Smith, M. Schiff, Y. Liu, and S. P. Dinesh-Kumar, "Efficient virus-induced gene silencing in *Arabidopsis*," *Plant Physiology*, vol. 142, no. 1, pp. 21–27, 2006.

[73] D. Orzaez, S. Mirabel, W. H. Wieland, and A. Granell, "Agroinjection of tomato fruits. A tool for rapid functional analysis of transgenes directly in fruit," *Plant Physiology*, vol. 140, no. 1, pp. 3–11, 2006.

[74] M. D. Wilkinson and M. Links, "BioMOBY: an open source biological web services proposal," *Briefings in Bioinformatics*, vol. 3, no. 4, pp. 331–341, 2002.

[75] S. Foissac, P. Bardou, A. Moisan, M.-J. Cros, and T. Schiex, "EUGÈNE'HOM: a generic similarity-based gene finder using multiple homologous sequences," *Nucleic Acids Research*, vol. 31, no. 13, pp. 3742–3745, 2003.

[76] K. Yano, M. Watanabe, N. Yamamoto, et al., "MiBASE: a database of a miniature tomato cultivar Micro-Tom," *Plant Biotechnology*, vol. 23, no. 2, pp. 195–198, 2006.

[77] Q. Dong, C. J. Lawrence, S. D. Schlueter, et al., "Comparative plant genomics resources at PlantGDB," *Plant Physiology*, vol. 139, no. 2, pp. 610–618, 2005.

[78] M. S. Boguski, T. M. J. Lowe, and C. M. Tolstoshev, "dbEST—database for "expressed sequence tags"," *Nature Genetics*, vol. 4, no. 4, pp. 332–333, 1993.

[79] N. D'Agostino, M. Aversano, L. Frusciante, and M. L. Chiusano, "TomatEST database: *in silico* exploitation of EST data to explore expression patterns in tomato species," *Nucleic Acids Research*, vol. 35, pp. D901–D905, 2007.

[80] C. Lin, L. A. Mueller, J. M. Carthy, D. Crouzillat, V. Pétiard, and S. D. Tanksley, "Coffee and tomato share common gene repertoires as revealed by deep sequencing of seed and cherry transcripts," *Theoretical and Applied Genetics*, vol. 112, no. 1, pp. 114–130, 2005.

[81] N. D'Agostino, M. Aversano, and M. L. Chiusano, "ParPEST: a pipeline for EST data analysis based on parallel computing," *BMC Bioinformatics*, vol. 6, supplement 4, p. S9, 2005.

[82] N. D'Agostino, A. Traini, L. Frusciante, and M. L. Chiusano, "Gene models from ESTs (GeneModelEST): an application on the *Solanum lycopersicum* genome," *BMC Bioinformatics*, vol. 8, supplement 1, p. S9, 2007.

[83] S. Griffiths-Jones, A. Bateman, M. Marshall, A. Khanna, and S. R. Eddy, "Rfam: an RNA family database," *Nucleic Acids Research*, vol. 31, no. 1, pp. 439–441, 2003.

[84] S. Ouyang and C. R. Buell, "The TIGR plant repeat databases: a collective resource for the identification of repetitive sequences in plants," *Nucleic Acids Research*, vol. 32, pp. D360–D363, 2004.

[85] L. D. Stein, C. Mungall, S. Shu, et al., "The generic genome browser: a building block for a model organism system database," *Genome Research*, vol. 12, no. 10, pp. 1599–1610, 2002.

[86] S. Torre, N. D'Agostino, M. L. Chiusano, L. Frusciante, A. Traini, and A. Barone, "BAC sequencing and annotation driven by experimental data: an application for a serine-threonine kinase on tomato chromosome 12," *Acta Horticulturae*, vol. 2007, pp. 449–456.

[87] Z.-K. Li, B.-Y. Fu, Y.-M. Gao, et al., "Genome-wide introgression lines and their use in genetic and molecular dissection of complex phenotypes in rice (*Oryza sativa* L.)," *Plant Molecular Biology*, vol. 59, no. 1, pp. 33–52, 2005.

Hindawi Publishing Corporation
International Journal of Plant Genomics
Volume 2008, Article ID 216513, 9 pages
doi:10.1155/2008/216513

Review Article

Genomic Resources and Tools for Gene Function Analysis in Potato

Glenn J. Bryan and Ingo Hein

Genetics Programme, Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, UK

Correspondence should be addressed to Glenn J. Bryan, glenn.bryan@scri.ac.uk

Received 22 November 2007; Revised 26 August 2008; Accepted 28 September 2008

Recommended by Pushpendra Gupta

Potato, a highly heterozygous tetraploid, is undergoing an exciting phase of genomics resource development. The potato research community has established extensive genomic resources, such as large expressed sequence tag (EST) data collections, microarrays and other expression profiling platforms, and large-insert genomic libraries. Moreover, potato will now benefit from a global potato physical mapping effort, which is serving as the underlying resource for a full potato genome sequencing project, now well underway. These tools and resources are having a major impact on potato breeding and genetics. The genome sequence will provide an invaluable comparative genomics resource for cross-referencing to the other Solanaceae, notably tomato, whose sequence is also being determined. Most importantly perhaps, a potato genome sequence will pave the way for the functional analysis of the large numbers of potato genes that await discovery. Potato, being easily transformable, is highly amenable to the investigation of gene function by biotechnological approaches. Recent advances in the development of Virus Induced Gene Silencing (VIGS) and related methods will facilitate rapid progress in the analysis of gene function in this important crop.

1. INTRODUCTION

Cultivated potato, the world's third most important human food crop, is a tetraploid outbreeder and suffers acutely from inbreeding depression. Genetic mapping is generally performed at the diploid level, using highly heterozygous clones as parents, and several diploid maps of potato have been generated [1], including one of the densest plant genetic maps [2]. Considerable progress has also been made in working at the tetraploid level [3, 4]. These efforts have led to the development of large numbers of molecular markers of all of the main types, which in some cases allow comparison of different potato maps or between potato and the closely related tomato. Genetic mapping has also led to knowledge of locations of many potato genes, notably those conferring resistance to many of the pests and pathogens that present a threat to potato [5] and genes influencing tuber traits [6]. Despite these advances, the lack of described mutational variation for potato is a disadvantage of its outbreeding mating habit, and renders genetic complementation problematic for the majority of genes. However, potato is relatively easy to transform, and so technologies such as overexpression and antisense technology are options for investigating gene function. Results of such experiments are not always so easy to interpret, and improved methods for functional analysis are critical to the future of potato breeding and genetics.

This article provides an overview of genomics resources currently available for potato, and the likely future developments in this area, paying particular emphasis to tools being developed for investigating gene function.

2. BASIC FACTS ABOUT THE POTATO GENOME

Cultivated potato has a chromosome number of $2n = 4x = 48$, and a haploid genome size of ~850 Mb, roughly six times that of *Arabidopsis thaliana* and twice the size of the rice genome [7]. Although small chromosome size has been a limitation for cytogenetic analysis in potato, notable advances have been made using pachytene chromosomes and extended DNA "fibres" for fluorescence in situ hybridization (FISH) [8]. The potato genome is very similar in size to its close relative tomato, and genetic maps of the two species show high levels of macrocolinearity [9].

Information on how well the two genomes are conserved at the microsyntenic level should start to become available as outputs from the respective genome projects accumulate. The tomato genome mainly comprises low-copy-number sequences, which diverged rapidly in evolutionary time [10]. Schweizer et al. [11], who characterised the potato genome in terms of the amounts of different classes of repetitive DNA, suggest that the more highly repeated sequences comprise only 4–7% of the potato genome, suggesting that it was relatively devoid of highly repetitive DNA sequences, thus supporting the earlier tomato study. It is also known that the majority of tomato heterochromatin is found in centromeric regions with almost all of the euchromatic DNA located distally in long uninterrupted tracts, a structural feature likely to be true of potato [12]. Gene isolation and recent BAC-end sequencing efforts are providing the first detailed glimpses of the genome structure in potato. Using BAC-end sequence and full BAC sequence data, it has also been shown that potato (34%) contains considerably less repetitive DNA than tomato (46%), this difference being consistent with relative genome sizes of the two crops (850 versus 1000 Mb, resp.) [13].

3. STRUCTURAL GENOMICS RESOURCES FOR POTATO

3.1. EST resources

The generation of large expressed sequence tag (EST) collections is a primary route for large-scale gene discovery. There have been several efforts to generate EST resources for potato [14–16]. The potato gene index (http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?) contains almost 220000 ESTs, assembled into more than 30 000 "contigs" with over 26 000 singletons. These efforts, while not exhaustive, comprise a major genomics resource for potato researchers, perhaps comprising between 50–70% of the total potato gene "repertoire." These ESTs will form an important source not only for the discovery of candidate genes and genetic markers, but also for the development of microarrays, until the whole genome sequence becomes available in potato. For instance, EST data from a number of different genotypes are also a rich source for the discovery of single nucleotide polymorphism (SNP) and simple sequence repeat (SSR) markers. For example, Tang et al. [17] demonstrate how large numbers of "eSNPs" can be mined from EST data using an SNP discovery pipeline (QualitySNP).

3.2. Large-insert genomic libraries and physical maps

Bacterial artificial chromosome (BAC) libraries have become the main vehicle for performing map-based gene cloning and physical mapping in potato. Several BAC libraries have been constructed from cultivated potato [18] and some of its wild relatives, for example, the diploids *Solanum bulbocastanum* [19], *Solanum pinnatisectum* [20], and the Mexican hexaploid *Solanum demissum* [21]. These libraries represent a potentially useful resource for the study of comparative genome organisation and evolution in potato and the wider Solanaceae. A BAC library has been constructed from the male parent (RH89-039-16) of the cross used to make the ultra-high-density (UHD) genetic map of potato with 10 000 loci [2], and is being used for construction of a genome-wide potato physical map. Other significant developments arising from the use of these BAC libraries include the use of BAC clones and fluorescence in situ hybridization (FISH) to develop chromosome-specific cytogenetic DNA markers for chromosome identification in potato [22].

4. GENE ISOLATION IN POTATO

4.1. Map-based approach

Mapping efforts in potato have also led to the generation of knowledge concerning the genetic architecture of a number of characters, including pest and disease resistance, tuber quality traits, dormancy, tuber shape, and colour. Also, several potato genes have been isolated using a map-based approach [18, 23, 24], with most of these aimed at isolation of major genes for resistance to the more serious pests and pathogens of potato, the late blight pathogen *Phytophthora infestans* (Mont. de Bary), potato cyst nematodes (PCN), and potato virus X (PVX). These activities have necessitated the development of dense genetic maps around the target resistance loci, as well as concomitant generation of genomic resources, such as BAC libraries. These gene cloning efforts have afforded early glimpses into the structure of the potato genome, through the sequencing of a considerable number of large-insert clones. For instance, a study of *Gpa2/Rx1* resistance gene "cluster" provided important information concerning the evolution and structure of *R* gene loci and has shown beyond any doubt that resistances to different pests/pathogens can be coded by structurally similar genes from the same gene cluster.

The *R3* locus, which maps to a cluster of genes for resistance against *P. infestans* and other resistance genes on the short arm of chromosome XI, has shown to comprise two very tightly linked resistance genes (*R3a* and *R3b*) with distinct specificities against *P. infestans* [25]. The *R3* locus was found to be syntenic with the *I2* locus of tomato, and a comparative approach was used to isolate *R3a*, which is constitutively expressed along with some of its paralogous genes [26]. It is highly likely that the same approach will allow the future isolation of other *P. infestans* resistance genes on the same chromosome. Similarly, there are now determined efforts to isolate genes from late blight resistance "hotspots" on other potato chromosomes. A notable example is the recent work on potato chromosome IV, whereby several resistance genes against *P. infestans* map to the same locus [27–29].

These are but a few of several successful map-based gene isolation efforts, but these illustrate how comparative genomics, either between different potato genotypes or between different Solanaceous plant species, can be used as a tool for accelerating the normally laborious task of gene isolation, and they bode well for the future of Solanaceae genomic research. As knowledge of the genome structure of potato and tomato increases, the isolation of such genes should become more facile.

4.2. Candidate gene approach

A candidate gene approach has also been used for isolating plant genes that underlie specific traits [30]. In potato, cloning of the gene *Gro1-4*, which confers resistance to pathotype Ro1 of the cyst nematode *Globodera rostochiensis*, has been achieved using a joint candidate gene/mapping approach [31]. The gene was found to colocalise in a large segregating population with a marker derived from a "resistance-gene-like" sequence. The marker was used to isolate 15 members of a closely related gene family from genomic libraries. By taking into account all available information (inheritance patterns in resistant and susceptible germplasm, mapping data, DNA sequence information), it was possible to reduce the number of candidates to three genes, which were subsequently tested for complementation of a susceptible phenotype by stable transformation. The identified functional gene, a member of the TIR-NBS-LRR class, differs from susceptible members of the same family by 29 amino acid changes. This approach may be used in future for isolation of other resistance genes/QTLs conferring partial and durable resistance to the major potato pests and pathogens.

Another example of the use of candidate gene approach in potato is the isolation of *P* gene that encodes anthocyanin biosynthetic enzyme flavonoid 3′,5′-hydroxylase (f3′5′h), and is responsible for the production of blue/purple anthocyanin pigments in tissues like tubers, flowers, or stems [32]. In this study, a Petunia f3′5′h gene was used to screen a potato cDNA library prepared from purple-coloured flowers and stems. Six positively hybridizing cDNA clones were sequenced and all appeared to be derived from a single gene that shared 85% sequence identity at the amino acid level with Petunia f3′5′h. The potato gene cosegregated with purple tuber colour in a diploid population and was found to be expressed in tuber skin only in the presence of the anthocyanin regulatory locus I. One of the f3′5′h cDNA clone that was placed under the control of a doubled CaMV 35S promoter was also used for transformation of the red-skinned cultivar "Desiree." Tuber and stem tissues that were coloured red in Desiree were purple in nine of 17 independently transformed lines, confirming the hypothesis that the transformed gene corresponded to the P locus.

In another study, DNA sequence variation was analysed at the invGE/GF locus (duplicate invertase genes *InvGE* and *InvGF*) on potato chromosome IX which colocalizes with a cold-sweetening QTL [33]. The study focused on 188 tetraploid potato cultivars, which were assessed for chip quality and tuber starch content. Two closely correlated invertase alleles, *invGE-f* and *invGF-d*, were associated with better chip quality in three breeding populations, and one allele (*invGF-b*) was associated with lower tuber starch content. The potato *invGE* gene was also found to be orthologous to the tomato invertase gene *Lin5*, causal for a fruit-sugar-yield QTL. These results suggested that natural variation for sugar yield in tomato fruits and that for sugar content in potato tubers are controlled by functional variants of orthologous invertase genes.

These few examples clearly demonstrate the potential of using the candidate gene approach in potato. It is also clear that the extensive knowledge of tuber biochemistry and the large number of potato gene sequences should enable its further application for tuber quality traits.

5. POTATO GENOME SEQUENCING

The ultra-high-density (UHD) genetic map of potato [2] forms the underlying framework for construction of a genome-wide physical map of the potato genome. Physical map construction is being carried out in two phases. First, approximately 73 000 clones from a BAC-library have been fingerprinted using a nonselective AFLP-based method. The fingerprint data has been used to assemble the RH BACs into roughly 7000 BAC contigs, with a similar number of "singletons" (i.e., single BAC clones). The second phase entails anchoring of the contigs and single BACs to the UHD map using a BAC pooling method, which should also reduce the number of contigs and increase the average contig size. Subsequent contiging will use a reduced stringency alignment approach which will reduce the number of contigs still further. The integrated genetic and physical map will be the main platform, which will be used for obtaining the DNA sequence of the potato genome. It is expected that approximately 1800 contigs will be anchored to the genetical map, and these scaffolds will be the starting point for genome sequencing. A BAC-end sequence resource, comprising more than 140 000 reads, has also been generated for the project [13]. The ongoing tomato and potato sequencing projects will have huge implications for those working in the Solanaceae, and will further sharpen the requirement for functional genomics tools.

6. ANALYSIS OF POTATO GENE EXPRESSION

A wide range of gene expression technologies have been used by potato researchers. Expression analysis is a discipline that is still very much in transition and it is likely to undergo significant development in the future, notably with recent developments in "next generation" sequencing (NGS) technologies, which have the potential to radically change the way gene discovery is performed.

6.1. cDNA-AFLP

The cDNA-AFLP technique has been used to study gene expression from stolon formation to sprouting in a range of different tissues during the potato tuber life cycle [34, 35]. Approximately 18 000 transcript-derived fragments (TDFs) were observed, and over 200 "process specific" TDFs belonging to different stages of potato tuber life cycle were isolated and sequenced. The sequence similarities of these TDFs to known genes give insights into the kinds of processes occurring during tuberisation, dormancy, and sprouting. This technique is extremely sensitive and can detect differences among gene family members indistinguishable by Northern blotting. A useful advance has been the realization that a large proportion of cDNA-AFLP fragments show genetic polymorphism in segregating populations and can be mapped as transcriptome-derived genetic markers [36]. Importantly, these markers show less centromeric clustering than AFLP markers derived directly from genomic DNA and appear to be targeted specifically to transcriptionally active regions of the genome. This method has been used to perform a large scale survey of genes differentially expressed during the tuber life cycle, and the isolation of some of their promoter regions [37]. Many genes expressed in the tuber life cycle are involved in defence, stress, storage, and signal transduction pathways. Twelve cis-acting elements were identified, and are known to be responsive to environmental stimuli known to play an important role during the tuber life cycle (light, sugars, hormones, etc.). More recently, a potato transcription map, based on cDNA-AFLP and containing approximately 700 TDFs, has been generated [38]. One of the disadvantages of cDNA-AFLP is that it does not provide gene sequence information and requires laborious isolation of gene fragments from polyacrylamide gels for sequence characterization.

6.2. SAGE

Serial analysis of gene expression (SAGE), which generates short cDNA sequence tags [39, 40] using a concatemerization-based method, has been used to examine global gene expression in potato tubers, generating 58 322 sequence tags (of length 19 nucleotides) of which 22 233 were unique [41]. Putative functions were assigned to almost 700 of those tags occurring at least ten times and roughly 70% matched each known potato EST sequence. This technology has the advantage over microarray technology in being an "open" technology, with the possibility of discovering "new" transcripts. Rapid amplification of complementary DNA ends (RACE) cloning was used to verify the reliability of SAGE tag annotation using EST sequences from more than one cultivar. Seventy two per cent of tags represented genes that participated in a known biological process, with the largest group (43%) consisting of transcripts active in physiological processes, about half of which were involved in metabolism. There were no transcripts found which were involved in photosynthesis. Of the 50 most abundant transcripts from the mature tuber, protease inhibitors were the dominant class, which is in good agreement with previous EST projects [14, 15].

The methodologies described briefly in this section are alternatives to the microarrays, which may ultimately be replaced by NGS methods. For example, Emrich et al. [42] recently demonstrated how such technologies can be used to extend significantly the EST resources for maize. The authors used a laser capture microdissection method to isolate rare transcripts from shoot apical meristems and then sequenced the corresponding cDNAs using 454 technology. This type of approach could be used in potato to identify transcripts not present in current EST databases or to extend the range of potato germplasm represented, currently limited to a few cultivars. All expression studies share the "problem" that they are only indicative of the function of particular genes or sets of genes in biological processes, and require functional analyses whereby the function of the candidate genes are compromised or exaggerated in some way (e.g., overexpression, silencing). This issue will be addressed in a subsequent section of this article.

7. MICROARRAYS: TOOLS FOR HIGH-THROUGHPUT GENE EXPRESSION ANALYSIS

7.1. cDNA microarrays

The available potato EST resources comprise an unknown but significant fraction of the gene complement of potato, and are derived from several genotypes, tissues, and environmental influences. A nonredundant set of 10 000 of these ESTs was used by the Institute for Genomic Research (TIGR) to develop a cDNA potato microarray that was made available to the research community at minimal cost. Moreover, the same organisation offered a transcription profiling service to allow the evaluation of these arrays by a wide range of users working on different Solanaceous plant species asking different biological questions. This allowed generation of massive microarray data that is publicly available (http://www.tigr.org/tdb/potato/profiling_service2.shtml#AProcedure). However, this platform had the disadvantage of containing a very small proportion of the potato gene repertoire. Moreover, as the "TIGR array" was based on spotted cDNAs, it was inherently difficult to achieve a high level of reproducibility. Rensink et al. [43] have used this platform to identify genes involved in abiotic stress responses, with more than 3000 genes found to be significantly up- or downregulated in response to at least one of the stress conditions used (cold, heat, salt). In another detailed study, expression of 1315 genes during tuber development was examined, where transient changes in gene expression were found to be relatively uncommon and several new genes were found to be differentially expressed during tuber development [44]. These studies, while informative, highlight the dilemma faced by plant molecular biologists in prioritizing genes for further study from a large number of candidate genes in the absence of genetic information and mutations in target trait genes.

7.2. Oligonucleotide microarrays

Long oligonucleotide arrays that have been manufactured by various technology providers have also been found useful in potato since the use of short oligonucleotide arrays may lead to misinterpretations due to high degree of allelic heterozygosity in this crop. For this purpose, the potato oligo chip initiative (POCI) has selected the Agilent "44K feature platform" system, which was made available for use in 2006. This system is very flexible and allows for redesign of the array as more gene sequence information becomes available. Kloosterman et al. [45] described the design of this platform and demonstrated its utility by analyzing different stages of tuber initiation and growth.

8. FUNCTIONAL STUDIES IN POTATO

Potato geneticists and breeders have generated a great deal of information about the location of genes and QTLs coding for important potato traits, including pest and disease resistance and tuber traits. The volume of gene sequence information, notably from cDNA sequencing and the genome project, will increase rapidly in the coming years. Developments in genetics and structural genomics are beginning to be matched by concomitant development of functional genomics tools. Potato has a strong need for a high-density gene map or a genome sequence, to place gene sequences in their genetic/genomic context. Relatively high-throughput methods are also needed for testing and assessing gene function. The availability of mutant populations of potato will also be of tremendous value in this regard [46]. Potato cultivars are highly heterozygous and contain very high levels of "genetic load." It has been estimated that there is one SNP approximately every 25 bp [47]. If individual alleles can be "isolated" in the homozygous condition, there is no telling what information they would yield about potato biology. The nonavailability of mutants may largely be overcome by recourse to use of diploid self-compatible potato clones for the development of mutant populations or by mining of variant alleles in heterozygous germplasm. Functional studies currently rely on the use of transformation-based techniques or use of viral vector-mediated gene delivery systems for the establishment of information regarding gene function. There have been some recent tantalising developments in functional genetics/genomics tools and resources for potato. Of course gene expression profiling or microarray studies have a role to play in the identification of a pool of candidate genes potentially involved in any given biological process. These methods, in combination with other functional genomics tools such as RNA interference (RNAi), virus-induced gene silencing (VIGS), and activation tagged lines, have the potential to facilitate the identification of the role of thousands of potato genes over the next several years. Furthermore, combining structural genetics approaches (such as QTL and candidate gene mapping) with functional genomics information (such as microarray-derived gene expression data for candidate genes) has great potential for the dissection of many complex, polygenic potato traits.

8.1. Virus-induced gene silencing (VIGS)

Virus-induced gene silencing (VIGS) is a powerful tool for plant functional genomics. VIGS exploits an RNA-mediated antiviral defense mechanism in plants. This phenomenon has been exploited for gene silencing through the use of virus vectors carrying host target genes that are directed against the corresponding plant mRNAs [48]. VIGS is increasingly used to generate transient loss-of-function assays, and is a powerful reverse-genetics tool in functional genomic programs as an alternative to stable transformation. In potato, two viral vectors, potato virus X (PVX) and tobacco rattle virus (TRV), have been successfully utilized for VIGS [49, 50]. Faivre-Rampant et al. [49] have shown that a binary PVX-based vector, pGR106, [51, 52] is effective in triggering VIGS of phytoene desaturase (PDS) in both diploid and cultivated tetraploid *Solanum* species. In this study, silencing was maintained throughout the foliar tissues and tubers and could also be triggered and sustained in in vitro micropropagated tetraploid potato for several cycles and on in vitro generated microtubers. Similarly, PDS silencing with TRV has been observed in cultivated potato, as well as the diploid wild species *S. bulbocastanum* and *S. okadae*, and the distantly related hexaploid *S. nigrum* [50]. In the same study, silencing of known resistance genes (e.g., *R1*, *Rx*, and *RB*) in normally resistant plants yielded a compatible interaction in detached leaf tests. A modification of the leaf inoculation used for both PVX- and TRV-based silencing was demonstrated for TRV in a so-called "agrodrench" method, in which soil adjacent to the plant root is drenched with an Agrobacterium suspension carrying the TRV-derived VIGS vectors [53]. TRV-based silencing of genes such as PDS, a 20S proteasome subunit (PB7) or Mg-protoporphyrin chelatase (Chl H) by agrodrench has been shown to be efficient for different members of the Solanaceae including *Nicotiana benthamiana*, tomato, pepper, tobacco, potato, and petunia.

N. benthamiana provides a particularly suitable model system for Solanaceae species, including potato, as it is highly amenable to manipulations such as VIGS and virus- or Agrobacterium-mediated overexpression of candidate genes (Figure 1) [54]. Indeed, many silencing studies have been conducted in *N. benthamiana* to demonstrate involvement of candidate genes involved in the plant disease resistance (including the hypersensitive response; HR), abiotic stress, cellular signaling, and secondary metabolite biosynthesis [55]. Recently, for example, Gilroy et al. [56], using a combination of VIGS and biochemical approaches, demonstrated that the cysteine protease cathepsin B is required for the HR. Silencing of cathepsin B in *N. benthamiana* prevented programmed cell death (PCD) and compromised disease resistance induced by *Erwinia amylovora* and *Pseudomonas syringae* pv. tomato (Pst) DC3000, two distinct nonhost bacterial pathogens. It also suppressed the HR triggered by transient coexpression of potato *R3a* and *Phytophthora infestans Avr3a* genes but did not compromise the HR triggered by recognition of *Cladosporium fulvum* AVR4 by tomato Cf-4. The ease of silencing in *N. benthamiana* makes it suitable for large scale VIGS experiments. A study of 192 cDNA-AFLP fragments, expressed during the HR following recognition of *Avr4* from *C. fulvum* by tomato *Cf-4*, was conducted in *N. benthamiana* and identified 15 *Avr4*-responsive tomato (ART) fragments that, when silenced, resulted in a compromised HR induced by both *Avr4* in Cf-4 transgenic plants and the *Inf1* gene from *P. infestans* [57]. In addition, silencing of HSP90, a nuclear GTPase, an L19 ribosomal protein, and a nucleotide binding-leucine rich repeat (NB-LRR)-type protein suppressed the HR [57]. Interestingly, silencing of the NB-LRR-type protein NRC1 not only affected the Cf-4/Avr4-induced HR and compromised Cf-4-mediated resistance to *C. fulvum*, but also revealed that this protein is required for the HR induced by the *R* proteins Cf-9, LeEix, Pto, Rx, and Mi [58].

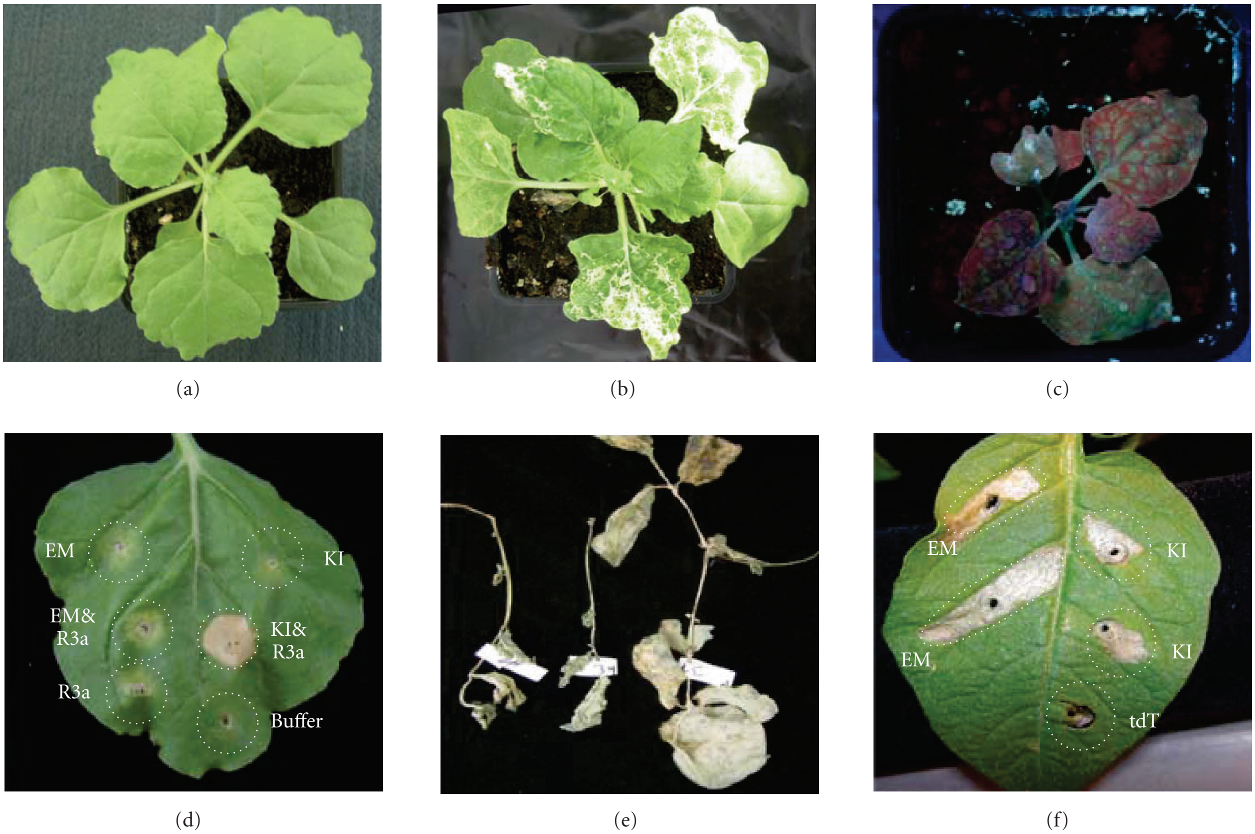


Figure 1: *N. benthamiana* plants, noninoculated (a), inoculated with pGR106::PDS to silence endogenous phytoene desaturase resulting in photo bleaching of leaves (b) and overexpressing GFP via a pGR106::GFP construct—viewed under UV light to show expression of GFP (c). Overexpression and coinfiltration of virulent Avr3a KI and avirulent Avr3a EM with the potato *R* gene R3a are shown in (d). Naturally occurring PVX resistance in *S. papita* (e) and the recognition of virulent Avr3a KI and avirulent Avr3a EM alleles but not from tdT, used as a control, in *S. chacoense* (f).

A recently developed TRV RNA2 vector, which utilizes ligation-independent cloning (LIC), has been employed to assess the function of 400 tomato ESTs in *N. benthamiana* [59]. The function of SlMADS1 and its *N. benthamiana* homologous sequences, NbMADS4-1 and -2, was shown during flowering and demonstrated that NbMADS4-1 and NbMADS4-2 act nonredundantly in floral development. Silencing of either gene resulted in loss of organ identity. These studies show the potential for use of *N. benthamiana* as a "proxy" species for high-throughput gene function analysis for potato and other Solanaceae.

8.2. *Virus and Agrobacterium tumefaciens-based overexpression*

In addition to their role in VIGS, virus vectors can be used for overexpressing genes in plants. The AgrobacteriumPVX-based binary vector pGR106, an efficient silencing vector for Solanum species, can also be used for overexpressing genes, as shown for GFP in Figure 1 [51, 52]. The search for novel sources of plant resistance, driven by knowledge of pathogen "effectors" with avirulent activities, rather than more traditional plant disease resistance breeding, has been coined "effectoromics" [60]. For example, overexpression of *P. infestans* effectors in potato represents an opportunity to seek vital and invariant components of the *P. infestans* pathogenicity apparatus that can be targeted for sustainable potato protection. Information emerging from effectoromic studies will be useful to identify the cognate host *R* genes as sources of durable disease resistance and to develop novel control strategies that are intrinsically difficult for the pathogen to overcome. The discovery of a conserved motif, RxLR, within many avirulence genes [61, 62] that is required for translocation of the effectors from pathogen haustoria into the plant cell [63] has had a tremendous impact on the prediction of pathogen effectors. Overexpression via pGR106 in *N. benthamiana* of 63 predicted *P. infestans* extracellular proteins (Pex) led to the discovery of two novel necrosis-inducing cDNAs, encoding extracellular proteins belonging to a large and complex protein family in *Phytophthora* [64]. Similarly, the recognition of the *P. infestans* effector *Avr3a* by the potato *R* gene *R3a* [26] could be demonstrated in *N. benthamiana* [61]. Coinfiltration of *N. benthamiana* leaves with an *A. tumefaciens* strain carrying a construct expressing *R3a* and a strain carrying a construct expressing the truncated avirulent *Avr3a* (*Avr3a* KI) sequence via PVX resulted in a confluent cell death response, not observed when overexpressing the truncated virulent *Avr3a* (*Avr3a* EM) sequence (Figure 1). Using the *P. infestans* elicitins INF1, INF2A, and INF2B, the same PVX system has been adapted and optimized to screen Solanum plants for response to pathogen elicitors [65]. Of 31 potato species tested, 11 clones of *Solanum huancabambense* and *Solanum microdontum* responded with HR-like symptoms, which were also observed following infiltration with purified recombinant INF1, INF2A, and INF2B.

Two similar studies have been reported that utilize the two *Avr3a* alleles described above to identify potentially novel resistance mechanisms within wild potato accessions [66, 67]. One study [66] utilized PVX to express the different *Avr3a* alleles in wild Solanum species, whereas the other [67] utilized Agrobacterium-only-based expression of the *Avr3a* alleles to circumvent the relative high level of resistance against PVX within the wild species tested (Figure 1). These studies identified similar sets of species that recognize both the EM and KI forms of AVR3a (unpublished data).

9. WHERE NEXT FOR POTATO?

Potato has entered an exciting new era, whereby the development of extensive genetic and genomic resources have opened up many new possibilities for studying important potato traits relevant to potato agronomy. Concomitant development of similar resources for other Solanaceous species, notably tomato, and a growing cohesiveness of the Solanaceae research community, as demonstrated by the "SOL vision" (http://www.sgn.cornell.edu/solanaceae-project/) bode well for future genomic research of potato and its close relatives. Development of biotechnological tools for assaying potato gene function is likely to progress rapidly in the coming years.

ACKNOWLEDGMENTS

The authors would like to acknowledge the financial support of the Scottish Government Rural and Environment Research and Analysis Directorate, and the European Union.

REFERENCES

[1] C. Gebhardt, "Molecular markers, maps, and population genetics," in *Potato Biology and Biotechnology: Advances and Perspectives*, D. Vreugdenhil, Ed., chapter 5, pp. 77–89, Elsevier, Amsterdam, The Netherlands, 2007.

[2] H. van Os, S. Andrzejewski, E. Bakker, et al., "Construction of a 10,000-marker ultradense genetic recombination map of potato: providing a framework for accelerated gene isolation and a genomewide physical map," *Genetics*, vol. 173, no. 2, pp. 1075–1087, 2006.

[3] C. A. Hackett, B. Pande, and G. J. Bryan, "Constructing linkage maps in autotetraploid species using simulated annealing," *Theoretical and Applied Genetics*, vol. 106, no. 6, pp. 1107–1115, 2003.

[4] J. E. Bradshaw, B. Pande, G. J. Bryan, et al., "Interval mapping of quantitative trait loci for resistance to late blight [*Phytophthora infestans* (Mont.) de bary], height and maturity in a tetraploid population of potato (*Solanum tuberosum* subsp. *tuberosum*)," *Genetics*, vol. 168, no. 2, pp. 983–995, 2004.

[5] I. Simko, S. Jansky, S. Stephenson, and D. Spooner, "Genetics of resistance to pests and disease," in *Potato Biology and Biotechnology: Advances and Perspectives*, D. Vreugdenhil, Ed., chapter 7, pp. 117–155, Elsevier, Amsterdam, The Netherlands, 2007.

[6] H. J. van Eck, "Genetics of morphological and tuber traits," in *Potato Biology and Biotechnology: Advances and Perspectives*, D. Vreugdenhil, Ed., pp. 91–115, Elsevier, Amsterdam, The Netherlands, 2007.

[7] K. Arumuganathan and E. D. Earle, "Nuclear DNA content of some important plant species," *Plant Molecular Biology Reporter*, vol. 9, no. 4, p. 415, 1991.

[8] T. Gavrilenko, "Potato cytogenetics," in *Potato Biology and Biotechnology: Advances and Perspectives*, D. Vreugdenhil, Ed., chapter 10, pp. 203–216, Elsevier, Oxford, UK, 2007.

[9] S. D. Tanksley, M. W. Ganal, J. P. Prince, et al., "High density molecular linkage maps of the tomato and potato genomes," *Genetics*, vol. 132, no. 4, pp. 1141–1160, 1992.

[10] D. Zamir and S. D. Tanksley, "Tomato genome is comprised largely of fast-evolving, low copy-number sequences," *Molecular & General Genetics*, vol. 213, no. 2-3, pp. 254–261, 1988.

[11] G. Schweizer, N. Borisjuk, L. Borisjuk, et al., "Molecular analysis of highly repeated genome fractions in *Solanum* and their use as markers for the characterization of species and cultivars," *Theoretical and Applied Genetics*, vol. 85, no. 6-7, pp. 801–808, 1993.

[12] Y. Wang, R. S. van der Hoeven, R. Nielsen, L. A. Mueller, and S. D. Tanksley, "Characteristics of the tomato nuclear genome as determined by sequencing undermethylated EcoRI digested fragments," *Theoretical and Applied Genetics*, vol. 112, no. 1, pp. 72–84, 2005.

[13] W. Zhu, S. Ouyang, M. Iovene, et al., "Analysis of 90 Mb of the potato genome reveals conservation of gene structures and order with tomato but divergence in repetitive sequence composition," *BMC Genomics*, vol. 9, article 286, pp. 1–14, 2008.

[14] M. Crookshanks, J. Emmersen, K. G. Welinder, and K. L. Nielsen, "The potato tuber transcriptome: analysis of 6077 expressed sequence tags," *FEBS Letters*, vol. 506, no. 2, pp. 123–126, 2001.

[15] C. M. Ronning, S. S. Stegalkina, R. A. Ascenzi, et al., "Comparative analyses of potato expressed sequence tag libraries," *Plant Physiology*, vol. 131, no. 2, pp. 419–429, 2003.

[16] B. Flinn, C. Rothwell, R. Griffiths, et al., "Potato expressed sequence tag generation and analysis using standard and unique cDNA libraries," *Plant Molecular Biology*, vol. 59, no. 3, pp. 407–433, 2005.

[17] J. Tang, B. Vosman, R. E. Voorrips, C. G. van der Linden, and J. A. M. Leunissen, "QualitySNP: a pipeline for detecting single nucleotide polymorphisms and insertions/deletions in EST data from diploid and polyploid species," *BMC Bioinformatics*, vol. 7, article 438, pp. 1–15, 2006.

[18] K. Kanyuka, A. Bendahmane, J. N. A. M. R. van der Voort, E. A. G. van der Vossen, and D. C. Baulcombe, "Mapping of intra-locus duplications and introgressed DNA: aids to map-based cloning of genes from complex genomes illustrated by physical analysis of the Rx locus in tetraploid potato," *Theoretical and Applied Genetics*, vol. 98, no. 5, pp. 679–689, 1999.

[19] J. Song, F. Dong, and J. Jiang, "Construction of a bacterial artificial chromosome (BAC) library for potato molecular cytogenetics research," *Genome*, vol. 43, no. 1, pp. 199–204, 2000.

[20] Q. Chen, S. Sun, Q. Ye, S. McCuine, E. Huff, and H.-B. Zhang, "Construction of two BAC libraries from the wild Mexican diploid potato, *Solanum pinnatisectum*, and the identification of clones near the late blight and Colorado potato beetle resistance loci," *Theoretical and Applied Genetics*, vol. 108, no. 6, pp. 1002–1009, 2004.

[21] H. Kuang, F. Wei, M. R. Marano, et al., "The R1 resistance gene cluster contains three groups of independently evolving, type I R1 homologues and shows substantial structural variation among haplotypes of *Solanum* demissum," *The Plant Journal*, vol. 44, no. 1, pp. 37–51, 2005.

[22] F. Dong, J. Song, S. K. Naess, J. P. Helgeson, C. Gebhardt, and J. Jiang, "Development and applications of a set of chromosome-specific cytogenetic DNA markers in potato," *Theoretical and Applied Genetics*, vol. 101, no. 7, pp. 1001–1007, 2000.

[23] A. Bendahmane, K. Kanyuka, and D. C. Baulcombe, "The Rx gene from potato controls separate virus resistance and cell death responses," *Plant Cell*, vol. 11, no. 5, pp. 781–791, 1999.

[24] E. A. G. van der Vossen, J. N. A. M. R. van der Voort, K. Kanyuka, et al., "Homologues of a single resistance-gene cluster in potato confer resistance to distinct pathogens: a virus and a nematode," *The Plant Journal*, vol. 23, no. 5, pp. 567–576, 2000.

[25] S. Huang, V. G. A. A. Vleeshouwers, J. S. Werij, et al., "The R3 resistance to Phytophthora infestans in potato is conferred by two closely linked R genes with distinct specificities," *Molecular Plant-Microbe Interactions*, vol. 17, no. 4, pp. 428–435, 2004.

[26] S. Huang, E. A. G. van der Vossen, H. Kuang, et al., "Comparative genomics enabled the isolation of the R3a late blight resistance gene in potato," *The Plant Journal*, vol. 42, no. 2, pp. 251–261, 2005.

[27] T.-H. Park, J. Gros, A. Sikkema, et al., "The late blight resistance locus Rpi-blb3 from *Solanum bulbocastanum* belongs to a major late blight R gene cluster on chromosome 4 of potato," *Molecular Plant-Microbe Interactions*, vol. 18, no. 7, pp. 722–729, 2005.

[28] T. H. Park, V. G. A. A. Vleeshouwers, D. J. Huigen, E. A.G. van der Vossen, H. J. van Eck, and R. G. F. Visser, "Characterization and high-resolution mapping of a late blight resistance locus similar to R2 in potato," *Theoretical and Applied Genetics*, vol. 111, no. 3, pp. 591–597, 2005.

[29] I. Hein, K. McLean, B. Chalhoub, and G. J. Bryan, "Generation and screening of a BAC library from a diploid potato clone to unravel durable late blight resistance on linkage group IV," *International Journal of Plant Genomics*, vol. 2007, Article ID 51421, 5 pages, 2007.

[30] S. Pflieger, V. Lefebvre, and M. Causse, "The candidate gene approach in plant genetics: a review," *Molecular Breeding*, vol. 7, no. 4, pp. 275–291, 2001.

[31] J. Paal, H. Henselewski, J. Muth, et al., "Molecular cloning of the potato *Gro1-4* gene conferring resistance to pathotype Ro1 of the root cyst nematode *Globodera rostochiensis*, based on a candidate gene approach," *The Plant Journal*, vol. 38, no. 2, pp. 285–297, 2004.

[32] C. S. Jung, H. M. Griffiths, D. M. De Jong, S. Cheng, M. Bodis, and W. S. De Jong, "The potato *P* locus codes for flavonoid $3', 5'$-hydroxylase," *Theoretical and Applied Genetics*, vol. 110, no. 2, pp. 269–275, 2005.

[33] L. Li, J. Strahwald, H.-R. Hofferbert, et al., "DNA variation at the invertase locus invGE/GF is associated with tuber quality traits in populations of potato breeding clones," *Genetics*, vol. 170, no. 2, pp. 813–821, 2005.

[34] C. W. B. Bachem, R. S. van der Hoeven, S. M. de Bruijn, D. Vreugdenhil, M. Zabeau, and R. G. F. Visser, "Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP: analysis of gene expression during potato tuber development," *The Plant Journal*, vol. 9, no. 5, pp. 745–753, 1996.

[35] C. Bachem, R. S. van der Hoeven, J. Lucker, et al., "Functional genomic analysis of potato tuber life-cycle," *Potato Research*, vol. 43, no. 4, pp. 297–312, 2000.

[36] B. Brugmans, A. F. del Carmen, C. W. B. Bachem, H. van Os, H. J. van Eck, and R. G. F. Visser, "A novel method for the construction of genome wide transcriptome maps," *The Plant Journal*, vol. 31, no. 2, pp. 211–222, 2002.

[37] L. M. Trindade, B. M. Horvath, R. van Berloo, and R. G. F. Visser, "Analysis of genes differentially expressed during potato tuber life cycle and isolation of their promoter regions," *Plant Science*, vol. 166, no. 2, pp. 423–433, 2004.

[38] E. Ritter, J. I. Ruiz de Galarreta, H. J. van Eck, and I. Sánchez, "Construction of a potato transcriptome map based on the cDNA-AFLP technique," *Theoretical and Applied Genetics*, vol. 116, no. 7, pp. 1003–1013, 2008.

[39] V. E. Velculescu, L. Zhang, B. Vogelstein, and K. W. Kinzler, "Serial analysis of gene expression," *Science*, vol. 270, no. 5235, pp. 484–487, 1995.

[40] S. Saha, A. B. Sparks, C. Rago, et al., "Using the transcriptome to annotate the genome," *Nature Biotechnology*, vol. 20, no. 5, pp. 508–512, 2002.

[41] K. L. Nielsen, K. Grønkjær, K. G. Welinder, and J. Emmersen, "Global transcript profiling of potato tuber using LongSAGE," *Plant Biotechnology Journal*, vol. 3, no. 2, pp. 175–185, 2005.

[42] S. J. Emrich, W. B. Barbazuk, L. Li, and P. S. Schnable, "Gene discovery and annotation using LCM-454 transcriptome sequencing," *Genome Research*, vol. 17, no. 1, pp. 69–73, 2007.

[43] W. Rensink, A. Hart, J. Liu, S. Ouyang, V. Zismann, and C. R. Buell, "Analyzing the potato abiotic stress transcriptome using expressed sequence tags," *Genome*, vol. 48, no. 4, pp. 598–605, 2005.

[44] B. Kloosterman, O. Vorst, R. D. Hall, R. G. F. Visser, and C. W. Bachem, "Tuber on a chip: differential gene expression during potato tuber development," *Plant Biotechnology Journal*, vol. 3, no. 5, pp. 505–519, 2005.

[45] B. Kloosterman, D. De Koeyer, R. Griffiths, et al., "The potato transcriptome: a new look at transcriptional changes during tuber development using the POCI array," *Functional & Integrative Genomics*, vol. 8, pp. 329–340, 2008.

[46] J. Muth, S. Hartje, R. M. Twyman, H.-R. Hofferbert, E. Tacke, and D. Prüfer, "Precision breeding for novel starch variants in potato," *Plant Biotechnology Journal*, vol. 6, no. 6, pp. 576–584, 2008.

[47] I. Simko, K. G. Haynes, and R. W. Jones, "Assessment of linkage disequilibrium in potato genome with single nucleotide polymorphism markers," *Genetics*, vol. 173, no. 4, pp. 2237–2245, 2006.

[48] R. Lu, A. M. Martín-Hernández, J. R. Peart, I. Malcuit, and D. C. Baulcombe, "Virus-induced gene silencing in plants," *Methods*, vol. 30, no. 4, pp. 296–303, 2003.

[49] O. Faivre-Rampant, E. M. Gilroy, K. Hrubikova, et al., "Potato virus X-induced gene silencing in leaves and tubers of potato," *Plant Physiology*, vol. 134, no. 4, pp. 1308–1316, 2004.

[50] G. Brigneti, A. M. Martín-Hernández, H. Jin, et al., "Virus-induced gene silencing in *Solanum* species," *The Plant Journal*, vol. 39, no. 2, pp. 264–272, 2004.

[51] L. Jones, A. J. Hamilton, O. Voinnet, C. L. Thomas, A. J. Maule, and D. C. Baulcombe, "RNA-DNA interactions and DNA methylation in post-transcriptional gene silencing," *Plant Cell*, vol. 11, no. 12, pp. 2291–2302, 1999.

[52] R. Lu, I. Malcuit, P. Moffett, et al., "High throughput virus-induced gene silencing implicates heat shock protein 90 in plant disease resistance," *The EMBO Journal*, vol. 22, no. 21, pp. 5690–5699, 2003.

[53] C.-M. Ryu, A. Anand, L. Kang, and K. S. Mysore, "Agrodrench: a novel and effective agroinoculation method for virus-induced gene silencing in roots and diverse Solanaceous species," *The Plant Journal*, vol. 40, no. 2, pp. 322–331, 2004.

[54] M. M. Goodin, D. Zaitlin, R. A. Naidu, and S. A. Lommel, "*Nicotiana benthamiana*: its history and future as a model for plant-pathogen interactions," *Molecular Plant-Microbe Interactions*, vol. 21, no. 8, pp. 1015–1026, 2008.

[55] M. R. Godge, A. Purkayastha, I. Dasgupta, and P. P. Kumar, "Virus-induced gene silencing for functional analysis of selected genes," *Plant Cell Reports*, vol. 27, no. 2, pp. 209–219, 2008.

[56] E. M. Gilroy, I. Hein, R. van der Hoorn, et al., "Involvement of cathepsin B in the plant disease resistance hypersensitive response," *The Plant Journal*, vol. 52, no. 1, pp. 1–13, 2007.

[57] S. H. E. J. Gabriëls, F. L. W. Takken, J. H. Vossen, et al., "cDNA-AFLP combined with functional analysis reveals novel genes involved in the hypersensitive response," *Molecular Plant-Microbe Interactions*, vol. 19, no. 6, pp. 567–576, 2006.

[58] S. H. E. J. Gabriëls, J. H. Vossen, S. K. Ekengren, et al., "An NB-LRR protein required for HR signalling mediated by both extra- and intracellular resistance proteins," *The Plant Journal*, vol. 50, no. 1, pp. 14–28, 2007.

[59] Y. Dong, T. M. Burch-Smith, Y. Liu, P. Mamillapalli, and S. P. Dinesh-Kumar, "A ligation-independent cloning tobacco rattle virus vector for high-throughput virus-induced gene silencing identifies roles for *NbMADS4-1* and -2 in floral development," *Plant Physiology*, vol. 145, no. 4, pp. 1161–1170, 2007.

[60] S. Kamoun and S. B. Goodwin, "Fungal and oomycete genes galore," *New Phytologist*, vol. 174, no. 4, pp. 713–717, 2007.

[61] M. R. Armstrong, S. C. Whisson, L. Pritchard, et al., "An ancestral oomycete locus contains late blight avirulence gene Avr3a, encoding a protein that is recognized in the host cytoplasm," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 21, pp. 7766–7771, 2005.

[62] A. P. Rehmany, A. Gordon, L. E. Rose, et al., "Differential recognition of highly divergent downy mildew avirulence gene alleles by *RPP1* resistance genes from two Arabidopsis lines," *Plant Cell*, vol. 17, no. 6, pp. 1839–1850, 2005.

[63] S. C. Whisson, P. C. Boevink, L. Moleleki, et al., "A translocation signal for delivery of oomycete effector proteins into host plant cells," *Nature*, vol. 450, no. 7166, pp. 115–118, 2007.

[64] T. A. Torto, S. Li, A. Styer, et al., "EST mining and functional expression assays identify extracellular effector proteins from the plant pathogen Phytophthora," *Genome Research*, vol. 13, no. 7, pp. 1675–1685, 2003.

[65] V. G. A. A. Vleeshouwers, J.-D. Driesprong, L. G. Kamphuis, et al., "Agroinfection-based high-throughput screening reveals specific recognition of INF elicitins in *Solanum*," *Molecular Plant Pathology*, vol. 7, no. 6, pp. 499–510, 2006.

[66] N. Champouret, H. Rietman, J. I. Bos, et al., "Functional allele mining: a new approach to identify R-gene homologues in *Solanum*," in *Proceedings of the 13th International Congress on Molecular Plant-Microbe Interactions*, p. 177, Sorrento, Italy, July 2007.

[67] I. Hein, J. Squires, P. Birch, and G. J. Bryan, "Screening wild potato accessions for resistance to the virulent allele of the *Phytophthora infestans* effector Avr3a," in *Proceedings of the 13th International Congress on Molecular Plant-Microbe Interactions*, p. 259, Sorrento, Italy, July 2007.

Hindawi Publishing Corporation
International Journal of Plant Genomics
Volume 2008, Article ID 528361, 17 pages
doi:10.1155/2008/528361

Review Article

Citrus Genomics

Manuel Talon[1] and Fred G. Gmitter Jr.[2]

[1] *Centro de Genómica, Instituto Valenciano de Investigaciones Agrarias (IVIA), 46113 Moncada, Valencia, Spain*
[2] *Citrus Research and Education Center (CREC), University of Florida, IFAS, Lake Alfred, FL 33850, USA*

Correspondence should be addressed to Fred G. Gmitter Jr., fgg@crec.ifas.ufl.edu

Received 9 October 2007; Accepted 15 March 2008

Recommended by Yunbi Xu

Citrus is one of the most widespread fruit crops globally, with great economic and health value. It is among the most difficult plants to improve through traditional breeding approaches. Currently, there is risk of devastation by diseases threatening to limit production and future availability to the human population. As technologies rapidly advance in genomic science, they are quickly adapted to address the biological challenges of the citrus plant system and the world's industries. The historical developments of linkage mapping, markers and breeding, EST projects, physical mapping, an international citrus genome sequencing project, and critical functional analysis are described. Despite the challenges of working with citrus, there has been substantial progress. Citrus researchers engaged in international collaborations provide optimism about future productivity and contributions to the benefit of citrus industries worldwide and to the human population who can rely on future widespread availability of this health-promoting and aesthetically pleasing fruit crop.

1. INTRODUCTION

Citrus is one of the most important and widely grown of the fruit crops, with total global production reported to be 105.4 million tons in 2004-2005 [1]. Citrus fruit is produced throughout the tropical and subtropical regions of the world, where the winter temperatures are adequate for tree survival and avoidance of freeze devastation, and where there is sufficient water and suitable soils to support tree growth and fruit production. The most significant production areas are found in the Americas (led by Brazil, the United States, Mexico, and Argentina), the Mediterranean basin (led by Spain, Italy, Egypt, and Turkey), and the south and east Asian regions (led by China, India, and Japan). Citrus production, whether for processed or fresh fruit products, from the largest producing countries is an important commodity for global trade and of tremendous economic value and impact. However, there is much citrus production of great importance to local national and regional economies and of value to the nutritional needs of people in less developed nations; this is born out, for example, by the fact that sweet oranges (*Citrus sinensis* L. Osb.) are reported to be grown in 114 countries, grapefruit (*Citrus paradisi* Macf.) and pummelos (*Citrus maxima* Merr.) in 74 countries, and lemons/limes (*Citrus limon* [L.] Burm. F./*Citrus aurantifolia* [Christm] Swing.) in 94 different countries [2].

In addition to the use as a food or beverage source, citrus products from some of the wild species not grown commercially are also of value as agents of traditional medicinal and sanitary utilization [3]. Several closely related genera have varying degrees of sexual compatibility with *Citrus*, some of which produce edible fruit for commerce (e.g., the kumquats, *Fortunella* [Swing.]) and others that possess traits of economic value for rootstock and scion improvement (e.g., the trifoliate orange or *Poncirus trifoliata* [L.] Raf.). Despite the diversity of fruit types, however, nearly 70% of the world's citrus production is sweet orange.

Given the tremendous extent and value of citrus production, it may be somewhat surprising on first consideration that nearly all of the major scion and rootstock cultivars utilized in much of the world have not arisen as a consequence of systematic and targeted breeding programs. Rather, they have arisen spontaneously as seedling and/or bud sport mutations or by introduction and trials of materials from one location to another [4–6]. The reasons for the low level impact of traditional breeding approaches to genetic improvement of this major fruit crop are related to the peculiarities of citrus reproductive biology and the fairly unique aspects of the taxonomic relationships of the major cultivar groups [7]. Citrus seedlings are subject to juvenile periods ranging from one to as many as 20 years, though typically they will flower and fruit within 3–7 years, depending on species. Even after first flowering, it is common for fruit traits to be atypical of later characteristics as scion lines mature. One consequence of juvenility is the obvious delay between hybridization and selection for desired characteristics; however, a secondary consequence is the requirement of large unit areas of land to grow substantially large individual hybrids, thereby increasing the cost of maintenance in the field and limiting the number of families and individuals within families that can be grown. Further, many of the commercial citrus types produce polyembryonic seeds through nucellar embryony, yielding seedlings that are essentially clones of the maternal parent. These embryos arise autonomously prior to anthesis and their development to maturity follows normal pollination and endosperm development [8]. These nucellar embryos most frequently grow much more vigorously than any zygotic embryos and, consequently, the frequency of true zygotics is extremely low. Finally, it is important to recognize that several of the so-called "species" of economic significance (e.g., *C. sinensis*, *C. paradisi*, and *C. limon*) are not biologically defined species; the cultivars in these groups represent accumulated somatic mutations identified over centuries through on-tree or nucellar seedling mutations [9]. Further, some cultivar groups within other species, such as the Clementine and Satsuma mandarins, are likewise the result of somatic mutations and not a consequence of hybridization. Market and consumer expectations and demands for specific commodities (e.g., sweet oranges, grapefruit, lemons, Clementines, and Satsumas) thereby limit the possibilities for genetic improvement within these cultivar groups because the commodities must meet the consumers' expectations and concepts related to fruit traits. These needs and the narrow germplasm bases actually represented within these cultivar groups, along with the reproductive factors, have precluded breeding as a strategy for cultivar development and improvement. The exceptions to this are pummelos (*C. maxima*), the development of new types of mandarin hybrids (using selections that produce monoembryonic seeds containing true zygotic embryos), and rootstock breeding, where hybridization and selection are viable and productive approaches. In these cases, however, limited genetic understanding of the inheritance and control of critical traits remains a substantial issue.

With globalization of citrus production and increased human travel throughout the world, particularly devastating citrus diseases have been rapidly spreading, thus threatening the viability and the very future of citrus production globally. This is the case with much of agriculture; increasing human populations and urban development have forced citrus production to less desirable regions where environmental factors present greater challenges to sustained production. Increased demands for water resources that follow increased populations and urbanization likewise limit the availability of water resources with adequate quality to maintain tree growth and production. There exist genetic resources to address most of these challenges. However, the genetic challenges and the lack of understanding of the fundamental mechanisms underlying these critical traits, as described above, present tremendous impediments to the progress needed to incorporate needed genes and alleles and to devise the appropriate strategies for the continued production and economically feasible availability of citrus fruits to the future world population. In this context, the advent of genomic science and the powerful new tools that are being developed and utilized for citrus improvement take on critical significance. This article will review the progress that has taken place thus far in the development and application of genomic information for citrus improvement and present the current status of research and future directions envisioned.

2. LINKAGE MAPPING

Citrus and the closely related genera are partially sexually compatible in varying degrees; they are primarily diploid with a few known triploids and occasional tetraploid forms ($2n = 2x = 18$), and they possess fairly small genomes (e.g., sweet orange has been said to be around 367 Mb, or approximately three times that of *Arabidopsis* [10]). As such, the citrus species should be amenable to many of the commonly used techniques and approaches related to genomic research, including genetic and physical mapping, full genome sequencing, and functional genomics studies aimed at unraveling the complexities of key traits of interest. Because of the valuable characteristics within some of the related genera that are absent from *Citrus*, particularly cold tolerance and resistance to citrus canker (caused by *Xanthomonas axonopodis* pv. *citri*) from kumquat (*Fortunella*) and multiple stress-tolerant and disease-resistance traits from *P. trifoliata* (including freeze avoidance, and resistance to citrus tristeza virus (CTV), *Phytophthora*, and citrus nematode [*Tylenchulus semipenetrans*]), many of the genetic mapping projects and some of the physical mapping as well have focused on *P. trifoliata* through intergeneric hybrids with *Citrus*.

Citrus breeders and geneticists have long desired to have linkage maps that empower selection schemes based on easily scored, neutral molecular markers rather than relying on frequently difficult, time-consuming, and inefficient approaches based on phenotypic characterizations. Indeed, genetic linkage maps have been produced across the past two decades with increasing value and resolution, as the evolution of new marker systems has taken place. The first published report of linkage mapping in citrus (using a small intragenic family of *Citrus*, and a larger *Citrus* × *Poncirus* family) was based on leaf isozymes [11]; five markers were found that defined two linkage groups, and significantly this was also the first report of linkage distortion which has been a common feature of the many intergeneric mapping efforts that followed subsequently. As RFLP technology became commonly applied in genetic studies, citrus scientists began to incorporate it together with isozyme methods, and new maps resulted first with

35 markers in 8 linkage groups covering 314 cM within a citrus backcross family [12], followed by maps with 52 and 35 markers each, defining 11 and 10 linkage groups and 533 and 351 cM, in an intergeneric backcross family and a population derived from crossing two individual intergeneric F1 hybrids, respectively [13, 14]. As new marker systems were developed, the maps produced from each of these families were further populated first by RAPDs [15], thereby increasing the number of markers from Durham's map from 52 to 189, decreasing the number of linkage groups to 9, and more than doubling genomic coverage to 1192 cM. Sankar and Moore [16] increased marker coverage in the same map to 310 markers through use of ISSRs. In similar fashion, Jarrell's map was improved through incorporation of SSR loci [17], and then further by ISSR marker development to 156 markers defining 16 linkage groups across 701 cM [18]. AFLP markers were first reported to be used for citrus genetic mapping in 1998 by de Simone et al. [19] and in 1999 by Ling et al. [20] (also elaborating the Durham map). Many other whole genome maps have been produced as well as trait specific maps identifying single gene and QTL regions of significance; these have been summarized by Chen et al. [21]. It is through the latter category of trait specific mapping that some of the promise of genomic science for citrus genetic improvement is being pursued and realized, including selection of disease resistant and environmental stress tolerant hybrids in rootstock breeding programs, and targeted gene cloning projects aimed at providing potential solutions to serious disease problems.

Although higher throughput and increased marker density became possible through application of RAPD, AFLP, and ISSR techniques, these systems were of limited value in comparative genomic studies and in utilization for marker-assisted selection (MAS) methods because of the dominant nature of the markers and their low portability among populations. RFLPs, SCARs, CAPS, SSRs, and SNPs are obviously much more desirable for broad applications, and the citrus research community has been developing these resources over time. SCAR markers for citrus were developed first by Deng et al. [22], based on RAPD markers that were closely linked to *Ctv*, a gene for CTV resistance from *P. trifoliata* [23]. Polymorphism could not be revealed by some of these SCAR markers without restriction of amplified products, thus they were converted to CAPS. In 1999, García et al. [24] likewise used CAPS together with RFLPs, RAPDs, and isozymes to map genes in *Citrus* and *Poncirus* associated with apomixis. Though codominant marker types such as RFLP, SSR, SCAR, CAPS, and SNPs were being used earlier, the numbers of such loci available were severely limited. Consequently, there has been a limited ability to interrelate maps and/or markers developed in different populations or targeting different QTLs, and there were very few chromosome-specific anchor markers that could enable comparative mapping efforts between different genetic resources of citrus. The processes for developing such markers previously were very time and labor intensive, and the efficiencies were extremely low. Further, some of these markers were developed in the absence of genome or transcriptome sequences, and as such they might be considered to be gene anonymous. However, the revolution in sequencing technologies, including the sequencing of BAC clones and fairly extensive EST libraries for citrus accessions under multiple conditions, has produced a very substantial resource for high-throughput development, verification, and utilization of molecular markers for citrus linkage mapping, which meets the desired criteria. These markers frequently are based on EST sequences and therefore represent specific genes, which functions may be known or estimated in some cases. For example, Omura et al. [25, 26] first reported development of 131 mapped CAPS markers derived from EST sequences. Using a backcross citrus population, these markers were assigned into nine linkage groups that accounted for 685 cM coverage, and they were found to be portable to another population. With the rapid increase in publicly available EST databases in the USA [27] and publicly available software programs, large-scale searches for various types of SSR motifs and efficient design of appropriate primers have made it possible to identify and map EST-SSRs in citrus [21, 28]. The first such map for sweet orange and *P. trifoliata* was published in 2007 [21], and it is being expanded collaboratively [29]. New international, collaborative EST-SSR mapping efforts are currently underway [30] using other citrus-based families as part of a plan intended to lead to the full-length sequence of a haploid citrus genome, to be integrated with physical and genetic maps based on BAC end sequencing, high-throughput marker saturation, and mapping traits of economic importance to genetic improvement of citrus. These extensive international efforts are being promoted and coordinated by the International Citrus Genome Consortium (ICGC), currently chaired by F. G. Gmitter of the University of Florida, USA. New technology will continue to enhance progress toward high-resolution and highly informative maps of citrus genomes in the future. Currently, efforts are underway in several labs around the world to utilize microarrays for mapping SNPs in various families and genetic backgrounds, and, as a full genome sequence comes forward for citrus, followed by additional resequencing of other genomes of interest, the genomic identification and locations of thousands of trait-relevant SNPs will become known and exploited for genetic improvement of the crop.

3. PHYSICAL MAPPING

The main challenge for a comprehensive and meaningful description of the genomes is the integration of the DNA marker-based genetic maps with physical maps, and eventually with DNA sequence of the whole genome, the ultimate physical map. Large genomic DNA insert-containing libraries are required for physical mapping, positional cloning, and genome sequencing of complex genomes. The physical mapping of complex genomes is based on the construction of a genomic library, and the determination of the overlaps between the inserts of the mapping clones in order to generate an ordered, cloned representation of nearly all the sequences present in the target genome.

For the generation of high-resolution physical maps, the construction of bacterial artificial chromosome (BAC) libraries containing clones with large DNA fragments appears to be indispensable. The BAC cloning system has become a dominant system over others to clone large genomic DNA inserts. BAC clone collections and BAC-based contig maps are indeed powerful tools having multiple applications in genomics such as supporting positional cloning or to aid large-scale assembly of whole genomes. In whole genome sequencing projects, BAC end sequences (BES, paired-end reads) are also of inestimable help for the integration of the physical map with the genome sequence. Furthermore, in many agricultural important species, BAC clones and physical maps are being rapidly developed since they are essential components in linking phenotypic traits to the responsible genetic variation, to integrate the genetic data, for the comparative analysis of genomes, and to speed up and improve potential and effectiveness of marker-assisted selection (MAS) for breeding.

3.1. BAC libraries

In citrus, Yang et al. [31] and Deng et al. [32] independently constructed two BAC libraries as part of a map-based, or positional, cloning strategy with the idea of identifying BAC clones spanning the genetic region identified as containing gene(s) for resistance to CTV. CTV is the causal agent of several diseases causing significant economic damage and losses to citrus worldwide. Broad spectrum resistance to CTV was previously associated with a single dominant gene, *Ctv*, characterized in *P. trifoliata*, a sexually compatible relative of citrus [23]. In order to clone this gene, Yang et al. [31] constructed a BAC library from an individual plant homozygous for *Ctv*. The library contained 45 000 clones with an average insert size of 80 kb. *Ctv* was initially mapped to a 282-kb region including a disease resistance gene cluster with seven members and eight retrotransposons clustered [33]. Sequence analysis of the *Ctv* surrounding genomic region located the locus into a 121-kb *Poncirus* region comprising 10 genes. All 10 genes were individually cloned in *Agrobacterium*-based binary vector and used to transform susceptible varieties [34] to test their resistance capability.

In a parallel effort, a BAC library was constructed from the genomic DNA of an intergeneric *Citrus* and *Poncirus* hybrid for molecular isolation of disease resistance genes, including *Ctv* [32]. The library of 24 000 clones with an average insert size of 115 kb was screened with DNA markers linked to the *Ctv* gene and citrus disease resistance gene candidate (RGC) sequences. A few clones were isolated with each of the CTV resistance gene-linked markers, and several hundred others were identified using previously cloned citrus RGC sequences as probes [35]. Further fingerprinting and assembly resulted in the identification of 25 contigs of 120–250 kb. Additional libraries were developed from the same intergeneric hybrid for the purpose of map-based cloning of *Ctv*. From these libraries, full contigs were constructed that spanned both the resistance allele from *Poncirus* and the susceptibility allele from the *Citrus* chromosome. These clones were fully sequenced and assembled. Comparisons of the resistance and susceptibility allelic genomic sequences revealed that the levels of similarity varied from region to region. Within the region where the most likely *Ctv* candidate genes were delimited, based on fine genetic mapping and predicted by various sequence analysis programs, there were 2 NBS-LRR candidate genes found in *Ctv* that were completely missing from the *Ctv* sequence. Based on the sequence analysis and the fine mapping results, it was concluded that either one or both of these unique sequences should be considered the first priority candidates for *Ctv*.

A further exploitation of this BAC library resulted in identification of other disease resistance gene-like DNA sequences, using a PCR approach with degenerate primers designed from conserved NBS (nucleotide-binding site) motifs [35]. In addition to the three amplified DNA fragment markers associated with the citrus tristeza virus resistance gene (*Ctv*), another fragment (Pt8a) was found to be associated with the major gene responsible for the citrus nematode resistance (*Tyr1*). In a similar approach, degenerate primers for the conserved motifs in the kinase domains of the plant disease resistant genes (R) of rice *Xa21* and tomato *Pto* were used in PCR amplification to identify resistance gene candidates. Twenty-nine sequences highly similar to the kinase domain of *Xa21* were cloned and characterized [36]). Using the BAC library, two full-length sequences, including upstream promoters and downstream terminating sequences, were identified. Markers derived from these *Xa21*-like sequences have been found linked to putative QTLs for citrus canker resistance segregating among hybrids derived from *Citrus ichangensis* Swing. with *Citrus limettoides* Tan. (Gmitter, unpublished data).

3.2. Physical mapping

Two communications at the PAG XV meeting in San Diego, 2007, reported progress on the construction of physical maps of citrus [37, 38]. The Spanish Citrus Genomic Consortium has constructed three BAC libraries from Clementine mandarin (EcoR I, Hind III, and MboI) containing a total of 57 000 clones with an average insert size of 120 kb (19x coverage). Half of these BAC clones were end-sequenced (29 Mb), and these sequences analyzed [37]. The sequence analysis revealed that most abundant known retroelements were LTR elements, especially Ty-1 Copia [39] and Gypsy [40] elements, while known DNA transposons were scarce. Basic local alignment search tool (BLAST) searches also identified about 14 000 clones with coding regions in a least one end and therefore putative euchromatin regions. BAC end sequences were also searched for single nucleotide polymorphisms (SNPs) and simple sequence repeats (SSRs) in the Clementine genome. These initial analyses identified more than 2800 sequence repeats in coding regions out of 7700 putative SSRs. Some 1.7% of the reads had high similarity with the sequence of the *Citrus sinensis* chloroplast genome [41], suggesting that about 70% of the chloroplast genome of Clementine was similarly recovered in this sequencing effort. In parallel, a physical map derived from the same 28 000-clone set of the Clementine BAC libraries is being constructed by restriction enzyme fragment fingerprinting, and work is in progress to place on it a sufficient number of genetically mapped markers to anchor and orientate the contigs (Ollitrault, personal communication). It is expected that the paired-end reads will also aid integration of the genetic and physical maps.

The Citrus Genome Analysis Team from Japan has also communicated the construction of a physical map of citrus by high-information-content fingerprinting (HICF) analysis of a BAC library from Satsuma mandarin (*Citrus unshiu* Marc.) consisting of 37 000 clones, with 13.3x of citrus genome [38]. More than 6000 BAC clones from the library were fluorescent labeled using the SNaPshot kit after digestion with BamHI, EcoRI, XbaI, and XhoI and assembly of BAC fingerprints by FPC resulted in approximately 1000 contigs (1.6x coverage). Consistent assembly among contigs obtained by fingerprint analysis and physical maps obtained by BAC walking with both molecular markers and BAC end sequences was observed. Further evaluations by additional clones, assignment of molecular markers for the contigs, and gap filling by BAC end sequencing to complete the physical map were also reported to be in progress [38].

A BAC library of Ridge Pineapple sweet orange was produced by Michael Bausher (USDA-ARS, Ft. Pierce, FL, USA) containing 18 432 clones (BamHI/Mbo I) with an average insert size of 145 kb, or an estimated 7x coverage. A total of 16 727 clones from this library have been fingerprinted and assembled into 472 contigs, as of August 2006. Access is freely available to the public at http://phymap.ucdavis.edu:8080/citrus. This resource was searched by EST-SSR overgo probes to identify BAC clones in six known heterozygous genomic regions containing polymorphic alleles of mandarin and/or pummelo ancestry (the putative ancestral genomic contributors to sweet orange). These BAC clones were then sequenced and assembled "blind" to assess the difficulties in assembling sequences of the heterozygous sweet orange.

The goal of genome-wide integrated physical and genetic maps is a priority in citrus genomics since it will provide the essential and powerful tools for research into the citrus genome, such as effective positional cloning, marker development, high-throughput EST mapping, and large-scale genome sequencing and assembly.

4. CITRUS SEQUENCING

One of the major goals of the International Citrus Genomics Consortium is to provide a high quality sequence of a citrus genome. Sequencing of the citrus genome will facilitate the comparison of herbaceous and woody perennial genomes and provide a valuable resource for studying significant biological questions of critical importance to genetic improvement of citrus. From a scientific perspective, citrus as a fruit tree developing nonclimacteric fruits possesses a combination of interesting biological characteristics such as apomixis, gametophytic self- and cross-incompatibility, juvenility, deciduousness/evergreen foliage, dormancy, seasonality, root/shoot interaction, oil glands, nutraceutical compounds, and plant-pathogen interactions. Additionally, citrus is the most economically significant fruit crop produced in the world, although citrus production is severely threatened by pest, disease, and environmental problems to which current commercial rootstock and scion cultivars are susceptible.

In February 2004, a proposal prepared and supported by the National Citrus Genomics Steering Committee (USA) and the International Citrus Genomics Consortium (ICGC, composed of researchers from Australia, Brazil, China, France, Israel, Italy, Japan, Spain, and USA) to sequence the genome of sweet orange was presented to the Joint Genome Institute (JGI). This institution reported at the beginning of 2007 to have produced a low coverage (ca. 1.2x) whole-genome shotgun sequence of *Citrus sinensis* (sweet orange) by sequencing ends of about 126 000 fosmid clones containing 40 kb inserts and 257 000 plasmid clones containing 8 kb inserts [42]. Total sequence coverage, available at http://harvest.ucr.edu/, was about 473 Mb, coverage apparently insufficient to provide quality assembling of the high heterozygous sweet orange.

In January 2007, the Steering Committee of the ICGC met at JGI in California, USA to reassess the present status of citrus genome research and to forge plans for future collaborative efforts. The first outcome of the meeting was a decision to shift focus to a haploid genome as the target for sequencing, rather than the previously stated target of sweet orange; this consideration was made to eliminate the difficulties associated with quality assembly of highly heterozygous diploid sweet orange genome. A haploid derived genome sequence should serve as the highest quality reference genome for all future genomic research efforts. It was required that the haploid (or di- or tri-haploid individual chosen) should be available for free distribution internationally, and that it should be pathogen-free, exhibit robust vegetative growth (as a partial guarantee against gross genome defects), and be relatively easy to maintain. Teams were established to verify chromosome number and to assess candidates for homozygosity using SSR markers representing good genome coverage based on linkage maps. Additionally, the candidates will be assessed using at least two available citrus microarray platforms in an effort to insure that there are no large deletions or other cytogenetic defects. Currently there are three candidates, all derived from Clementine mandarin, that are being evaluated according to this plan. It is envisaged that the international collaboration will be supported from various agencies and sources in different partner nations, that the sequence information will be quickly deposited and shared freely among the participating laboratories, and that the goal will be achieved once there is 8 to 10x coverage. Currently, there are funding commitments from USA, Spain, France, Italy, and China.

The complete chloroplast genome sequence of *Citrus sinensis* was recently provided by Bausher et al. [41]. It is 160,129 bp in length and contains 133 genes (89 protein-coding, 4 rRNAs, and 30 distinct tRNAs). The genome included 29 direct and inverted repeats 30 bp or longer, and comparison of protein-coding sequences with expressed sequence tags revealed six putative RNA edits. Phylogenetic analyses provide strong support for the monophyly of eurosids II and for the placement of Citrus (Sapindales) sister to a clade including the Malvales/Brassicales.

5. FUNCTIONAL GENOMICS

5.1. EST sequencing

The first sets of ESTs (expressed sequence tags) from any citrus material came from the pioneering work of Omura and coworkers who reported about 3000 partial sequences of cDNA clones from libraries derived from seeds, and from developing and mature fruit and albedo tissue, during the second half of the 1990s [43]. Later, a set of 6500 ESTs derived from whole seedlings of sweet orange was developed by Bausher et al. [44], and a new contribution of 600 sequences from *Citrus unshiu* of the Japanese team was also reported [45]. Since then, various groups (including Roose and Close at University of California at Riverside [UCR], Dandekar at University of California at Davis [UCD], and the Spanish Citrus Genomics Consortium at Valencia) have contributed to EST sequencing efforts using several species, mostly *C. sinensis* (sweet orange), *C. clementina* (Clementine mandarin), *C. paradisi* (grapefruit), *Poncirus trifoliata,* and other hybrids (*C. sinensis* × *Poncirus trifoliata*, Carrizo citrange). The total resource has reached 232 808 citrus sequences in the National Center for Biotechnology Information (NCBI) EST database as of May 2007. This EST collection includes a wide representation of sequences from many cDNA libraries derived from multiple reproductive (flowers, ovaries, fruits, seeds) and vegetative (roots, leaves, buds) organs and tissues (pulp flesh, flavedo, abscission zones) at different developmental stages and challenged with biotic (*Phytophthora*, citrus tristeza virus, herbivory, *Penicillium*) and abiotic (salinity, iron deficiency, water deficit) agents, and elicitor and hormonal treatments.

Although a compressive analysis of all ESTs in public databases has not been performed, several subsets of the data have been partially analyzed. Forment et al. [46], for example, generated 25 cDNA libraries covering different conditions and from 22 635 high-quality ESTs identified 11 836 putative unigenes. A third of these unique sequences was reported not to have Arabidopsis orthologues. From a deeper analysis of a collection of 54 000 single-pass ESTs, derived mostly from a normalized full-length cDNA library (41 000 ESTs) and nine additional standard libraries representing particular treatments and tissues from several selected varieties and rootstocks, Terol et al. [47] identified 13 000 putative unigenes with significant BLAST hits. Further analyses and comparisons with *Arabidopsis* suggested the occurrence of citrus paralogues, putative conserved orthologues, single copy genes, duplication events, and increased number of genes for specific pathways. Interestingly, the sequences of the genes belonging to these different species were essentially identical, suggesting that their differential behavior cannot be attributed to major sequence divergences. Nearly 17% (2250 total) of the predicted citrus unigenes had no detectable similarity to *Arabidopsis* genes, and of these, 647 unigenes produced significant hits only to *Citrus* species, suggesting that these clusters might be putative *Citrus* exclusive genes [47]. This work also contributed over 8500 clones carrying putative full-length cDNA sequences. Full-length sequences and clones that are valuable tools since they can facilitate proper prediction of gene structures provide a useful resource for functional analysis and may greatly facilitate annotation of the full genome sequence. BLAST searches against sequenced citrus ESTs are possible through several open database projects (i.e., http://harvest.ucr.edu/; http://cgf.ucdavis.edu;/ http://bioinfo.ibmcp.upv.es/genomics/cfgpDB/) or data deposited in GenBank. Although predictions from EST clustering tend to overestimate the total number of genes, these citrus EST sequences are apparently derived from perhaps 40 to 50 000 genes, a number more similar to that reported in *Populus* than that in *Arabidopsis*.

In addition, Machado et al. [48] reported at the PAG-XV meeting that Brazilian researchers have developed a huge EST sequencing effort. According to this communication, the CitEST (http://citest.centrodecitricultura.br/) Brazilian database including more than 260 000 valid reads contained unigene sets from several citrus species but mainly sweet orange, mandarin, and *Poncirus trifoliata*. These ESTs were generated from several libraries under biotic (*Xylella fastidiosa*, CTV, Citrus Leprosis Virus, *Phytophthora*, mite) and abiotic (drought) stresses, and during fruit development.

5.2. Microarrays platforms

The importance of microarray technology for transcript profiling approaches in functional genomics is increasing exponentially in practically all plant systems and, in particular, in many agricultural crops. In citrus, the first transcript profiling data was reported by Shimada et al. [49] who constructed a cDNA microarray to monitor expression of mRNA from 2213 genes during fruit development. Since then, several citrus DNA microarray platforms were developed. The Spanish Citrus Genomic Consortium developed a first generation cDNA microarray containing 12 672 probes corresponding to 6875 putative unigenes of a 22 000-EST collection [46]. Subsequently, a second-generation microarray comprising 12 000 unigenes was released and, shortly afterwards, the Consortium produced the current version, a higher density citrus microarray composed of 24 000-element cDNA array containing 20 000 unigenes, based on nearly 90 000 high-quality sequences generated from 52 different cDNA libraries.

Similarly, a citrus 22 K oligoarray containing 21 495 independent ESTs from Citrus species has been recently developed in Japan. The information regarding this platform is available at the website http://www.fruit.affrc.go.jp/index-e.html. This tool has already produced very useful information [50]. In 2006, Affymetrix developed and released a citrus GeneChip containing 960 444 total 25-mer oligos in an 11 micron format (http://www.affymetrix.com/analysis/index.affx), this product came through close collaboration with Close and Roose (UCR) and was based on the NCBI citrus EST collection that was available at the time of its design. About two-third of the content was designed for gene expression analysis using 30 264 probe sets. Most of the remaining one-third was designed to genotype 3219 genes using 5023 SNPs identified in ESTs from *Citrus sinensis* and other citrus species. The citrus chip also contained probe sets for detection of several pathogens and commonly used transgenes and a representation of the region of the *P. trifoliata* genome containing *Ctv*, the CTV resistance allele. In January 2007 at PAG, it was communicated that the GeneChip Citrus Genome Array is being used in Israel, for example, to analyze transcription profiles of bud sprouting as related to alternate bearing behavior [51]. Additional work is currently underway in various labs using the Affymetrix product for high-throughput linkage mapping, and to assess gene expression under various physiological and pathogen and/or pest challenges, it is anticipated that there will be several reports published within the next year on these research projects.

Several other communications presented at PAG XV, for instance, Deng et al. [52] and elsewhere reported other research projects using cDNA citrus microarrays or smaller custom arrays based on subtractive libraries. Some of them include analysis of transcriptional responses of 1731 genes to herbivory [53], of 312 subtracted genes in abscission zones (Tadeo and Talon, unpublished data), and the investigation of citrus canker resistance in kumquat (*Fortunella spp*) using an array with 2254 elements [54].

5.3. Gene expression and transcriptome profiling

In a recent review, Jansson and Douglas [55] explained the usefulness of *Populus* as a new plant system model offering new insights in many physiological processes that cannot be easily studied in *Arabidopsis* or rice, the two main models for plant biology. The strength of this proposition, sustained by the completion of the whole poplar genome sequence and the development of several genetics and genomics tools, holds promise to elucidate major tree-specific traits such as wood formation, long-term perennial growth, biotic interactions, and others. The availability of a second tree model can provide contrasting data on plant genome evolution, gene family structure, and other pivotal tree traits. Citrus as a fruit tree not only will promote achievement of these goals but more importantly offers a suitable system to study "fruit growth and quality," a fundamental plant trait for which *Arabidopsis*, rice, and poplar are not useful systems. Although these model plant systems, including tomato, are crucial to understand plant growth and development, the dramatic developmental differences found across species are channeling many efforts to genomic and post-genomic studies of crop plant species, rather than retaining focus solely on these model species.

Citrus possesses an enormous unexplored potential to reveal relevant plant growth processes and some responses that probably cannot be studied in any other plant. Although functional genomics in citrus is currently in its infancy, the particular citrus biology suggests that citrus may contain a reservoir of genes with peculiar and unique functions. One of the first steps to assign functions to unknown genes is the large-scale gene analysis of the transcriptome. In citrus, before microarray availability, gene expression in developmental and environmentally regulated processes, as in many other systems, was mostly studied through differential display techniques (i.e., DDRT-PCR). Genes involved in many processes were also identified after subtractive hybridization of cDNA libraries constructed from two different conditions. The main targets in citrus research have been those physiological processes that sustain major commercial traits. Below, we summarize the knowledge gained in these several areas.

5.3.1. Fruit growth and ripening

While in tomato (a climacteric fruit) great strides have been made in the areas of ethylene regulation, carotenoid accumulation, and cell wall metabolism, in nonclimacteric citrus fruit the general information is substantially less. Mature citrus fruits release low amounts of ethylene but respond to exogenous ethylene by accelerating respiration, chlorophyll degradation, and carotenoid deposition. In these fruits, very low rates of ethylene production have been associated with constitutive expression of the 1-aminocyclopropane-1-carboxylate synthase 2 (CsACS2) and ethylene receptor CsETR1 genes, indicating that citrus possesses a system I machinery. However, it has been reported that a climacteric-like rise in ethylene production, preceded by induction of the genes for CsACS1, ACC oxidase1, and the ethylene receptor CsERS1, characteristic of a system II-like, appears to be present in young fruitlets [56]. It is well known that ethylene accelerates the molecular changes in the carotenoid biosynthesis naturally occurring during maturation (see below), while gibberellins and nitrates, two ripening retardants, reduce expression of early carotenoid biosynthetic genes and repress pheophorbide a oxygenase (PaO) expression [57, 58], a gene involved in chlorophyll disappearance. Other characteristic genes induced by ethylene are ferredoxins or *thi,* a gene involved in thiamine biosynthesis. The large ABA amounts found in the peel of citrus fruit during maturation appear to be synthesized by two 9-cis-epoxycarotenoid dioxygenases (NCED) with differential spatial and environmental expression [59, 60].

In citrus, an initial small-scale EST sequencing project from mature fruit resulted in the identification of 20% of the sequences as encoding for metallothionein [61]. The abundance of these kinds of genes was confirmed later in more developed citrus arrays. Later, Shimada et al. [49] used a citrus cDNA microarray containing 2213 independent genes to examine gene expression during fruit development and reported that the expression profile in the different tissues of the fruit, flesh, albedo, and flavedo was rather different. Recently, a comprehensive transcriptome analysis using a citrus 22 K oligoarray was performed to identify ethylene-responsive genes in mandarin fruit [50]. In the 72 hours after ethylene treatment, 1493 genes were shown to be modulated by the hormone. Ethylene repressed the transcription of most genes involved in photosynthesis, chloroplast biogenesis, and sugar metabolism, while it induced the transcription of several genes related to resistance, defense, stress, amino acid synthesis, protein degradation, and secondary metabolism. The sensitivity and responsive patterns to exogenous ethylene were significantly different among carotenoid biosynthesis genes (see below). Furthermore, most of the ethylene biosynthesis genes and its signal transduction components did not show any significant expression change after ethylene treatment. Interestingly, a type II ethylene receptor (ETR2) showed higher sensitivity to exogenous ethylene than two other type I ethylene receptors (*CsETR1* and *CsERS1*), suggesting that ETR2 might be associated with low ethylene sensitivity in mature fruit [50].

During the last decades, research on citrus fruit flavor that depends upon multiple compounds, mostly sugars, acids, and flavanones, has received considerable attention because of both the uniqueness of the physiological processes sustaining this trait and the potential importance of these components to human health. To date, the most comprehensive study on the transcriptome profiling of the citrus fruit flesh was presented by Cercós et al. [62] who examined gene expression with the first generation Spanish cDNA microarray during development and ripening of self-incompatible *Citrus clementina*. They reported that as many as 2243 putative unigenes showed significant expression changes while functional classification revealed that genes encoding for regulatory proteins were most significantly overrepresented approximately within the middle of the rapid fruit growth phase; this suggested that fruits at this stage were reprogramming developmental commands to face the complex cellular modifications during ripening. Most pivotal changes were related to carbohydrate build up, acid reduction, modifications in secondary metabolism, carotenoid accumulation, and chlorophyll decreases. Alterations of the transcriptome associated with carbon accumulation were expected since it was known that expression of several homologues of pivotal genes implicated in carbon metabolism (e.g., phosphoenolpyruvate carboxykinase, ADP-glucose pyrophosporylase, sucrose synthase, and sucrose phosphatase synthase) and transport (i.e., sucrose transporters) during fruit growth considerably changed [63, 64]. In general, these genes appear to belong to small families including few members, showing differential spatial and temporal expression.

On the other hand, developing citrus fruits accumulate a considerable amount of citric acid in the vacuoles of the juice sac cells, although before ripening this high concentration is considerably reduced. The rate of change and final acid levels are perceived as major components for citrus fruit quality. Research in gene regulation of acid metabolism, however, has not led to a full understanding of this essential process. There is considerable evidence, nevertheless, obtained comparing acidless and acidic varieties that activity and expression of citrate synthase were not responsible for these differences [65]. Another gene characterized in citrus was NADP(+)-isocitrate dehydrogenase (NADP-IDH), encoding for an enzyme involved in citrate metabolism. Recently, a citrate transporter gene has been reported encoding a novel vacuolar citrate/symporter that is able to mediate citrate vacuolar efflux through the electroneutral cotransport of H+ and citrate ions [66]. Interestingly, the transcriptomic study together with the analyses of selected metabolites suggested the occurrence of specific metabolic alternatives during citric acid catabolism [62]. Microarray data suggested that citrate was sequentially metabolized to glutamate that was finally catabolized through the gamma-aminobutyrate (GABA) shunt. This observation was of special relevance since it linked an efficient major proton-consuming reaction with high acid levels. This work provides a convincing explanation for the strong reduction of both citrate and cytoplasmic acidity that takes place in citrus fruit flesh during development and ripening.

Transcript profiling also revealed down-regulation patterns of gene expression for anthocyanin and flavonoid biosynthesis, confirming previous observations. Thus, it was known that in common oranges there was a differential repression of some of the enzymes of anthocyanin biosynthetic pathway, namely chalcone synthase (CHS), anthocyanidin synthase (ANS), and UDP-glucose-flavonoid 3-O-glucosyltransferase (UFGT) [67], in contrast to "blood" pigmented oranges. However, anthocyanin and gene expression associated with anthocyanin synthesis increased at low temperature [68]. Flavanones, a flavonoid subgroup, that greatly contribute to the bitter flavor of grapefruit and other citrus, have also been the subject of intensive work and pivotal genes of this biosynthetic pathway such as CHS, chalcone isomerase (CHI), flavanone 3-hydroxylase (F3H), dihydroflavonol 4-reductase (DFR), and flavonol synthase (FLS) have been isolated and characterized [69, 70]. In an elegant work, Frydman et al. [71] demonstrated that the key flavor-determining step of citrus flavanone biosynthesis was catalyzed by rhamnosyltransferases. They demonstrated that 1,2 rhamnosyltransferases catalyzed biosynthesis of the bitter neohesperidosides, while 1,6 rhamnosyltransferases catalyzed biosynthesis of the tasteless rutinosides. Bitter species, such as grapefruit and pummelo, accumulated bitter flavanone-7-O-neohesperidosides (naringin, the major flavonoid glycoside in grapefruit) responsible, in part, for their characteristic juice flavor, while nonbitter species, such as mandarin and orange, accumulated only tasteless flavanone-7-O-rutinosides.

Bitterness in citrus also is associated with the presence of limonoids, triterpene derivatives that confer the scent to fresh lemon and oranges. Kita et al. [72] isolated a cDNA clone encoding limonoid UDP-glucosyltransferase (limonoid GTase) that regulated the conversion of limonoid aglycones such as limonin, a bitter compound, to their nonbitter glucosides.

In addition to limonoids, citrus fruits possess unique aromas rarely found in other fruit species produced by other terpenes. This is also an area of high research interest. Monoterpenes (*d*-limonene terpinene and pinene) and other low-abundance sesquiterpenes (valencene, nootkatone, and α- and β-sinensal) stand out in citrus as important aroma and also flavor compounds. Lücker et al. [73] and Shimada et al. [74] isolated various monoterpene synthases (*d*-limonene, γ-terpinene, β-pinene synthase, β-ocimene, and cineole synthase), and it has also been shown that their metabolic engineering produced new aromas in tobacco [75]. Monoterpene synthesis takes place in epithelial cells surrounding the secretory cavities that contain the oil glands in the flavedo [76]. Regarding sesquiterpene production, Sharon-Asa et al. [77] identified a sesquiterpene synthase-encoding gene, regulating the conversion of farnesyl diphosphate to a single sesquiterpene, valencene. They reported the transcript that was responsive to ethylene naturally accumulated only towards fruit maturation. Other putative sesquiterpene synthases, such as β-farnesene synthase, have also been cloned.

Work is also in progress to characterize induced mutants that exhibit altered fragrance (*alf*) and abnormal number of oil glands in the flavedo [78]. Transcriptome analysis of fruits from these mutants showed changes in expression profile of genes encoding enzymes involved in the biosynthesis of volatile compounds derived from isoprenoid and phenylpropanoid pathways. In fruits of *alf*, several genes with different biological functions were down-regulated although genes coding for a new putative terpene synthase (TPS) and an O-methyltransferase (OMT), apparently involved in secondary metabolism of volatile compounds, had the highest differences in expression. In the mutant with lower number of glands, transcript profiling also revealed strong down-regulation of genes encoding enzymes from the phenylpropanoid and isoprenoid biosynthetic pathways. In a similar approach, Ishikawa et al. [79] used a 22 K citrus microarray to analyze gene expression in a mutant that shows smoother rind and decreased numbers of oil glands. The authors reported that the genes of the nonmevalonate pathway of isoprenoid synthesis and monoterpene synthases were down-regulated in the mutant.

Tetraterpenes are also crucial components of citrus fruit that contains one of the greatest arrays of carotenoids found in any plant. The simultaneous carotenoid accumulation and chlorophyll reduction occurring during natural ripening indeed determines the color of the fruit peel, a most valuable characteristic of perceived fruit quality. Many pivotal genes of the carotenoid pathway have been cloned in citrus (phytoene synthase (CitPSY), phytoene desaturase (CitPDS), ζ-carotene (car) desaturase (CitZDS), carotenoid isomerase (CitCRTISO), lycopene β-cyclase (CitLCYb), β-ring hydroxylase (CitHYb), zeaxanthin (CitZEA) epoxidase (CitZEP), lycopene β-cyclase (CitLCYb), and lycopene ε-cyclase (CitLCYe), and their expression has been correlated with the accumulation of carotenoids in fruit [80, 81]. It was reported that the transition of peel color from green to orange, and the change from β,ε-carotenoid to β,β-carotenoid accumulation was accompanied by the disappearance of CitLCYe and the increase in CitLCYb transcripts. As fruit maturation progressed, a concomitant increase in the expression of CitPSY, CitPDS, CitZDS, CitLCYb, CitHYb, and CitZEP led to massive β,β-xanthophyll accumulation. Cercós et al. [62] showed that expression of carotenoid biosynthetic genes in fruit flesh followed rather similar changes. Mutations of flesh color are being investigated in China using a citrus cDNA array with 6000 unigenes [52].

In contrast to carotenoid accumulation, there have been fewer studies of the chlorophyll degradation processes in citrus. Previous work on the regulation of catabolism showed that chlorophyllase (*CLH*) was constitutively expressed during natural fruit development [82]. Recent results suggest that *CLH* functions as a rate-limiting enzyme in chlorophyllcatabolism controlled via post-translational regulation [83]. It is also known that pheophorbide a oxygenase (*PaO*) and geranylgeranyl reductase expression, correlated with chlorophyll degradation [57]. Recent work upon “nan,” a stay-green mutant of Navel orange that produces fruit with abnormal brown flavedo, showed that typical ripening-related chlorophyll (Chl) degradation was impaired in this mutant. Transcript and proteomic profilings revealed that a citrus orthologue of a number of *SGR* (*stay green*) genes was expressed at substantially lower levels in “nan” both prior to and during ripening [84]. The “*nan*” mutation also resulted in the suppressed expression of numerous photosynthesis-related genes and in the induction of genes associated with oxidative stress. The transcriptome of other selected citrus mutants is also being investigated to identify gene functions related to fruit quality that in citrus are barely accessible through genetic approaches. To this end, three collections of induced mutated lines (EMS, gamma rays and fast neutrons) have been generated, comprising 10 000 potential [84] mutants.

With the exception of thermostable pectin methylesterase activity [85, 86] that greatly reduces citrus juice quality, cell wall metabolism in citrus has been studied as related to fruit abscission, a major component of final yield. One of the strategies for the identification of abscission-related genes followed by Dr. Burns’ team (University of Florida) was based upon the isolation of ethylene-induced genes in the calyx, the laminar, and the floral abscission zones. The role of ethylene on the regulation of abscission has been widely illustrated for decades, and several works have shown that ethylene is the primary effector activating the abscission pathway in citrus [87, 88]. Differential display and subtractive cDNA library screening were also used to search for abscission-related metabolism changes. Important components of the citrus abscission process were thus associated with expression and/or activity of pivotal enzymes of cell wall metabolism (glucanases, polygalacturonases, galactosidases, and other hydrolases; [89, 90]), hormonal synthesis, and signal transduction (i.e., ACC synthases and oxydases) and secondary metabolism/PR proteins (i.e., phenylalanine ammonia lyase, chitinases). Yuan et al. [91] also demonstrated that differential expression of ACC synthase 1 and ACC oxidase genes was associated with reduction of ethephon-enhanced leaf abscission by guanfacine, a G-protein-coupled alpha-(2A)-adrenoreceptor selective antagonist, and suggested a link between G-protein-related signalling and abscission. Interestingly, guanfacine had little effect on ethephon-enhanced fruit loosening. In spite of this information, major regulators of the abscission process in citrus are still mostly unknown although both custom manufactured and large-scale microarrays, in some instances, coupled to laser assisted microdissection (LAM) are currently being used in Florida and Spain to gain new insight into this process. Part of these transcriptomic profiling studies has been summarized in a recent Ph.D. dissertation presenting a model of leaf abscission events occurring at the laminar abscission zone [60]. The two-stage model proposes a first phase of activation, mostly characterized by the activation of signalling pathways (hormones, phospholipids, calcium, and oxygen reactive species). In a second stage, the execution phase, degradation of the cell wall by hydrolytic enzymes would be culminated and sugar-nucleotide metabolism for cell elongation induced. The process would end with the promotion of a double defensive program intended to protect the living zone remaining attached to the plant including deposition of physical barriers (callose and lignin) and induction of pathogen resistance.

Several other microarray studies on citrus growth and ripening are under development and have not been published yet. For instance, Dr. Sadka is investigating with the GeneChip Citrus Genome Array (Affymetrix) the transcriptome modifications occurring during the induction of flower bud differentiation using "on" and "off" trees [51], taking advantage of the alternate bearing behavior, a process regulating differentially carbohydrate-related gene expression [92]. In Brazil, the sequencing carried out at the Centro APTA Citros "Sylvio Moreira"—IAC (Brazil) that has generated one of the most important databases for this genus in the world is being extensively used to produce "*in silico*" analyses. This approach is yielding information not only related to fruit growth and development (terpene production, cell wall metabolism, etc.) but also in the biotic stress field [48].

5.3.2. *Responses to pathogenic and environmental stresses*

In citrus, gene expression associated with the responses to biotic and abiotic stresses has targeted a limited number of genes in spite of the economical importance of the citrus diseases and environmental constrains. Multiple pathogens provoke a range of citrus disorders, mostly fungal (leaf spot, *Alternaria*; mold, *Penicillium*; post-bloom fruit drop, *Colletotrichum acutatum*; root rot, *Phytophthora*), bacterial (canker, *Xanthomonas axonopodis*; citrus variegated chlorosis, *Xylella fastidiosa*; Huanglongbing or greening, *Candidatus* Liberibacter), and viral diseases (citrus tristeza virus, CTV; citrus leprosis virus, CiLV). Environmental stresses include cold temperatures, drought, flooding, salinity, and high and low soil pH, among others.

In response to the inoculation with conidia of *Alternaria*, at least two cytosolic antifungal miraculins with protease inhibitor activity were strongly up regulated. Actually, induction of miraculin expression is one of the most prominent responses observed in microarray experiments performed in open field experiments. Both miraculin genes responded to methyl jasmonate and were antagonized by salicylate [93]. Several studies with the green mold pathogen, *Penicillium digitatum*, also indicated that genes such as thioredoxins, the *gnsl* gene (beta-1,3-endoglucanase activity), and chitinases are major components of the molecular mechanisms involved in activation of pathogen defense in citrus. Other responsive genes reported in citrus were epoxide hydrolase and hidroperoxide lyase. It has also been shown that the fungus *Colletotrichum acutatum* altered hormonal homeostasis increasing both levels of ethylene, indole-3-acetic acid, cis-jasmonic acid (JA) and salicylic acid (SA), and associated gene expression [94].

Gandía et al. [95] have recently presented data on the transcriptional response of citrus to infection with severe and mild isolates of citrus tristeza virus. These studies concluded that gene expression was only significantly altered with the severe isolate. Changes detected in the citrus transcriptome after infection with this isolate were predominantly associated with symptom expression (chlorophyllases, SAM transferases, ACC oxidase, and lipid transfer proteins), defense mechanism, and general responses to stress (miraculins, superoxide dismutases, glutation transferees, NBS-LRR resistance genes, thioredoxin, protease inhibitors, ubiquitin ligases, etc.).

To study the mechanisms of canker resistance in kumquat, a custom microarray using 2254 ESTs from subtractive libraries is being utilized to determine the response to infective bacteria in an incompatible interaction [54]. The macroscopic phenotype, a delayed hypersensitive response in the inoculated leaves, was accompanied by altered expression of 1245 genes. This study identified major components of the incompatible interaction, reactive oxygen species (ROS) production, and programmed cell death (PCD). In addition, a number of common defense mechanisms besides a number of resistance genes and putative receptors were also identified.

Citrus plants are also very liable to infestation by aphids, whitefly, and other insects as well as being susceptible to herbivory. Mozoruk et al. [53] described how nylon filter cDNA arrays were used to analyze the transcriptional changes of 1731 citrus unigenes that resulted from herbivory by a xylem-feeding leafhopper, *Homalodisca coagulata*. Insect feeding led to a significant expression change in 50 transcripts broadly functioning in direct defense, defense signalling, ROS scavenging, transport, cell wall modification, photosynthesis, and abiotic stress. The authors also noted that the transcript profile recorded greatly resembled that induced by wounding, likely through JA-independent pathways. In contrast to similar studies with aphids, SA-dependent pathogenesis related genes were weakly induced.

Although transcriptional profiling using microarrays has developed into the most prominent tool for functional genomics, none has yet reported on the effects and responses of citrus to the major environmental constraints (salinity, flooding, water deficit, chilling, and iron deficiency). High-throughput analyses of gene expression in citrus challenged with major abiotic stresses, however, are currently underway in several laboratories around the world and will soon produce valuable information that might eventually lead to discovery of novel genes and functions. For instance, it is known that in citrus, physiological disturbances produced by salinity are associated with leaf chloride build up rather than with sodium accumulation, as observed in many plants [96, 97]. Genes in principle associated with the response of citrus to salinity were initially obtained from a cDNA expression library of citrus salt-treated cell suspensions. These genes, homologues to phospholipid hydroperoxides, glutathione peroxidases [98], olesins, Lea5, or lypoxygenases, were involved in the oxidative response rather than in the specific response to salinity. Other genes involved in oxidative stress well known in citrus are glutathione S-transferases [99] and copper/zinc-superoxide dismutases. However, recent microarray analyses are providing much-needed insights into chloride tolerance mechanisms and short- and long-term adaptation of citrus to salinity. Several teams are engaged in a Euro-Mediterranean MPC INCO project, started in 2006, between Spain, Morocco, Tunisia, Turkey, and France focused on tolerances to salinity and iron deficiency associated with alkaline soils. One of the major components of this project is the large-scale study and genome-wide acquisition of quantitative biological information on gene expression from multiple tolerant and susceptible genotypes. Work on transcriptomic comparisons in this area is confirming that Cl^- is the most important ion involved in the genetic response of citrus to salinity. In addition, major metabolic regulation changes are also apparent during salinity acclimatization in tolerant rootstocks. In contrast, flooding is mostly characterized by the rise of oxidative stress.

Chilling resistance in citrus is another area that has received much attention but lacks current comprehensive gene expression analyses provided by microarrays. Since most commercially important citrus varieties are cold-sensitive and therefore susceptible to freezing, *Poncirus trifoliata* (L.), an interfertile *Citrus* relative that can tolerate temperatures as low as -26°C after acclimation, is being used for improving cold tolerance in citrus rootstocks and as a source for the identification of cold-regulated genes. In general, many studies have been performed through subtractive hybridization [100] and DDRT-PCR [101] comparing expression in sensitive and resistant varieties. It has been shown, for example, that expression of a *C-repeat-binding factor* (*CBF*) and one of its targets, *COR19*, a cold-induced gene, accumulated both earlier and to higher levels in *Poncirus*. Moreover, COR19, COR11 [102], and COR15 were found to belong to an unusual group 2 LEA gene family responsive to low temperature. These dehydrins differ from most other plant dehydrins in having an unusual K-segment similar to that of gymnosperms and in having a serine cluster (S-segment) at an unusual position at the carboxy-terminus [103]. Citrus, however, also possesses the typical plant angiosperm-type K-segment consensus sequence. Other up-regulated transcripts that may play a role in cold sensitivity are a novel RING-H2 finger gene, AP2 domain containing genes and CTL, and a homologue of a low-temperature-responsive gene from *Arabidopsis*. During postharvest storage, chilling injury in citrus fruit can be reduced by previous short heat treatments that activate different molecular responses. Genes differentially expressed in the chilling response have mostly been related to lipid membrane and cell wall enzymes, to main regulators of secondary metabolism and hormonal homeostasis, and to oxidative and general stress responses [104, 105].

Although the main applications of microarrays to date are in transcriptome profiling analyses, microarrays can also be used to study DNA variation. Oligonucleotide arrays are particularly suited for the detection of single nucleotide mismatches during hybridization and, hence, for the discovery of novel DNA variants or the determination of known variants. The citrus GeneChip, for example, was designed to genotype 3219 genes using 5023 SNPs. The 20 K Spanish Consortium microarrays have been used to identify heterozygous deletions in fast neutron irradiated citrus mutants through array-based comparative genomic hybridization (array-CGH) and to study gene colinearity. Preliminary CGH yielded several candidate genes that were in haploid gene dosage. After comparison with the *Arabidopsis* and *Populus* genomes, it was observed that *Populus* orthologues of *Citrus* deleted genes grouped in two duplicated chromosomes in contrast to *Arabidopsis* orthologues that were distributed in several chromosomes (Ríos and Talon, unpublished data).

5.4. Genetic transformation

Citrus transformation procedures, in general, follow *Agrobacterium tumefaciens* protocols, and subsequent regeneration through organogenesis and somatic embryogenesis are also rather typical and straightforward [106]. Transformation efficiency of young material is usually low, 15–20%, and a major achievement to overcome the juvenility limitation was the direct transformation of adult material [107].

In citrus, genetic transformation is mostly being explored as an alternative to classical genetic breeding and not many examples can be found in the literature illustrating the use of genetic transformation for functional genomics. For example, there is interest in modulating the growth habit of rootstocks since this might eventually affect the development of the scion and facilitate diverse cultural practices (e.g., pruning, pesticide applications, and harvesting). Thus, it was known that the ectopic overexpression in tobacco of a citrus GA 20-oxidase, a regulatory step of gibberellin biosynthesis in citrus, [108] enhanced gibberellin content and shoot growth [109]. Later, Fagoaga et al. [110] generated transgenic Carrizo rootstocks overexpressing this GA 20-oxidase and confirmed that the gene controls gibberellin flux through the pathway since taller (sense) and shorter (antisense) phenotypes correlating with higher and lower levels of active GA_1 were obtained. In these transgenic lines, however, cell division was more affected than cell elongation, in contrast to the effects observed in herbaceous plants [111]. In another example, an antisense construction with a citrus ACC synthase gene repressed ACC increase after a chilling treatment. A pectin methylesterase gene (*Cs-PME4*) isolated from sweet orange to prevent juice cloud separation was also introduced via protoplasts and subsequent regeneration through somatic embryogenesis [112].

Generally, characteristics related to commercial valuable traits are modified through the use of transgenes. To accelerate flowering time, Carrizo seedlings constitutively overexpressing the *Arabidopsis* floral-regulatory genes *LEAFY* (*LFY*) or *APETALA1* (*AP1*) were generated [113]. Both kind of transgenic citrus produced fertile flowers in their first year considerably shortening the juvenile phase. Consistently with the role of *LFY* and *AP1*, juvenility in citrus was positively correlated with CsTFL (homolog to TERMINAL FLOWER) transcript accumulation and negatively correlated with LEAFY and APETALA1 RNA levels [114]. In a similar approach but with a citrus gene, it was showed that transgenic *Poncirus* carrying the *CiFT* gene (homolog to *FLOWERING LOCUS T*), another flowering time gene, also exhibited early flowering although this phenotype was accompanied with several pleiotropic effects [115]. It is possible that the early flowering *AP1* and *CiFT* transgenic citrus could be used as rapid cycling genotypes for functional genomics studies. In a further example, Carrizo rootstock constitutively expressing a Δ^1-pyrroline-5-carboxylate synthetase mutant gene from *Vigna*, showed higher water deficit tolerance [116]. Regarding tolerance to stresses, however, a huge amount of work has been centered on resistance to biotic stresses, a matter of major relevance in citrus industry. Thus, tolerance or resistance to *Phytophthora citrophthora*, the most widely spread oomycete in citrus growing areas, was generated by introducing the gene P23, that codes for a pathogenesis-related protein induced in tomato. These results provided evidence for the antifungal activity in vivo of the P23 pathogenesis-related protein against *P. citrophthora* [117].

A great effort is also being developed to understand the basis of the tolerance to citrus tristeza virus (CTV), the causal agent of the most important virus disease in citrus. The strategy is generally supported by the concept of pathogen-derived resistance (PDR), based on expression of viral sequences interfering with the virus life cycle in plants. CTV resistant transformants have been obtained by genetically engineering the *p25* and *p23* genes from CTV [118]. However, it still remains to be elucidated if transgenic citrus plants expressing CTV-derived sequences are a plausible alternative to cross protection to control CTV strains in the field. In an alternative strategy, heterologous expression of plant-derived resistance genes is promoted to confer resistance against CTV. General resistance to CTV has been found in *Poncirus trifoliata*, and a region containing the resistance gene (*Ctv*) has been characterized. Furthermore, work is under way for other pivotal diseases such as citrus mosaic virus (CiMV), citrus canker, and citrus blight.

5.5. Reverse genetics

In addition to genetic transformation, the capability to perform reverse genetic analyses is crucial to develop functional studies. The creation of transgenic lines is a powerful and straightforward way to determine gene function. However, in citrus, high-throughput transgenic programs such as the generation of RNA interference knockouts, activation tagging through enhancer elements, gene-trap T-DNA insertions, or transposable tagging systems have not yet been developed. The capacity for the maintenance and characterization of many transgenic lines of a perennial tree with both a long juvenile phase, large individual plant size, and a complex reproductive biology has probably hindered these developments. In *Populus*, however, activation tagging and insertional mutagenesis approaches are been explored despite logistical challenges in working with transgenic trees, a direction that may well be followed by citrus researchers in the near future.

5.6. Tilling/fast neutrons

Since gene disruption is the most effective method to analyze gene functions and no efficient tagging or insertional methods are available in citrus, strategies based on genome-wide mutagenesis such as TILLING (targeted induced local lesions in genomes) and fast neutron mutagenesis are being explored further [84]. These approaches are nontransgenic and have particular interest for the industry where the debate on GMOs has restricted their application in crop improvement. TILLING identifies individuals carrying point mutations while the fast neutron mutagenized population is searched for gene deletions using PCR amplification. Both approaches, at the moment, are of limited usefulness as strategies for reverse genetics in citrus because of the lack of genomic sequences and the large amounts of space required for mutated populations of suitable size. ECOTILLING, however, on natural citrus variants and microarray-based detection of deletions on fast neutron citrus mutants in a more direct genetics strategy are very straightforward approaches. However, unless a high-throughput transformation protocol is developed for citrus, functionally analyzing all genes with tagging approaches or genome-wide mutagenesis and screening are not realistic strategies.

5.7. Viral-induced gene silencing

Viral-induced gene silencing (VIGS), on the other hand, is an attractive and very promising alternative in citrus. Knocking out the expression of a gene by VIGS does not require genetic transformation and has proven to be a very efficient tool for function analysis of plant genes. VIGS is particularly suitable for woody plants like citrus with long juvenile periods that require long periods between transformation and fruiting. In a hopeful work, Dr. Guerri and colleagues at IVIA have recently showed that VIGS might be possible in citrus using *Citrus leaf blotch virus* (CLBV) as a viral vector (Dr. Guerri, personal communication). These workers cloned a full-length cDNA of the CLBV genome [119] in a binary vector under the control of the 35S promoter and demonstrated that tobacco and citrus plants Agro-infiltrated with this construct became infected and replicated CLBV normally. Recently, they have showed that tobacco plants Agro-infiltrated with a CLBV chimeric construct carrying a fragment of the phytoene desaturase gene developed photobleaching symptoms and reduced the cognate transcripts. Parallel experiments in citrus are planned. Availability of the CLBV-based vector will certainly open new possibilities to study functional genomics in citrus.

5.8. Proteomics/metabolomics

Other powerful approaches for functional genomics studies such as proteomics and metabolomics to comprehensively analyze proteomes and phenotypes have just begun for citrus. For example, Blumwald and coworkers (UCD) are using two main approaches, namely 2D gel analyses coupled with MALDI-TOF-TOF from juice sac cell vacuoles and LC^2-MS-MS analyses of ER/Golgi, plasma membrane, tonoplast, mitochondria, and soluble enriched fractions from citrus juice sac cells to define the "citrus fruit proteome." Current work is in progress but they have already reported the identification of over 1500 proteins involved in sugar metabolism, citrate cycle, signalling, transport, and processing and have characterized changes in protein expression during development [120]. In a further example, proteome changes in the fruit albedo during postharvest ageing were studied through 2D-PAGE, and relevant proteins were also identified through mass spectrometry determinations [121]. This proteomic survey indicated that major changes in protein content (ATP synthase beta subunit, ascorbate peroxidase, translationally controlled tumor protein, cysteine protease, etc.) were apparently related to the activation of programmed cell death.

Numerous analyses of citrus metabolites, especially of ripening and matured fruits, have been reported in the past. However, new methods to characterize the metabolic phenotypes of representative lines from mutants and natural varieties must be developed. Metabolic profiling and metabolomic procedures using state-of-the-art gas chromatography-mass spectrometry or fast gas chromatography-time-of-flight mass spectrometry need to be setup.

The final objective of citrus functional genomics is to identify candidate genes, alleles, and genotypes improving citrus fruit quality, correlating phenotypic analyses, metabolomic profiling, and gene expression. At completion, genes and alleles with major functions in nutritional quality could be selected and genotypes with improved fruit composition searched among existing collections or generated.

6. CONCLUSION AND FUTURE PROSPECTS

This paper has reviewed various aspects of the current status of citrus genome research, including the development of fundamental tools, the applications currently under way and envisaged leading to solutions to seemingly intractable problems facing the citrus industries of the world, the opportunities of improving further the perceived and real value of citrus fruit and products, and the challenges that remain not only for genomic research but for making progress in truly incorporating new knowledge into new plant materials. The international citrus research community has been growing closer together, and new international alliances are making the achievement of truly great advances possible; this is essential, as no one group or even nation has sufficient resources to address all the needs for tool development and deployment, and many of the problems faced are global in nature. It is clearly evident that by combining research resources and by adopting the principle of depositing information in the public domain, freely available to global research partners, the promise of genome research to improve citrus plants, production, and protection from diseases, and enhanced product quality and value, can be realized. The free availability of these tools and materials is truly the key to the success in genomics research. Citrus is a very important tree fruit crop throughout the world not only is it of great economic significance but it is also of great value for human nutrition and well-being. In addition, it possesses many unique characteristics of great biological interest. Consequently, the benefits of an expanded and focused effort into all aspects of citrus genomics will be of great benefit to humanity in general as well as to the realm of plant science. Citrus will have a first genome sequenced in the very near future; this will not be the end of the process but the beginning of many more citrus genome sequencing projects to add layers of valuable information to the already developed and developing tools to understand the functions and interrelationships of genes, their products, and their interactions with the environment. Through the acquisition of this knowledge and its application to the field, citrus will continue to be an economically valuable fruit crop plant and a source of important health and nutrition benefits to people throughout the world.

ACKNOWLEDGMENTS

Work at Centro de Genómica was supported by INIA grants RTA04-013 and 05-247, INCO contract 015453, and Ministerio de Educación y Ciencia grant AGL2007-65437-C04-01/AGR. The help and expertise of A. Almenar, E. Blázquez, A. Boix, I. López, A. López García-Usach, I. Sanchís, and M. Sancho are gratefully acknowledged. Work at the University of Florida's Citrus Research and Education Center was supported in the past several years by various grants from USDA-NRI, USDA-CSREES, and from the Florida Citrus Production Research Advisory Council. In addition to cited authors, assistance and expertise in care and management of plant materials and technical assistance by M. Holt, L. Wang, Y. Wang, and M. K. Wendell and manuscript preparation by B. A. Thompson are gratefully acknowledged.

REFERENCES

[1] Food Agriculture Organization of the United Nations, "Developments in international citrus trade in 2004-2005," 2006, http://www.fao.org/es/esc/en/20953/20990/highlight_28187en.html.

[2] Food Agriculture Organization of the United Nations, "Annual Statistics," Rome, Italy, 2003.

[3] F. G. Gmitter Jr. and X. Hu, "The possible role of Yunnan, China in the origin of contemporary *Citrus* species (Rutaceae)," *Economic Botany*, vol. 44, no. 2, pp. 267–277, 1990.

[4] R. K. Soost and J. W. Cameron, "Citrus," in *Advances in Fruit Breeding*, J. Janick and J. N. Moore, Eds., pp. 507–540, Purdue University Press, West Lafayette, Ind, USA, 1975.

[5] J. W. Cameron and R. K. Soost, "Citrus," in *Evolution of Crop Plants*, N. W. Simmonds, Ed., pp. 261–264, Longman, New York, NY, USA, 1984.

[6] R. K. Soost and M. L. Roose, "Citrus," in *Fruit Breeding, Vol 1: Tree and Tropical Fruits*, J. Janick and J. N. Moore, Eds., pp. 257–324, John Wiley & Sons, New York, NY, USA, 1996.

[7] F. G. Gmitter Jr., J. W. Grosser, and G. A. Moore, "Citrus," in *Biotechnology of Perennial Fruit Crops*, F. A. Hammerschlag and R. E. Litz, Eds., pp. 335–369, CABI, Wallingford, Conn, USA, 1992.

[8] A. Wakana and S. Uemoto, "Adventive embryogenesis in *Citrus*. I. The occurrence of adventive embryos without

pollination or fertilization," *American Journal of Botany*, vol. 74, no. 4, pp. 517–530, 1987.

[9] F. G. Gmitter Jr., "Origin, evolution and breeding of the grapefruit," in *Plant Breeding Reviews*, J. Janick, Ed., vol. 13, pp. 345–363, John Wiley & Sons, New York, NY, USA, 1995.

[10] K. Arumuganathan and E. D. Earle, "Nuclear DNA content of some 479 important plant species," *Plant Molecular Biology Reporter*, vol. 9, no. 3, pp. 208–218, 1991.

[11] A. M. Torres, T. Mau-Lastovicka, T. E. Williams, and R. K. Soost, "Segregation distortion and linkage of *Citrus* and *Poncirus* isozyme genes," *Journal of Heredity*, vol. 76, no. 4, pp. 289–294, 1985.

[12] P.-C. Liou, "A molecular study of the *Citrus* genome through analysis of restriction fragment length polymorphism and isozyme mapping," Ph.D. dissertation, University of Florida, Gainesville, Fla, USA, 1990.

[13] R. E. Durham, P. C. Liou, F. G. Gmitter Jr., and G. A. Moore, "Linkage of restriction fragment length polymorphisms and isozymes in *Citrus* ," *Theoretical and Applied Genetics*, vol. 84, no. 1-2, pp. 39–48, 1992.

[14] D. C. Jarrell, M. L. Roose, S. N. Traugh, and R. S. Kupper, "A genetic map of citrus based on the segregation of isozymes and RFLPs in an intergeneric cross," *Theoretical and Applied Genetics*, vol. 84, no. 1-2, pp. 49–56, 1992.

[15] Q. Cai, C. L. Guy, and G. A. Moore, "Extension of the genetic linkage map in *Citrus* using random amplified polymorphic DNA (RAPD) markers and RFLP mapping of cold-acclimation responsive loci," *Theoretical and Applied Genetics*, vol. 89, no. 5, pp. 606–614, 1994.

[16] A. A. Sankar and G. A. Moore, "Evaluation of inter-simple sequence repeat analysis for mapping in Citrus and extension of the genetic linkage map," *Theoretical and Applied Genetics*, vol. 102, no. 2-3, pp. 206–214, 2001.

[17] J. M. H. Kijas, M. R. Thomas, J. C. S. Fowler, and M. L. Roose, "Integration of trinucleotide microsatellites into a linkage map of Citrus," *Theoretical and Applied Genetics*, vol. 94, no. 5, pp. 701–706, 1997.

[18] M. L. Roose, D. Feng, F. S. Cheng, R. I. Tayyar, C. T. Federici, and R. S. Kupper, "Mapping the *Citrus* genome," in *Proceedings of the 1st International Society for Horticultural Sciences (ISHS '00)*, R. Goren and E. E. Goldschmidt, Eds., vol. 535, pp. 25–32, Acta Horticulturae, Leuven, Belgium, September 2000.

[19] M. de Simone, M. P. Russo, G. Puelo, et al., "Construction of genetic maps for *Citrus aurantium* and *C. latipes* based on AFLP, RAPD and RFLP markers," *Fruits*, vol. 53, no. 6, pp. 383–390, 1998.

[20] P. Ling, C. Yu, Z. Deng, et al., "Citrus genome mapping with AFLP markers," in *Plant & Animal Genome VII Conference*, p. 189, San Diego, Calif, USA, January 1999.

[21] C. Chen, K. D. Bowman, Y. A. Choi, et al., "EST-SSR genetic maps for *Citrus sinensis* and *Poncirus trifoliata*," *Tree Genetics & Genomes*, vol. 4, no. 1, pp. 1–10, 2007.

[22] Z. Deng, S. Huang, S. Xiao, and F. G. Gmitter Jr., "Development and characterization of SCAR markers linked to the citrus tristeza virus resistance gene from *Poncirus trifoliata*," *Genome*, vol. 40, no. 5, pp. 697–704, 1997.

[23] F. G. Gmitter Jr., S. Y. Xiao, S. Huang, X. L. Hu, S. M. Garnsey, and Z. Deng, "A localized linkage map of the citrus tristeza virus resistance gene region," *Theoretical and Applied Genetics*, vol. 92, no. 6, pp. 688–695, 1996.

[24] R. García, M. J. Asíns, J. Forner, and E. A. Carbonell, "Genetic analysis of apomixis in *Citrus* and *Poncirus* by molecular markers," *Theoretical and Applied Genetics*, vol. 99, no. 3-4, pp. 511–518, 1999.

[25] M. Omura, T. Ueda, M. Kita, et al., "EST mapping of Citrus," in *Proceedings of the International Society of Citriculture IX Congress*, vol. 1, pp. 71–74, Orlando, Fla, USA, December 2000.

[26] M. Omura, T. Ueda, M. Kita, et al., "Extension of *Citrus* linkage map by CAPS marker," in *Plant & Animal Genome IX Conference*, p. 538, San Diego, Calif, USA, January 2001.

[27] T. J. Close, S. Wanamaker, M. Lyon, G. W. Mei, C. Davies, and M. L. Roose, "A GeneChip® for citrus," in *Plant & Animal Genome XIV Conference*, p. 82, San Diego, Calif, USA, January 2006.

[28] C. Chen, P. Zhou, Y. A. Choi, S. Huang, and F. G. Gmitter Jr., "Mining and characterizing microsatellites from citrus ESTs," *Theoretical and Applied Genetics*, vol. 112, no. 7, pp. 1248–1257, 2006.

[29] M. T. Lyon, C. T. Federici, Y. Kacar, et al., "SST-based linkage maps for sweet orange and trifoliate orange," in *Plant & Animal Genome XIV Conference*, p. 480, San Diego, Calif, USA, January 2007.

[30] F. Luro, G. Constantino, C. Billot, et al., "Genetic maps of Clementine mandarin and intergeneric Clementine × Poncirus using genomic and EST microsatellite markers," in *Plant & Animal Genome XIV Conference*, p. 487, San Diego, Calif, USA, January 2007.

[31] Z.-N. Yang, X.-R. Ye, S. Choi, et al., "Construction of a 1.2-Mb contig including the citrus tristeza virus resistance gene locus using a bacterial artificial chromosome library of *Poncirus trifoliata* (L.) Raf," *Genome*, vol. 44, no. 3, pp. 382–393, 2001.

[32] Z. Deng, Q. Tao, Y.-L. Chang, et al., "Construction of a bacterial artificial chromosome (BAC) library for citrus and identification of BAC contigs containing resistance gene candidates," *Theoretical and Applied Genetics*, vol. 102, no. 8, pp. 1177–1184, 2001.

[33] Z.-N. Yang, X.-R. Ye, J. Molina, M. L. Roose, and T. E. Mirkov, "Sequence analysis of a 282-kilobase region surrounding the citrus tristeza virus resistance gene (Ctv) locus in *Poncirus trifoliata* L. Raf," *Plant Physiology*, vol. 131, no. 2, pp. 482–492, 2003.

[34] M. Rai, "Refinement of the Citrus tristeza virus resistance gene (*Ctv*) positional map in *Poncirus trifoliata* and generation of transgenic grapefruit (*Citrus paradisi*) plant lines with candidate resistance genes in this region," *Plant Molecular Biology*, vol. 61, no. 3, pp. 399–414, 2006.

[35] Z. Deng, S. Huang, P. Ling, et al., "Cloning and characterization of NBS-LRR class resistance-gene candidate sequences in citrus," *Theoretical and Applied Genetics*, vol. 101, no. 5-6, pp. 814–822, 2000.

[36] Z. Deng and F. G. Gmitter Jr., "Cloning and characterization of receptor kinase class disease resistance gene candidates in Citrus," *Theoretical and Applied Genetics*, vol. 108, no. 1, pp. 53–61, 2003.

[37] J. Terol, M. A. Naranjo, and M. Talon, "BAC end sequencing analysis of 28,000 genomic clones from *Citrus clementina*," in *Plant & Animal Genome XV Conference*, p. 80, San Diego, Calif, USA, January 2007.

[38] T. Shimizu, T. Shimada, F. Nishikawa, et al., "Constructing a physical map of *Citrus unshiu*," in *Plant & Animal Genome XV Conference*, p. 481, San Diego, Calif, USA, January 2007.

[39] L. Rico-Cabanas and J. A. Martínez-Izquierdo, "CIRE1, a novel transcriptionally active *Ty1-copia* retrotransposon

from *Citrus sinensis*," *Molecular Genetics and Genomics*, vol. 277, no. 4, pp. 365–377, 2007.

[40] G. P. Bernet and M. J. Asíns, "Identification and genomic distribution of gypsy like retrotransposons in *Citrus* and *Poncirus*," *Theoretical and Applied Genetics*, vol. 108, no. 1, pp. 121–130, 2003.

[41] M. G. Bausher, N. D. Singh, S.-B. Lee, R. K. Jansen, and H. Daniell, "The complete chloroplast genome sequence of *Citrus sinensis* (L.) Osbeck var 'Ridge Pineapple': organization and phylogenetic relationships to other angiosperms," *BMC Plant Biology*, vol. 6, article 21, pp. 1–11, 2006.

[42] M. L. Roose, R. P. Niedz, F. G. Gmitter Jr., et al., "Analysis of a 1.2x whole genome sequence of *Citrus sinensis*," in *Plant & Animal Genome XV Conference*, p. 81, San Diego, Calif, USA, January 2007.

[43] S. Hisada, T. Akihama, T. Endo, T. Moriguchi, and M. Omura, "Expressed sequence tags of Citrus fruit during rapid cell development phase," *Journal of the American Society for Horticultural Science*, vol. 122, no. 6, pp. 808–812, 1997.

[44] M. Bausher, R. Shatters, J. Chaparro, P. Dang, W. Hunter, and R. Niedz, "An expressed sequence tag (EST) set from *Citrus sinensis* L. Osbeck whole seedlings and the implications of further perennial source investigations," *Plant Science*, vol. 165, no. 2, pp. 415–422, 2003.

[45] T. Shimada, M. Kita, T. Endo, et al., "Expressed sequence tags of ovary tissue cDNA library in *Citrus unshiu* Marc," *Plant Science*, vol. 165, no. 1, pp. 167–168, 2003.

[46] J. Forment, J. Gadea, L. Huerta, et al., "Development of a citrus genome-wide EST collection and cDNA microarray as resources for genomic studies," *Plant Molecular Biology*, vol. 57, no. 3, pp. 375–391, 2005.

[47] J. Terol, A. Conesa, J. M. Colmenero, et al., "Analysis of 13000 unique *Citrus* clusters associated with fruit quality, production and salinity tolerance," *BMC Genomics*, vol. 8, article 31, pp. 1–22, 2007.

[48] M. A. Machado, A. M. Amaral, J. F. Astua, et al., "Analysis of citrus transcriptome: CitEST in Brazil," in *Plant & Animal Genome XV Conference*, p. 77, San Diego, Calif, USA, January 2007.

[49] T. Shimada, H. Fuiii, T. Endo, et al., "Toward comprehensive expression profiling by microarray analysis in citrus: monitoring the expression profiles of 2213 genes during fruit development," *Plant Science*, vol. 168, no. 5, pp. 1383–1385, 2005.

[50] H. Fujii, T. Shimada, A. Sugiyama, et al., "Profiling ethylene-responsive genes in mature mandarin fruit using a citrus 22K oligoarray," *Plant Science*, vol. 173, no. 3, pp. 340–348, 2007.

[51] R. Goldberg-Moeller, L. Shlizerman, N. Zur, E. Or, and A. Sadka, "The use of genomic approach to identify factors involved in the induction of flower bud differentiation in Citrus," in *Plant & Animal Genome XV Conference*, p. 78, San Diego, Calif, USA, January 2007.

[52] X. Deng, N. Tao, Q. Liu, J. Ye, Y. Cheng, and W. Guo, "Research on the molecular mechanism of flesh color mutant of sweet orange," in *Plant & Animal Genome XV Conference*, p. 79, San Diego, Calif, USA, January 2007.

[53] J. Mozoruk, L. E. Hunnicutt, R. D. Cave, W. B. Hunter, and M. G. Bausher, "Profiling transcriptional changes in *Citrus sinensis* (L.) Osbeck challenged by herbivory from the xylem-feeding leafhopper *Homalodisca coagulata* (Say) by cDNA macroarray analysis," *Plant Science*, vol. 170, no. 6, pp. 1068–1080, 2006.

[54] A. A. Khalaf, J. Jones, G. A. Moore, and F. G. Gmitter Jr., "Microarray expression analysis in Nagami kumquat in response to canker infection," in *Plant & Animal Genome XV Conference*, p. 485, San Diego, Calif, USA, January 2007.

[55] S. Jansson and C. J. Douglas, "*Populus*: a model system for plant biology," *Annual Review of Plant Biology*, vol. 58, pp. 435–458, 2007.

[56] E. Katz, P. M. Lagunes, J. Riov, D. Weiss, and E. E. Goldschmidt, "Molecular and physiological evidence suggests the existence of a system II-like pathway of ethylene production in non-climacteric Citrus fruit," *Planta*, vol. 219, no. 2, pp. 243–252, 2004.

[57] E. Alós, M. Cercós, M.-J. Rodrigo, L. Zacarías, and M. Talón, "Regulation of color break in citrus fruits. Changes in pigment profiling and gene expression induced by gibberellins and nitrate, two ripening retardants," *Journal of Agricultural and Food Chemistry*, vol. 54, no. 13, pp. 4888–4895, 2006.

[58] M. J. Rodrigo and L. Zacarías, "Effect of postharvest ethylene treatment on carotenoid accumulation and the expression of carotenoid biosynthetic genes in the flavedo of orange (*Citrus sinensis* L. Osbeck) fruit," *Postharvest Biology and Technology*, vol. 43, no. 1, pp. 14–22, 2007.

[59] M.-J. Rodrigo, B. Alquezar, and L. Zacarías, "Cloning and characterization of two 9-*cis*-epoxycarotenoid dioxygenase genes, differentially regulated during fruit maturation and under stress conditions, from orange (*Citrus sinensis* L. Osbeck)," *Journal of Experimental Botany*, vol. 57, no. 3, pp. 633–643, 2006.

[60] J. Agustí, M. Zapater, D. J. Iglesias, M. Cercós, F. R. Tadeo, and M. Talón, "Differential expression of putative 9-*cis*-epoxycarotenoid dioxygenases and abscisic acid accumulation in water stressed vegetative and reproductive tissues of citrus," *Plant Science*, vol. 172, no. 1, pp. 85–94, 2007.

[61] T. Moriguchi, M. Kita, S. Hisada, T. Endo-Inagaki, and M. Omura, "Characterization of gene repertoires at mature stage of citrus fruits through random sequencing and analysis of redundant metallothionein-like genes expressed during fruit development," *Gene*, vol. 211, no. 2, pp. 221–227, 1998.

[62] M. Cercós, G. Soler, D. J. Iglesias, J. Gadea, J. Forment, and M. Talón, "Global analysis of gene expression during development and ripening of citrus fruit flesh. A proposed mechanism for citric acid utilization," *Plant Molecular Biology*, vol. 62, no. 4-5, pp. 513–527, 2006.

[63] A. Komatsu, Y. Takanokura, M. Omura, and T. Akihama, "Cloning and molecular analysis of cDNAs encoding three sucrose phosphate synthase isoforms from a citrus fruit (*Citrus unshiu* Marc.)," *Molecular and General Genetics*, vol. 252, no. 3, pp. 346–351, 1996.

[64] A. Komatsu, T. Moriguchi, K. Koyama, M. Omura, and T. Akihama, "Analysis of sucrose synthase genes in citrus suggests different roles and phylogenetic relationships," *Journal of Experimental Botany*, vol. 53, no. 366, pp. 61–71, 2002.

[65] A. Sadka, E. Dahan, E. Or, M. L. Roose, K. B. Marsh, and L. Cohen, "Comparative analysis of mitochondrial citrate synthase gene structure, transcript level and enzymatic activity in acidless and acid-containing *Citrus* varieties," *Australian Journal of Plant Physiology*, vol. 28, no. 5, pp. 383–390, 2001.

[66] T. Shimada, R. Nakano, V. Shulaev, A. Sadka, and E. Blumwald, "Vacuolar citrate/H^+ symporter of citrus juice cells," *Planta*, vol. 224, no. 2, pp. 472–480, 2006.

[67] P. S. Cotroneo, M. P. Russo, M. Ciuni, G. R. Recupero, and A. R. Lo Piero, "Quantitative real-time reverse transcriptase-

PCR profiling of anthocyanin biosynthetic genes during orange fruit ripening," *Journal of the American Society for Horticultural Science*, vol. 131, no. 4, pp. 537–543, 2006.

[68] A. R. Lo Piero, I. Puglisi, P. Rapisarda, and G. Petrone, "Anthocyanins accumulation and related gene expression in red orange fruit induced by low temperature storage," *Journal of Agricultural and Food Chemistry*, vol. 53, no. 23, pp. 9083–9088, 2005.

[69] F. Wellmann, R. Lukačin, T. Moriguchi, L. Britsch, E. Schiltz, and U. Matern, "Functional expression and mutational analysis of flavonol synthase from *Citrus unshiu*," *European Journal of Biochemistry*, vol. 269, no. 16, pp. 4134–4142, 2002.

[70] T. Moriguchi, M. Kita, S. Hasegawa, and M. Omura, "Molecular approach to citrus flavonoid and limonoid biosynthesis," *Food, Agriculture and Environment*, vol. 1, no. 1, pp. 22–25, 2003.

[71] A. Frydman, O. Weisshaus, M. Bar-Peled, et al., "Citrus fruit bitter flavors: isolation and functional characterization of the gene *Cm1,2RhaT* encoding a 1,2 rhamnosyltransferase, a key enzyme in the biosynthesis of the bitter flavonoids of citrus," *The Plant Journal*, vol. 40, no. 1, pp. 88–100, 2004.

[72] M. Kita, Y. Hirata, T. Moriguchi, et al., "Molecular cloning and characterization of a novel gene encoding limonoid UDP-glucosyltransferase in *Citrus*," *FEBS Letters*, vol. 469, no. 2-3, pp. 173–178, 2000.

[73] J. Lücker, M. K. El Tamer, W. Schwab, et al., "Monoterpene biosynthesis in lemon (*Citrus limon*) cDNA isolation and functional analysis of four monoterpene synthases," *European Journal of Biochemistry*, vol. 269, no. 13, pp. 3160–3171, 2002.

[74] T. Shimada, T. Endo, H. Fujii, et al., "Molecular cloning and functional characterization of four monoterpene synthase genes from *Citrus unshiu* Marc," *Plant Science*, vol. 166, no. 1, pp. 49–58, 2004.

[75] J. Lücker, W. Schwab, B. van Hautum, et al., "Increased and altered fragrance of tobacco plants after metabolic engineering using three monoterpene synthases from lemon," *Plant Physiology*, vol. 134, no. 1, pp. 510–519, 2004.

[76] Y. Yamasaki and K. Akimitsu, "In situ localization of gene transcriptions for monoterpene synthesis in irregular parenchymic cells surrounding the secretory cavities in rough lemon (*Citrus jambhiri*)," *Journal of Plant Physiology*, vol. 164, no. 11, pp. 1436–1448, 2007.

[77] L. Sharon-Asa, M. Shalit, A. Frydman, et al., "Citrus fruit flavor and aroma biosynthesis: isolation, functional characterization, and developmental regulation of *Cstps1*, a key gene in the production of the sesquiterpene aroma compound valencene," *The Plant Journal*, vol. 36, no. 5, pp. 664–674, 2003.

[78] E. Carrera, O. Ruíz-Rivero, M. J. Rodrigo, L. Zacarías, and M. Talón, "Altered volatile composition and expression of genes of phenylpropanoid and isoprenoid biosynthesis in the peel flavedo of fruit from citrus mutants," in *Proceedings of the 8th International Meeting: Biosynthesis and Function of Isoprenoids in Plants, Microorganisms and Parasites (TERPNET '07)*, Strasbourg, France, April-May 2007.

[79] S. Ishikawa, T. Shimizu, H. Fujii, et al., "Microarray profiling of gene expression in mutant cultivar of Satsuma mandarin during fruit ripening development," in *Plant & Animal Genome XV Conference*, p. 719, San Diego, Calif, USA, January 2007.

[80] M. Kato, Y. Ikoma, H. Matsumoto, M. Sugiura, H. Hyodo, and M. Yano, "Accumulation of carotenoids and expression of carotenoid biosynthetic genes during maturation in citrus fruit," *Plant Physiology*, vol. 134, no. 2, pp. 824–837, 2004.

[81] M.-J. Rodrigo, J. F. Marcos, and L. Zacarías, "Biochemical and molecular analysis of carotenoid biosynthesis in flavedo of orange (*Citrus sinensis* L.) during fruit development and maturation," *Journal of Agricultural and Food Chemistry*, vol. 52, no. 22, pp. 6724–6731, 2004.

[82] D. Jacob-Wilk, D. Holland, E. E. Goldschmidt, J. Riov, and Y. Eyal, "Chlorophyll breakdown by chlorophyllase: isolation and functional expression of the Chlase1 gene from ethylene-treated *Citrus* fruit and its regulation during development," *The Plant Journal*, vol. 20, no. 6, pp. 653–661, 1999.

[83] S. Harpaz-Saad, T. Azoulay, T. Arazi, et al., "Chlorophyllase is a rate-limiting enzyme in chlorophyll catabolism and is posttranslationally regulated," *Plant Cell*, vol. 19, no. 3, pp. 1007–1022, 2007.

[84] E. Alos, J. Agustí, J. Brumos, et al., "Transcriptomic analyses of natural and induced mutants of citrus reveal key genes limiting physiological processes of major relevance to the citrus industry," in *Proceedings of the 5th Plant Genomics European Meetings*, Venice, Italy, October 2006.

[85] C. J. Nairn, D. J. Lewandowski, and J. K. Burns, "Genetics and expression of two pectinesterase genes in Valencia orange," *Physiologia Plantarum*, vol. 102, no. 2, pp. 226–235, 1998.

[86] T. M. I. E. Christensen, J. E. Nielsen, J. D. Kreiberg, P. Rasmussen, and J. D. Mikkelsen, "Pectin methyl esterase from orange fruit: characterization and localization by in-situ hybridization and immunohistochemistry," *Planta*, vol. 206, no. 4, pp. 493–503, 1998.

[87] A. Gómez-Cadenas, F. R. Tadeo, M. Talón, and E. Primo-Millo, "Leaf abscission induced by ethylene in water-stressed intact seedlings of Cleopatra mandarin requires previous abscisic acid accumulation in roots," *Plant Physiology*, vol. 112, no. 1, pp. 401–408, 1996.

[88] A. Gómez-Cadenas, J. Mehouachi, F. R. Tadeo, E. Primo-Millo, and M. Talón, "Hormonal regulation of fruitlet abscission induced by carbohydrate shortage in citrus," *Planta*, vol. 210, no. 4, pp. 636–643, 2000.

[89] J. K. Burns, D. J. Lewandowski, C. J. Nairn, and G. E. Brown, "Endo-1,4-β-glucanase gene expression and cell wall hydrolase activities during abscission in Valencia orange," *Physiologia Plantarum*, vol. 102, no. 2, pp. 217–225, 1998.

[90] Z. Wu and J. K. Burns, "A β-galactosidase gene is expressed during mature fruit abscission of 'Valencia' orange (*Citrus sinensis*)," *Journal of Experimental Botany*, vol. 55, no. 402, pp. 1483–1490, 2004.

[91] R. Yuan, Z. Wu, I. A. Kostenyuk, and J. K. Burns, "G-protein-coupled α_{2A}-adrenoreceptor agonists differentially alter citrus leaf and fruit abscission by affecting expression of ACC synthase and ACC oxidase," *Journal of Experimental Botany*, vol. 56, no. 417, pp. 1867–1875, 2005.

[92] C. Y. Li, D. Weiss, and E. E. Goldschmidt, "Effects of carbohydrate starvation on gene expression in citrus root," *Planta*, vol. 217, no. 1, pp. 11–20, 2003.

[93] S. Tsukuda, K. Gomi, H. Yamamoto, and K. Akimitsu, "Characterization of cDNAs encoding two distinct miraculin-like proteins and stress-related modulation of the corresponding mRNAs in *Citrus jambhiri* Lush," *Plant Molecular Biology*, vol. 60, no. 1, pp. 125–136, 2006.

[94] K. A. Lahey, R. Yuan, J. K. Burns, P. P. Ueng, L. W. Timmer, and K.-R. Chung, "Induction of phytohormones and differential gene expression in citrus flowers infected

by the fungus *Colletotrichum acutatum*," *Molecular Plant-Microbe Interactions*, vol. 17, no. 12, pp. 1394–1401, 2004.

[95] M. Gandía, A. Conesa, G. Ancillo, et al., "Transcriptional response of *Citrus aurantifolia* to infection by *Citrus tristeza virus*," *Virology*, vol. 367, no. 2, pp. 298–306, 2007.

[96] R. Romero-Aranda, J. L. Moya, F. R. Tadeo, F. Legaz, E. Primo-Millo, and M. Talon, "Physiological and anatomical disturbances induced by chloride salts in sensitive and tolerant citrus: beneficial and detrimental effects of cations," *Plant, Cell & Environment*, vol. 21, no. 12, pp. 1243–1253, 1998.

[97] J. L. Moya, A. Gómez-Cadenas, E. Primo-Millo, and M. Talón, "Chloride absorption in salt-sensitive Carrizo citrange and salt-tolerant Cleopatra mandarin citrus rootstocks is linked to water use," *Journal of Experimental Botany*, vol. 54, no. 383, pp. 825–833, 2003.

[98] T. Beeor-Tzahar, G. Ben-Hayyim, D. Holland, Z. Faltin, and Y. Eshdat, "A stress-associated citrus protein is a distinct plant phospholipid hydroperoxide glutathione peroxidase," *FEBS Letters*, vol. 366, no. 2-3, pp. 151–155, 1995.

[99] A. R. Lo Piero, I. Puglisi, and G. Petrone, "Gene isolation, analysis of expression, and in vitro synthesis of glutathione *S*-transferase from orange fruit [*Citrus sinensis* L. (Osbeck)]," *Journal of Agricultural and Food Chemistry*, vol. 54, no. 24, pp. 9227–9233, 2006.

[100] M. Şahin-Çevik and G. A. Moore, "Identification and expression analysis of cold-regulated genes from the cold-hardy *Citrus* relative *Poncirus trifoliata* (L.) Raf," *Plant Molecular Biology*, vol. 62, no. 1-2, pp. 83–97, 2006.

[101] C.-K. Zhang, P. Lang, F. Dane, et al., "Cold acclimation induced genes of trifoliate orange (*Poncirus trifoliata*)," *Plant Cell Reports*, vol. 23, no. 10-11, pp. 764–769, 2005.

[102] Q. Y. Cai, G. A. Moore, and C. L. Guy, "An unusual group 2 LEA gene family in citrus responsive to low temperature," *Plant Molecular Biology*, vol. 29, no. 1, pp. 11–23, 1995.

[103] R. Porat, D. Pavoncello, S. Lurie, and T. G. McCollum, "Identification of a grapefruit cDNA belonging to a unique class of citrus dehydrins and characterization of its expression patterns under temperature stress conditions," *Physiologia Plantarum*, vol. 115, no. 4, pp. 598–603, 2002.

[104] M. T. Sanchez-Ballesta, Y. Lluch, M. J. Gosalbes, L. Zacarias, A. Granell, and M. T. Lafuente, "A survey of genes differentially expressed during long-term heat-induced chilling tolerance in citrus fruit," *Planta*, vol. 218, no. 1, pp. 65–70, 2003.

[105] M. Sapitnitskaya, P. Maul, G. T. McCollum, et al., "Postharvest heat and conditioning treatments activate different molecular responses and reduce chilling injuries in grapefruit," *Journal of Experimental Botany*, vol. 57, no. 12, pp. 2943–2953, 2006.

[106] L. Peña, M. Cervera, C. Fagoaga, et al., "Citrus," in *A Compendium of Transgenic Crop Plants*, C. Kole and T. C. Hall, Eds., vol. 5, Wiley-Blackwell, Malden, Mass, USA, 2008.

[107] M. Cervera, J. Ju árez, A. Navarro, et al., "Genetic transformation and regeneration of mature tissues of woody fruit plants bypassing the juvenile stage," *Transgenic Research*, vol. 7, no. 1, pp. 51–59, 1998.

[108] M. Talón, L. Zacarias, and E. Primo-Millo, "Gibberellins and parthenocarpic ability in developing ovaries of seedless mandarins," *Plant Physiology*, vol. 99, no. 4, pp. 1575–1581, 1992.

[109] A. M. Vidal, C. Gisbert, M. Talón, E. Primo-Millo, I. López-Díaz, and J. L. García Martínez, "The ectopic overexpression of a citrus gibberellin 20-oxidase alters the gibberellin content and induces an elongated phenotype in tobacco," *Physiologia Plantarum*, vol. 112, no. 2, pp. 251–260, 2001.

[110] C. Fagoaga, F. R. Tadeo, D. J. Iglesias, et al., "Engineering of gibberellin levels in citrus by sense and antisense overexpression of a GA 20-oxidase gene modifies plant architecture," *Journal of Experimental Botany*, vol. 58, no. 6, pp. 1407–1420, 2007.

[111] M. Talón, F. R. Tadeo, and J. A. D. Zeevaart, "Cellular changes induced by exogenous and endogenous gibberellins in shoot tips of the long-day plant *Silene armeria*," *Planta*, vol. 185, no. 4, pp. 487–493, 1991.

[112] W. Guo, Y. Duan, O. Olivares-Fuster, et al., "Protoplast transformation and regeneration of transgenic Valencia sweet orange plants containing a juice quality-related pectin methylesterase gene," *Plant Cell Reports*, vol. 24, no. 8, pp. 482–486, 2005.

[113] L. Peña, M. Martín-Trillo, J. Juárez, J. A. Pina, L. Navarro, and J. M. Martínez-Zapater, "Constitutive expression of *Arabidopsis LEAFY or APETALA1* genes in citrus reduces their generation time," *Nature Biotechnology*, vol. 19, no. 3, pp. 263–267, 2001.

[114] L. J. Pillitteri, C. J. Lovatt, and L. L. Walling, "Isolation and characterization of a *TERMINAL FLOWER* homolog and its correlation with juvenility in citrus," *Plant Physiology*, vol. 135, no. 3, pp. 1540–1551, 2004.

[115] T. Endo, T. Shimada, H. Fujii, Y. Kobayashi, T. Araki, and M. Omura, "Ectopic expression of an *FT* homolog from *Citrus* confers an early flowering phenotype on trifoliate orange (*Poncirus trifoliata* L. Raf.)," *Transgenic Research*, vol. 14, no. 5, pp. 703–712, 2005.

[116] H. B. C. Molinari, C. J. Marur, J. C. B. Filho, et al., "Osmotic adjustment in transgenic citrus rootstock Carrizo citrange (*Citrus sinensis* Osb. x *Poncirus trifoliata* L. Raf.) overproducing proline," *Plant Science*, vol. 167, no. 6, pp. 1375–1381, 2004.

[117] C. Fagoaga, C. López, A. Hermoso de Mendoza, et al., "Post-transcriptional gene silencing of the p23 silencing suppressor of *Citrus tristeza virus* confers resistance to the virus in transgenic Mexican lime," *Plant Molecular Biology*, vol. 60, no. 2, pp. 153–165, 2006.

[118] A. Domínguez, A. Hermoso de Mendoza, J. Guerri, et al., "Pathogen-derived resistance to *Citrus tristeza virus* (CTV) in transgenic Mexican lime (*Citrus aurantifolia* (Christ.) Swing.) plants expressing its *p*25 coat protein gene," *Molecular Breeding*, vol. 10, no. 1-2, pp. 1–10, 2002.

[119] M. C. Vives, L. Galipienso, L. Navarro, P. Moreno, and J. Guerri, "The nucleotide sequence and genomic organization of *Citrus leaf blotch virus*: candidate type species for a new virus genus," *Virology*, vol. 287, no. 1, pp. 225–233, 2001.

[120] E. Katz, M. Fon, Y. J. Lee, B. S. Phinney, A. Sadka, and E. Blumwald, "The citrus fruit proteome: insights into citrus fruit metabolism," *Planta*, vol. 226, no. 4, pp. 989–1005, 2007.

[121] I. Lliso, F. R. Tadeo, B. S. Phinney, C. G. Wilkerson, and M. Talón, "Protein changes in the albedo of citrus fruits on postharvesting storage," *Journal of Agricultural and Food Chemistry*, vol. 55, no. 22, pp. 9047–9053, 2007.