Cereal Genomics

Cereal Genomics

edited by

P.K. Gupta
Ch Charan Singh University,
Meerut, India

and

R.K. Varshney
Institute of Plant Genetics and Crop Plant Research (IPK),
Gatersleben, Germany

KLUWER ACADEMIC PUBLISHERS
NEW YORK, BOSTON, DORDRECHT, LONDON, MOSCOW

eBook ISBN: 1-4020-2359-6
Print ISBN: 1-4020-2358-8

Created in the United States of America

Visit Springer's eBookstore at: http://ebooks.springerlink.com
and the Springer Global Website Online at: http://www.springeronline.com

Table of Contents

PREFACE

Cereals make an important component of daily diet of a major section of human population, so that their survival mainly depends on the cereal grain production, which should match the burgeoning human population. Due to painstaking efforts of plant breeders and geneticists, at the global level, cereal production in the past witnessed a steady growth. However, the cereal production in the past has been achieved through the use of high yielding varieties, which have a heavy demand of inputs in the form of chemical fertilizers, herbicides and insecticides/pesticides, leading to environmental degradation. In view of this, while increasing cereal production, one also needs to keep in mind that agronomic practices used for realizing high productivity do not adversely affect the environment. Improvement in cereal production in the past was also achieved through the use of alien genetic variation available in the wild relatives of these cereals, so that conservation and sustainable use of genetic resources is another important area, which is currently receiving the attention of plant breeders.

The work leading to increased cereal production in the past received strong support from basic research on understanding the cereal genomes, which need to be manipulated to yield more from low inputs without any adverse effects as above. Through these basic studies, it also became fairly apparent that the genomes of all cereals are related and were derived from the same lineage, million of years ago. The availability of molecular markers and the tools of genomics research have provided unprecedented power for dissecting these genomes further for resolving details, which could not be examined earlier due to low resolving power of the then available techniques and approaches. Genomics is an area of research, which has undergone exponential growth during the last more than one decade, and has led to generation of knowledge at an accelerated pace. In each cereal, this has led to the development of large repertoires of molecular markers, construction of high density molecular maps (both genetic and physical maps), generation of large number of expressed sequence tags (ESTs), and sequencing of regions carrying specific genes. This has also demonstrated that there is synteny/collinearity among cereal genomes, although rearrangements of DNA segments are frequent. Rice genome has also been sequenced, and a comparison of ESTs of wheat with rice genome sequences suggested that perhaps rice genome sequence would not always and under all conditions prove useful to predict and discover genes in other cereals. This has led to the realization that gene-rich regions of other cereals will also have to be sequenced to bring about the desired improvement in all these cereals. Identification and cloning of QTLs including those involved in epistatic and environmental interactions, or those involved in controlling the level of expression of genes (eQTLs) is another

area, which is receiving increased attention of geneticists and statisticians world over. These approaches are supposed to help plant breeding through both, the marker assisted selection (MAS) and transgenic cereals.

The data and knowledge generated through cereal genomics research during the last more than a decade is enormous, so that entire issues of journals like Plant Molecular Biology, Plant Cell, Plant Physiology, and Functional and Integrative Genomics (F&IG) have been devoted to cereal genomics. However, on this subject, no book was available so far, where the students, teachers and young research workers could find all information on cereal genomics at one place. The present book hopefully should fill this gap and may prove useful not only for teaching, but also in the initial years of research career of students working in the area of cereal genomics.

The editors are grateful to the authors of different chapters (see Appendix I), who not only reviewed the published research work in their area of expertise but also shared their unpublished results to make the articles up-to-date. We also appreciate their cooperation in meeting the deadlines and in revising their manuscripts, whenever required. While editing this book, the editors also received strong support from many reviewers (see Appendfix II), who willingly reviewed the manuscripts for their love for the science of cereal genomics, and gave useful suggestions for improvement of the manuscripts. However, the editors owe the responsibility for any errors that might have crept in inadvertently during the editorial work.

The book was edited during the tenure of PKG as UGC Emeritus Fellow/INSA Senior Scientist at CCS University, Meerut (India) and that of RKV as a Post-doctoral Research Scientist at IPK, Gatersleben (Germany). The cooperation and help received from Noeline Gibson and Jacco Flipsen of Kluwer Academic Publishers during various stages of the development and completion of this project is duly acknowledged. Sachin, Pawan and Shailendra helped PKG in the editorial work. The editors also recognize that the editorial work for this book has been quite demanding and snatched away from them some of the precious moments, which they should have spent together with their respective families.

The editors hope that the book will prove useful for the targeted audience and that the errors, omissions and suggestions, if any, will be brought to their notice, so that a future revised and updated edition, if planned, may prove more useful.

P. K. Gupta
R.K. Varshney

Chapter 1

CEREAL GENOMICS: AN OVERVIEW

Pushpendra K. Gupta[1,*] and Rajeev K. Varshney[2]

[1]*Ch Charan Singh University, Meerut- 250004, U.P., India;* [2]*Institute of Plant Genetics and Crop Plant Research (IPK), Correnssstrasse 3, D-06466 Gatersleben, Germany.*
[*]Author for correspondence: pkgupta36@vsnl.com

1. INTRODUCTION

Cereals are widely cultivated and produce annually, 1800 to 1900 million tonnes of food grains worldwide (see Table 1; FAO website http://apps.fao.org). Cereals also represent 60% of the calories and proteins consumed by human beings. They include a variety of crops including rice, maize, wheat, oats, barley, rye, etc., but excluding millets like pearl millet and other minor millets. In the past, cereals have been a subject of intensive cytogenetic investigations that are now extended further in the genomics era using powerful tools of molecular biology. The progress in cereal genomics research during the last two decades, involving the use of molecular markers for a variety of purposes, and the whole genome sequencing in rice has been remarkable indeed. The results of genomics research resolved many aspects, which the conventional cytogenetics failed to resolve. For instance, this involved initially the preparation of molecular maps, which were utilized extensively for comparative genomics and cytogenomics. Cereal genomes have also been subjected to both structural and functional genomics research, which during the last two decades covered both basic and applied aspects. As a result, not only we understand better the genomes of major cereals and the mechanisms involved in the function of different cereal genes, but we have also utilized information generated from genomics research in producing better transgenic crops, which will give higher yields, sometimes with value addition. Transgenic cereals will also be available, which will be resistant to major pests and diseases and will be adapted to changing environmental conditions. In this book, while we have tried to cover for cereal crops, the areas involving molecular markers, whole genome sequencing (WGS) and

P.K. Gupta and R.K. Varshney (eds.), Cereal Genomics, 1-18.

Table 1. Some details about important cereals, discussed in the book

Cereal species	Biological Name	FAO data (2002)[1]		Chromsome number	Genome size (Mbp)[2]	Availability of resources	Number of ESTs available in public domain[3]
		Yield (Tonnes/ Hectare)	Production (Million Tonnes)				
Barley	*Hordeum vulgare*	2.5	132	2*n*=2*x*=14	5,000	Both genetic and physical maps (based on translocation breakpoints) available	356,848
Maize	*Zea mays*	4.3	603	2*n*=2*x*=20	2,500	Extensive genetic (including transcript map) and BAC-based physical maps available	393,719
Oats	*Avena sativa*	1.9	27	2*n*=6*x*=42	11,400	Genetic maps (but not saturated) available	574
Rice	*Oryza sativa*	3.9	580	2*n*=2*x*=24	430	Extensive genetic (including transcript) and YAC/BAC-based physical maps, as well as 4 drafts of complete genome sequences available	283,935
Rye	*Secale cereale*	2.2	21	2*n*=2*x*=14	8,400	Genetic maps (but not saturated) available	9,194
Sorghum	*Sorghum bicolor*	1.3	55	2*n*=2*x*=20	750	Integrated cytogenteic, genetic and physical maps avaialble	161,813
Wheat	*Triticum aestivum*	2.7	568	2*n*=6*x*=42	16,000	Extensive genetic as well as delition lines-based physical maps available	549,926

[1]As per FAO wesbsite http://apps.fao.org; [2]As per Bennett and Leitch (1995); [3]As per dbEST release 030504- http://www.ncbi.nlm.nih.gov/dbEST/dbEST summary.html

the functional genomics, by choice we have not included the transgenics, which involve only an application of genomics research and do not make a part of genomics research itself. We realize that cereal genomics has been the subject of several special issues or sections of journals like PNAS, USA (Volume 95 issue no.5, March 1998), Plant Molecular Biology (Volume 48 issue no. 1-2, January 2002), Plant Physiology (Volume 125 issue no. 4, March 2001; Volume 130 issue no. 4, December 2002), Functional & Integrative Genomics (Volume 3 issue no. 1-2, March 2003; Volume 4 issue no.1, March 2004), etc. Several review articles also adequately cover the present status of research on this important subject (Gale and Devos, 1998; Goff, 1999; Bennetzen, 2000; Devos and Gale, 2000; Wise, 2000; Edwards and Stevenson, 2001; Lagudah *et al.,* 2001; Feuillet and Keller, 2002; Rafalski, 2002; Appels *et al.*, 2003; Bennetzen and Ma, 2003; Paterson *et al.*, 2003, etc). However, the subject has not been presented elsewhere in the form of a book that may be used for teaching and research, and hence the present effort of this book on cereal genomics.

2. MOLECULAR MARKERS: DEVELOPMENT AND USE IN GENOMICS RESEARCH

2.1. Construction of Molecular Maps

A variety of molecular markers is now known and their development and use has been a subject of intensive research in all cereals. The subject has been adequately covered in several recent reviews (Kumar, 1999; Koebner *et al.*, 2001; Gupta *et al.*, 2002), but none of these reviews adequately covers all aspects. In the present book, 12 of the 20 chapters are devoted to the discussions on different types of molecular markers and their uses. The availability of molecular marker maps of cereal genomes, with varying density and resolution, has also facilitated comparative genomics studies. The group led by Mike Gale at John Innes Center (JIC), Norwich (UK), is regularly updating these results. Availability of a number of marker assays provides ample opportunity for exercising choice of a suitable marker system based on intended objective, convenience and costs. In this book, the different molecular marker systems and their evaluation in cereal genomics is being discussed in Chapter 2 by Daryl Somers from Winipeg (Canada). Rajeev Varshney and Andreas Boerner from Gatersleben and Viktor Korzun from Bergen have covered the methods for the preparation of molecular maps and the progress made so far in preparation of these maps in Chapter 3. It will be noticed that while significant progress has been made in the preparation of molecular maps in cereals, most of these maps are based on

RFLPs, the first molecular marker system that became available in early 1980s (for details see Philips and Vasil, 1994, 2001). Progress has also been made in the preparation of microsatellite maps. Although in the past it has been expensive and cumbersome to generate microsatellites, they have now been generated in almost all cereal species (Gupta and Varshney, 2000). For instance, a microsatellite map of maize genome with 900 SSR markers became available recently (Sharopova *et al.*, 2002; total ~1800 SSR mapped loci available as on October 2001; http://www.agron.missouri.edu/ssr.html), and another integrated microsatellite map with ~1000 SSR loci is now available in bread wheat (D. Somers, Canada, personal communication). Moreover, in recent years a large amount of sequence data has been/is being generated in many cereals from several genome sequencing and EST sequencing projects (see Table 1). Therefore the available sequence data is also being exploited for development of microsatellite markers (Kantety *et al.*, 2002; Varshney *et al.*, 2002; Gao *et al.*, 2003). In this direction, more than 2000 SSR loci have already been mapped in rice (McCouch *et al.*, 2002) and work is in progress in other cereals like barley (Varshney *et al.*, unpublished) and wheat (M. Sorrells, USA, personal communication). Availability of sequence data also provided the next generation of markers, i.e. single nucleotide polymorphism (SNP, popularly pronounced as snip). However, SNP discovery is costly, so that the construction of SNP maps in cereals will take time. This is in sharp contrast to the situation in human genome, where a large number of microsatellites and at least two million SNPs have already been mapped. A programme for the preparation of a HapMap for the human genome, utilizing haplotypes, based on SNPs, was also initiated in late 2002 under the SNP Consortium (http://snp.cshl.org/; The International HapMap Consortium, 2003). Among cereals, SNPs are now being discovered and will be extensively used in future for genotyping in crops like barley, rice, maize and wheat. EST-based SSRs, SNPs, or RFLPs, are also being used for construction of 'functional or transcript maps' in many cereals like rice (Kurata *et al.*, 1994; Harushima *et al.*, 1998; Wu *et al.*, 2002), maize (Davis *et al.*, 1999), barley (A. Graner, Germany, personal communication). EST–based markers are also being used, *firstly*, to assess functional diversity, and *secondly*, to anchor genic regions in BAC/YAC-based physical maps, etc. Physical maps have also been prepared in some cereals like rice (Tao *et al.*, 2001; Chen *et al.*, 2002), maize (Yim *et al.*, 2002; http://www.agron. missouri.edu/maps.html), sorghum (Klein *et al.*, 2000; Childs *et al.*, 2001; http://sorghumgenome.tamu.edu/), etc. However, construction of BAC/YAC-based physical maps is difficult in large and complex genomes like those of wheat and barley. Therefore, for physical mapping, translocation breakpoints (aneuploid stocks) were used in barley (Künzel *et al.*, 2002) while deletion lines (another aneuploid stocks)- are being used in wheat (B.S. Gill, USA, personal communication;

http://wheat.pw.usda.gov/NSF/progress_mapping. html). At SCRI in UK, efforts are underway for prepartion of local physical maps in barley by using radiation hybrids (RH) or HAPPY mapping (Waugh *et al.*, 2002; Thangavelu *et al.*, 2003).

Physical maps are important for chromosome walking in map-based cloning projects (see later) and have also facilitated identification of gene rich regions in large genomes, which are too big to be used for whole genome sequencing. These gene rich regions will be the target for genome sequencing in these crops e.g. wheat (see later) where BAC libraries are already available. This aspect of gene distribution, gene density and gene islands has been discussed by Kulvinder Gill, Washington (USA) in Chapter 12. It has been shown that cereal genomes possess gene-rich and gene-poor regions (Akhunov *et al.*, 2003). Moreover physical location, structural organization and gene densities of the gene-rich regions are similar across the cereal genomes (Feuillet and Keller, 1999; Sandhu and Gill, 2002). Despite this, cereal genomes greatly vary in their size, which is attributed to the presence of varying amounts of repetitive sequences in these genomes (Heslop-Harrison, 2000). Repetitive sequences can be found in the genome either in tandem arrays or in a dispersed fashion. Therefore repetitive sequences can be classified into 3 categories: (i) transposable including retrotransposon elements, which are mobile genetic elements; (ii) microsatellite sequences, which are tandemly repeated DNA sequences (also called simple sequence repeats, SSRs); and (iii) special classes such as telomeric/ centromeric sequences or rDNA units, etc. First two classes constitute a major proportion of the repetitive sequences, present in the genome. During the last decade, a number of studies have been carried out to study physical organization of retrotransposons in several plant genomes including cereals (for references see Kumar and Bennetzen, 1999; Wicker *et al.*, 2002). The fraction of the genome contributed by retrotransposons increases with genome size from rice, the smallest cereal genome (430 Mb, ~14% LTR retrotransposons, Tarchini *et al.*, 2000), through maize (2500 Mb, 50-80% LTR retrotransposons, SanMiguel *et al.*, 1996) to barley (5000 Mb, >70% LTR retrotransposons, Vicient *et al.*, 1999). Contribution and organization of retrotransposons and microsatellites in cereal genomes has been discussed by Alan Schulman from Helsinki (Finland) and the editors of this book in Chapter 4.

2.2. Comparative Genomics

The availability of molecular maps for all major cereal genomes facilitated studies on comparative mapping, where rice genome was used as the anchor

genome and all cereal genomes could be expressed in the form of -30 odd genomic blocks derived from the rice genome (Moore *et al.*, 1995; Gale and Devos, 1998). These studies also led to the conclusion that maize genome really consists of two genomes of five chromosomes each and is therefore an archaic tetraploid (Devos and Gale, 2000). Comparative studies revealed a good conservation of markers within large chromosomal segments of the cereal genomes**.** However, mapping of resistance gene analogs (RGAs) isolated from rice, barley and foxtail millet showed limited orthology (Leister *et al.*, 1999). In recent years, isolation and sequencing of large genomic DNA fragments (100-500 kb) from many cereals provided further insights about the conservation of gene order between different cereal genomes at the sub-megabase level i.e. micro-colinearity (Chen *et al.*, 1997, 1998; Dubcovsky *et al.*, 2001), although evidence for disruption of this collinearity is also available in some cases (Tikhonov *et al.*, 1999; Tarachini *et al.*, 2000; Li and Gill, 2002). The comparative genomic studies also resolved a large number of duplications, translocations, and inversions that accompanied the evolution of these cereal genomes and could not be earlier resolved by conventional tools of cytogenetics research (Paterson *et al.*, 2000; Benntezen, 2000; Bennetzen and Ma, 2003). Thus, the detailed studies on comparative genomics of cereals facilitated researchers to utilize information generated from one cereal genome for that of the other (see Bellgard *et al.*, 2004). These aspects related to comparative mapping and genomics have been covered by Andrew Paterson from Athens (USA) in Chapter 5.

2.3. QTL Analysis: Its Use in Study of Population Structure and for Crop Improvement

Molecular markers have also been used for the study of population structures in the progenitors of our major cereals, which has been discussed by Eviatar Nevo, Haifa (Israel) in Chapter 6. Another important area of genomics research is QTL analysis (including QTL interval mapping), which has been utilized for mapping of QTL for a variety of economic traits and for developing markers that are closely associated with these QTLs in different cereal crops. One of the offshoots of this research is also the identification of traits and the associated QTL, which led to the domestication of cereals (Paterson *et al.*, 1995; Heun *et al.*, 1997; Badr *et al.*, 2000; Ozkan *et al.*, 2002; Salamini *et al.*, 2002; Peng *et al.*, 2003). This aspect has been discussed in Chapter 7 by Francesco Salamini and his colleagues from Cologne (Germany), Lodi and Milan (Italy). One of the major benefits of QTL analysis would also be to identify QTL for resistance against biotic and abiotic stresses and the associated molecular markers (e.g. Ordon *et al.*,

1998; Friedt *et al.*, 2003; Ribaut *et al.*, 2002), so that this information will be utilized in future either for the marker-aided selection or for the isolation of these QTL through map-based cloning. Exploitation of molecular markers for identification of genes/QTLs for disease resitance in barley and wheat has been discussed in Chapter 8 by Ahmed Jahoor and his colleagues from Roskilde and Horsens (Denmark). Similarly, identification of genes/QTLs conferring tolerance to abiotic stresses have been discussed in Chapter 9 by Roberto Tuberosa and Silvio Salvi from Bologna (Italy).

2.4. Marker -Assisted Selection (MAS) and Map-Based Cloning (MBC)

Marker-assisted selection (MAS) is a powerful tool for indirect selection of difficult traits at the seedling stage during plant breeding, thus speeding up the process of conventional plant breeding and facilitating the improvement of difficult traits that can not be improved upon easily by the conventional methods of plant breeding (Ribaut and Hoisington, 1998). It has been realized that despite extensive research in this area, MAS has not been put to practice in actual plant breeding, to the extent earlier anticipated (see http://www.fao.org/biotech/logs/c10logs.htm). The reasons for lack of activity involving MAS, and the future possibilities of using MAS in wheat and barley breeding have recently been discussed in some reviews (Koebner *et al.*, 2001; Koebner and Summers, 2003; Thomas, 2002, 2003), etc. It has however been recognized that with the availability of a large repertoire of SSR markers in majority of cereals and with the developments of SNPs at an accelerated pace, MAS will be effectively used in future to supplement the conventional plant breeding. In Chapter 10 of this book, Robert Koebner from Norwich (UK) is dealing with the present status and future prospects of MAS in cereals.

Production of transgenic cereals requires isolation of important genes for agronomic traits (including those for resistance to diseases). Identification of closely linked markers with such genes provides the starting point for map-based cloning (MBC) of these genes. There are three major requirements for map-based gene isolation (Ordon *et al.*, 2000; Wise, 2000): (i) a high resolution genetic map spanning the gene of interest, (ii) availability of a large-insert genomic YAC or BAC library, (iii) multiple independent mutant stocks and (iv) an efficient transformation system for use in functional complementation. All these resources either have become available or their generation is in progress in almost all the cereals (Table 1). Some of the resistance genes that have been isolated include the following, i.e. *Mlo*

(Büschges *et al.*, 1997), *Mla* (Wei *et al.*, 1999), *Rar1* (Lahaye *et al.*, 1998; Shirasu *et al.*, 1999), *Rpg1* (Hovrath *et al.*, 2003) in barley; *xa-21* (Song *et al.*, 1995), *xa-1* (Yoshimura *et al.*, 1998), *PiB* (Wang *et al.*, 1999); *Pi-ta* (Bryan *et al.*, 2000) in rice; and *Lr10* (Stein *et al.*, 2000; Feuillet *et al.*, 2003), *Lr21* (Huang *et al.*, 2003) and *Pm3* (Yahiaoui *et al.*, 2003) in wheat. In rice, progress is underway to isolate *xa-5* (Blair *et al.*, 2003), genes required from rice yellow mosaic virus (RYMV movement) (Albar *et al.*, 2003). In recent years some studies have targeted cloning of QTLs also by using MBC approach (see Yano, 2001). For instance, a *fruit weight2.2* (*fw2.2*) QTL in tomato (Fraray *et al.*, 2000) and two major photoperiod sensitivity QTLs, *Hd1* (Yano *et al.*, 2000) and *Hd6* (Takahashi *et al.*, 2001) in rice have recently been isolated. As an example for QTL isolation in cereals other than rice, *Vgt1* responsible for transition from vegetative to the reproductive phase in maize is looked upon as one possible target (Salvi *et al.*, 2002). With the availability of resources and expertise, developed recently, it is expected to isolate some important QTLs in other cereals also in the near future. In Chapter 11, Nils Stein and Andreas Graner from Gatersleben (Germany) have discussed methodology and progress in the area of MBC. Gene tagging using transposon induced mutant populations together with cDNA approaches are also gaining importance in some cereals like maize (Bensen *et al.*, 1995; Das and Martienssen, 1995; for review see Osborne and Baker, 1995 and Walbot, 2000), rice (Hirochika, 1997; Izawa *et al.*, 1997; Zhu Z.G. *et al.*, 2003), barley (Scholz *et al.*, 2001), etc.

3. LARGE- SCALE GENOME/ TRANSCRIPTOME SEQUENCING AND ITS UTILIZATION

3.1. Methods and Progress of Whole Genome Sequencing (WGS) in Cereals

Among higher plants, Arabidopsis genome is the first to be fully sequenced (TAGI, 2000). However, during the last two years (2001-2003) four drafts of rice genome sequences have also become available and are being extensively utilized for a variety of purposes (Barry, 2001; Goff *et al.*, 2002; Yu *et al.*, 2002; IRGSP- http://rgp.dna.affrc.go.jp/IRGSP/). The high quality sequences for three rice chromosomes 1, 4 and 10 have also been completed and published (Sasaki *et al.* 2002; Feng *et al.* 2002; The Rice Chromosome 10 Sequencing Consortium, 2003). As we know, two different approaches are available for whole genome sequencing, one of them involving preparation of physical maps of BACs first, and then sequence the genome BAC-by BAC, and the other involving whole genome shotgun (WGS) approach

pioneered by James C. Venter, Rockville (USA) for the human genome (Venter *et al.*, 1998). Both these approaches have been utilized for whole genome sequencing of rice genome. In Chapter 13, Yeisoo Yu and Rod Wing from Tucson (USA) presents a detailed account on the methodology and the progress made in this direction. After completion of sequencing of Arabidopsis and rice genomes, efforts are underway to characterize and annotate all the genes in these two genomes. Information generated from these genomes will continue to prove very useful in different aspects of genomics research in other cereals. Takuji Sasaki and Baltazar Antonio from Tsukuba (Japan) have discussed the rice genome as a model system for cereals in Chapter 18. Similarly, in Chapter 17, Klaus Mayer and his colleagues from MIPS, Neuherberg (Germany) have discussed the Arabidopsis genome and its use in cereal genomics. However, some cereals like barley have a unique property of malting, which makes it different from other cereals. This suggests that at least for this particular trait, barley genome has genes, which are absent in Arabidopsis and rice genomes. Similar unique genes for some other traits may be available in other cereals also. Therefore, several large-scale EST sequencing projects were initiated in barley, wheat, sorghum, maize, etc. and as a result large amount of data has been generated (Table 1). Due to importance of maize as a cereal crop, the Maize Genome Sequencing Project is also underway (Chandel and Brendel, 2002; http://www.maizegenome.org/). It is believed that the maize genome sequence will be useful for annotation of rice genome in the same manner as mouse genome proved useful for the human genomes (Gregory *et al.*, 2002). However, barley and wheat are also important cereals, but whole genome sequencing has not been planned for these two cereals due to their large genome size (Table1). However, in an ITMI meeting (Winipeg, Canada, June 1-4, 2002) Bikram Gill and other cereal workers discussed the concept of IGROW (International Genome Research on Wheat) to lead wheat genome sequencing and improvement effort for the next 10 years. IGROW has the following objectives: (i) identification of gene-rich regions by using BAC-library and *Cot*-based procedures, and (ii) shotgun sequencing of the identified gene-rich regions of the wheat genome (http://wheat.pw.usda.gov/ggpages/awn/48/Textfiles/IGROW.html). In November 2003, another meeting of IGROW sponsored by the National Science Foundation and the United States Department of Agriculture was held to discuss the need and a strategy for sequencing the wheat genome consisting of 16,000 Mbp. It was argued that the wheat genome sequence would provide a model for structural and functional changes that accompany polyploidy and that model species cannot be used to study the unique traits in wheat (Gill and Appels, 2004).

3.2 Bioinformatics and its Use in Development and Use of Cereal Databases

Extensive data on all aspects of cereal genomics are now available at GrainGenes (http://wheat.pw.usda.gov/). An independent database Gramene (http://www.gramene.org/) has also been created, which has major emphasis on rice genome and its relationship with other cereal genomes. Other independent databases are also available for some individual cereal crops, like rice (Oryzabase http://www.shigen.nig.ac.jp/rice/oryzabase/) and maize (ZeaDB http://www.zmdb.iastate.edu/; MaizeDB http://www.agron. missouri.edu/). Similar databases have yet to be developed for other cereals like wheat, barley and oats. Tools of bioinformatics are already being used for development and use of these databases for mining useful information. Some efforts are also underway at the University of California, Berkley (USA) to create a database (CereGenDB) for coding sequences that are conserved not only between rice and other cereals, but also between cereals and Arabidopsis. These aspects have been discussed in Chapter 14 by Dave Matthews, Olin Anderson and their colleagues from Ithaca and Albany (USA).

3.3. Functional Genomics and its Utility for Crop Improvement

During the last 5 years, as mentioned above,a large amount of sequence data, has been generated from many genome/EST sequencing projects in cereals (Table 1). Available sequence data are being already utilized for a variety of purposes, including annotation of these genomic sequences. An important area of research in the field of functional genomics in cereals is the study of expression patterns in time and space (see Schena, 1999; Kehoe *et al.*, 1999). These studies are being related with the whole genome sequences with and without known functions and also with the information available about the structure and function of proteins available in the databases. In this direction, mainly cDNA clones corresponding to the ESTs, have been utilized to prepare cDNA macro/micro-arrays (Richmond and Somerville, 2000). Exploitation of microarray technology for gene expression studies in cereals is still in its infancy (see Sreenivasulu *et al.*, 2002b). Nevertheless some progress has been made in identification of genes involved in embryo/seed development, seed germination, grain filling, etc. in some cereals like maize (Lee *et al.*, 2002), barley (Sreenivasulu *et al.*, 2002a; Potokina *et al.*, 2002), rice (Zhu T. *et al.*, 2003). Studies have also been targeted towards genes that are responsible for stress tolerance (Kawasaki *et al.*, 2001; Bohnert *et al.*,

2001; Ozturk *et al.*, 2003). These aspects are discussed in two chapters; Chapter 15 by Peter Langridge and his colleagues from Adelaide (Australia); and Chapter 16 by Nese Sreenivasulu and others from Gatersleben (Germany) and Hyderabad (India). Recently, a new approach, called 'genetical genomics' has also been proposed, where expression profiling of individual genes is combined with QTL mapping in a segregating (mapping) population (Jansen and Nap, 2001; Jansen, 2003). We believe that availability of large EST collections for genome-wide expression profiling and analytical tools available for molecular marker analysis in different cereals will accelerate the use of the 'genetical genomics' approach for identification of genes for different agronomic traits to be used for crop improvement programmes.

4. GENOMICS RESEARCH IN CROP IMPROVEMENT

The current investment on cereal genomics research eventually has to give returns in the form of improved cereal crops. For this purpose, we need to achieve success, both in using molecular marker aided selection (MAS) and in the development of improved transgenic cereals, which would give not only increased yield, but will also give value added cereals with improved nutritional quality. According to most cereal workers, both these expectations of cereal genomics research should be realized in the near future. Availability of complete genome sequence of rice along with the EST-sequencing projects in other cereals provides enough resources to utilize them in the wet lab as well as for *in silico* mining, with an objective to improve crops. Utilization of sequence data in post-genomic era is being discussed in Chapter 19 by Mark Sorrells, Ithaca (USA). Wanlong Li and Bikram Gill from Manhattan (USA) discuss the future role of genomics research for cereal improvement in the concluding chapter of this book (Chapter 20).

5. SUMMARY AND OUTLOOK

Significant progress in the field of cereal genomics has already been made in almost all cereals. For instance, availability of a variety of molecular markers facilitated preparation of high-density maps in almost all cereals. This activity proved useful in identification of molecular markers linked with genes/QTLs for a variety of economic traits including those conferring tolerance to biotic and abiotic stresses. In some cases, molecular markers have also been used for MAS, and for map-based cloning (MBC) of genes. However, in cereals other than rice and maize (eg., wheat and barley), further

research at an accelerated pace is required in both areas i.e. MAS and MBC. Data on genomic and EST sequences from a variety of sources are also being analysed to understand the genomes and the transcriptomes of different cereals. For instance, in the field of functional genomics, availability of ESTs from different parts of a plant in an individual cereal crop gave a momentum to research involving identification and annotation of genes for different biological processes. The genes, identified thus, now need to be integrated to the genetic and physical maps including those based on BAC clones derived from genomic and cDNA libraries. This will then allow effective use of DNA markers in MAS as well as in MBC projects with an ultimate objective of crop improvement in cereals. Additional work however is needed in the area of association mapping and linkage disequilibrium (LD), which have proved extremely fruitful in human genetics (Jannink and Walsh, 2002). Since in this area, only a beginning has been made in cereals (Thornsberry *et al.*, 2002; Remington *et al.*, 2002; see Rafalski and Morgante, 2004), and its impact is yet to be realized, this aspect has not been covered in this book. Similarly, not much work in the area of cereal proteomics is available so far to find a place in this book. We hope that more work on cereals will be done in these areas, by the time this book becomes available to the readers. In the next decade, it will be a major challenge in each cereal crop to build integrated databases, combining information on genome and proteome sequences and their maps, mRNA and protein expression profiles, mutant phenotypes, metabolism and allelic variation. This integrated information on trancriptome(s)/genome(s) will then be accessible on-line. We believe that in future, we will be talking of a 'cereal gene' rather than of a gene for a particular cereal species in order to improve the quality as well as quantity of cereal grain worldwide, which is a major challenge indeed, to be realized in the next few years.

6. REFERENCES

Akhunov E.D., Goodyear A.W., Geng S., Qi L.-L., Echalier B., Gill B.S., Miftahudin, Gustafson J.P., Lazo G., Chao S.M., *et al.* The organization and rate of evolution of wheat genomes are correlated with recombination rates along chromosome arms. Genome Res 2003, 13: 753-763

Albar L., Ndjiondjop M.-N., Esshak Z., Berger A., Pinel A., Jones M., Fargette D., Ghesquière A. Fine genetic mapping of a gene required for rice yellow mosaic virus cell to cell movement. Theor Appl Genet 2003; 107 (2): 371-378

Appels R., Francki M., Chibbar R. Advances in cereal functional genomics. Funct Integr Genomics 2003; 3: 1-24.

Badr A., Muller K., Schafer-Pregl R., El Rabey H., Effgen S., Ibrahim H.H., Pozzi C., Rohde W., Salamini F. On the origin and domestication history of barley (*Hordeum vulgare*). Mol Biol Evol 2000; 17: 499-510

Barry G.F. The use of Monsanto draft rice genome sequence in research. Plant Physiol 2001; 125: 1164-1165

Bellgard M., Ye J., Gojobori T., Appels R. The bioinformatics challenges in comparative analysis of cereal genomes- an overview. Funct Integr Genom 2004; 4: 1-11

Bennetzen J.L. Comparative sequence analysis of plant nuclear genomes: Microcolinearity and its many exceptions. Plant Cell 2000; 12:1021-1029

Bennetzen J.L., Ma J. The genetic colinearity of rice and other cereals on the basis of genomic sequence analysis. Curr Opin Plant Biol 2003; 6:128-33

Bensen R.J., Johal G.S., Crane V.C., Tossberg J.T., Schnable P.S., Meeley R.B., Briggs S.P. Cloning and characterization of the maize *An1* gene. Plant Cell 1995; 7: 75-84

Blair M.W., Garris A.J., Iyer A.S., Chapman B., Kresovich S. McCouch S.R. High resolution mapping and candidate gene identification at the *xa5* locus for bacterial blight resistance in rice (*Oryza sativa* L.). Theor Appl Genet 2003; 107: 62-73

Bohnert, H.J and the Plant Stress Consortium. A genomics approach towards salt stress tolerance. Plant Physiol Biochem 2001; 39: 295-311

Bryan G.T., Wu K.S., Farrall L., Jia Y.L., Hershey H.P., McAdams S.A., Faulk K.N., Donaldson G.K., Tarchini R., Valent B. A single amino acid difference distinguishes resistant and susceptible alleles of the rice blast resistance gene *Pi-ta*. Plant Cell 2000; 12: 2033-2045

Buschges, R., Hollricher, K., Panstruga, R., Simons, G., Wolter, M., Frijters, A., vanDaelen, R., vanderLee, T., Diergaarde, P., Groenendijk, J. *et al.* The barley *mlo* gene: A novel control element of plant pathogen resistance. Cell 1997; 88: 695-705

Chandel V.L., Brendel V. The maize genome sequencing project. Plant Physiol 2002; 130: 1594-1597

Chen M., Presting G., Barbazuk W.B., Goicoechea J.L., Blackmon B., Fang G., Kim H., Frisch D., Yu Y., Sun S., *et al.* An integrated physical and genetic map of the rice genome. Plant Cell 2002; 14: 537- 545

Chen M., SanMiguel P. Bennetzen J.L. Sequence organization and conservation in *sh2/a1*-homologous regions of sorghum and rice. Genetics 1998; 148:435-443

Chen M., SanMiguel P., deOliveira A.C., Woo S.S., Zhang H., Wing R.A., Bennetzen J.L. Microcolinearity in *sh2*-homologous regions of the maize, rice, and sorghum genomes. Proc Natl Acad Sci USA 1997; 94: 3431-3435

Childs K.L., Klein R.R., Klein P.E., Morishige D.T., Mullett J.E. Mapping genes on an integrated sorghum genetic and physical mapping using cDNA selection technology. Plant Jour 2001; 27: 243-255

Das L., Martienssen R. Site-selected transposon mutagenesis at the *Hcl106* locus in maize. Plant Cell 1995; 7: 287-294

Davis G.L., McMullen M.D., Baysdorfer C., Musket T., Grant D. Staebell M., Xu G., Polacco M., Koster L., Melia-Hancock S., *et al.* A maize map standard with sequenced core markers, grass genome reference points and 932 expressed sequence tagged sites (ESTs) in a 1736-locus map. Genetics 1999; 152: 1137- 1172

Devos K.M. Gale M.D. Genome relationships: the grass model in current research. Plant Cell 2000; 12: 637-646

Dubcovsky J., Ramakrishna W., SanMiguel P.J., Busso C.S., Yan L.L., Shiloff B.A., Bennetzen J.L. Comparative sequence analysis of colinear barley and rice bacterial artificial chromosomes. Plant Physiol 2001; 125: 1342-1353.

Edwards K.J., Stevenson D. Cereal genomics. Adv Bot Res Adv Plant Pathol 2001; 34: 1-22

Feng Q., Zhang Y., Hao P., Wang S., Fu G., Huang Y., Li Y., Zhu J., Liu Y., Hu X., *et al.,* Sequence and analysis of rice chromosome 4. Nature 2002; 420: 316-320

Feuillet C., Keller B. High gene density is conserved at syntenic loci of small and large grass genomes. Proc Nat Acad Sci USA 1999; 96: 8265-8270

Feuillet C., Keller B. Comparative genomics in the grass family: molecular characterization of grass genome structure and evolution. Ann Bot 2002; 89:3-10

Feuillet C., Travella S., Stein N., Albar L., Nublat A., Keller B. Map-based isolation of the leaf rust disease resistance gene *Lr10* from the hexaploid wheat (*Triticum aestivum* L.) genome. Proc Nat Acad Sci USA 2003; in press.

Fraray A., Nesbitt T.C., Frary A., Grandillo S., Knapp E., Cong B., Liu J., Meller J., Elber R., Alpert K.B., *et al. fw2.2*: a quantitative trait locus key to the evolution of tomato fruit size. Science 2000; 289:85-88

Friedt W., Werner K., Pellio B., Weiskorn C., Krämer M., Ordon F. Strategies of breeding for durable disease resistance in cereals. Prog Bot 2003; 64: 138-167

Gale M.D., Devos K.M. Comparative genetics in the grasses. Proc Natl Acad Sci USA 1998; 95: 1971-1974

Gao L.F., Tang J.F., Li H.W., Jia J.Z. Analysis of microsatellites in major crops assessed by computational and experimental approaches. Mol Breed 2003; 12: 245-261

Gill B.S., Appels R. NSF/USDA/IGROW workshop on wheat genome sequencing. Genetics 2004; in press

Goff S.A. Rice as a model for cereal genomics. Curr Opin Plant Biol 1999; 2: 86-89

Goff S.A., Ricke D., Lan T.H., Presting G., Wang R.L., Dunn M., Glazebrook J., Sessions A., Oeller P., Varma H., *et al.* A draft sequence of the rice genome (*Oryza sativa* L. ssp *japonica*). Science 2002; 296: 92-100

Gregory S.G., Sekhon M., Schein J., Zhao S., Osoegawa K., Scott C.E., Evans R.S., Burridge P.W., Cox T.V., Fox C.A., *et al.* A physical map of mouse genome. Nature 2002; 418:743-750

Gupta P.K., Varshney, R.K. The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica 2000; 113: 163-185

Gupta P.K., Varshney R.K., Prasad M. Molecular markers: principles and methodology. In: S.M. Jain, D.S. Brar, B.S. Ahloowalia (eds.), *Molecular Techniques in Crop Improvement*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2002; pp 9-54

Harushima Y., Yano M., Shomura A., Sato M., Shimono T., Kuboki Y., Yamamoto T., Lin S.Y., Antonio B.A., Parco A., *et al.* A high density rice genetic linkage map with 2275 markers using a single F2 population. Genetics 1998; 148: 479-494

Heslop-Harrison J.S. Comparative genome organization in plants: From sequence and markers to chromatin and chromosomes. Plant Cell 2000; 12: 617-635

Heun M., SchaferPregl R., Klawan D., Castagna R., Accerbi M., Borghi B., Salamini F. Site of einkorn wheat domestication identified by DNA fingerprinting. Science 1997; 278: 1312-1314

Hirochika H. Retrotransposons of rice: their regulation and use for genome analysis. Plant Mol Biol 1997; 35: 231-240

Horvath H., Rostoks N., Brueggman R., Steffenson B., von Wettstein D., Kleinhofs A. Genetically engineered stem rust resistance in barley using the *Rpg1* gene. Proc Natl Acad Sci USA 2003; 100: 364-369

Huang L., Brooks S.A., Li W., Fellers J.P., Trick H.N., Gill B.S. Map-based cloning of leaf rust resistance gene *Lr21* from the large and polyploid genome of bread wheat. Genetics 2003; 164: 655-664

Izawa T., Ohnishi T., Nakano T., Ishida N., Enoki H., Hashimoto H., Itoh K., Terada R., Wu C.Y., Miyazaki C., *et al.* Transposon tagging in rice. Plant Mol Biol 1997; 35: 219-229

Jannink J.-L., Walsh B. Association mapping in plant populations. In M.S. Kang (ed.), *Quantitative Genetics, Genomics and Plant Breeding*, CAB International, USA, 2002, pp. 59- 68

Jansen R.C. Studying complex biological systems using multifactorial perturbation. Nature Rev Genet 2003; 4: 145-151

Jansen R.C., Nap J.-P. Genetical genomics: the added value from segregation. Trends Genet 2001, 17: 388-391

Kantety R.V., Rota M.L., Matthews D.E., Sorrells M.E. Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. Plant Mol Biol 2002; 48: 501-510

Kawasaki S., Borchert C., Deyholos M., Wang H., Brazille S., Kawai K., Galbraith D., Bohnert H.J. Gene expression profiles during the initial phase of salt stress in rice. Plant Cell 2001; 13: 889-905

Kehoe D., Villand P., Sommerville S. DNA microarrays for studies of higher plants and other photosynthetic organisms. Trends Plant Sci 1999; 4:38-41

Koebner R., Powell W., Donini P. Contributions of DNA molecular marker technologies to the genetics and breeding of wheat and barley. Plant Breeding Reviews 2001, 21: 181-220

Koebner R.M.D., Summers R.W. 21st century wheat breeding: plot selection or plate detection? Trends Biotech 2003; 21: 59-63

Klein P.E., Klein R.R., Cartinhour S.W., Ulanch P.E., Dong J., Obert J.A., Morishge D.T., Schlueter S.D., Childs K.L., Ale M., *et al.* A high throughput AFLP based method for constructing integrated genetic and physical maps: progress toward a sorghum genome map. Genome Res 2000; 10: 789-807

Kumar L.S. DNA markers in plant improvement: An overview. Biotech Adv 1999; 17: 143-182

Kumar A. Bennetzen J.L. Plant retrotranspsons. Annu Rev Genet 1999; 33:479-532

Künzel G., Korzun L., Meister A. Cytologically integrated physical restriction fragment length polymorphism maps for the barley genome based on translocation breakpoints. Genetics 2000; 154: 397-412

Kurata, N., Nagamura, Y., Yamamoto, K., Harushima, Y., Sue, N., Wu, J., Antonio, B.A., Shomura, A., Shimizu, T., Lin, *et al.* A 300 kolobase interval genetic map of rice including 883 e×pressed sequences. Nature Genet 1994; 8: 365-372

Lagudah E.S., Dubcovsky J., Powell W. Wheat genomics. Plant Physiol Biochem 2001; 39: 335-344

Lahaye T., Shirasu K., Schulze-Lefert P. Chromosome landing at the barley *Rar1* locus. Mol Gen Genet 1998; 260: 92-101

Lee J.-M., Williams M.E., Tingey S.V., Rafalski J.-A. DNA array profiling of gene expression changes during maize embryo development. Funct Integr Genom 2002; 2: 13-27

Leister D., Kurth J., Laurie D.A., Yano M., Sasaki T., Graner A., Schulze-Lefert P. RFLP and physical mapping of resistance gene homologues in rice (*Oryza sativa*) and barley (*Hordeum vulgare*). Theor Appl Genet 1999; 98:509-520

Li W., Gill B.S. The colinerity of the *Sh2/ A1* orthologous region in rice, sorghum and maize is interrupted and accompanied by genome expansion in the triticeae. Genetics 2002; 160: 1153-1162

McCouch S.R.,Teytelman L., Xu Y., Lobos K.B., Clare K., Walton K., Fu B., Maghirang R., Li Z., Xing Y., *et al.* Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.). DNA Res 2002; 9: 199-207

Moore G., Devos K.M., Wang Z., Gale M.D. Grasses, line up and form a circle. Curr Biol 1995; 5: 737-739

Ordon F., Köhler H., Pellio B., Friedt W. From genetic towards physical distances: high resolution mapping of plant resistance genes. Prog Bot 2000; 61:37-52

Ordon F., Wenzel W., Friedt W. II Recombination: molecular markers for resistance genes in major grain crops. Prog Bot 1998; 59: 49-79

Osborne B.I., Baker B. Mover and shakers – maize transposons as tools for analysing other plant genomes. Curr Opin Plant Biol 1995; 7: 406-413

Ozkan H., Brandolini A., Schafer-Pregl R., Salamini F. AFLP analysis of a collection of tetraploid wheats indicates the origin of emmer and hard wheat domestication in southeast Turkey. Mol Biol Evol 2002; 19: 1797-1801

Ozturk Z.N., Talame V., Deyholos M., Michalowski C.B., Galbraith D.W., Gozukirmizi N., Tuberosa R., Bohnert H.J. Monitoring large-scale changes in transcript abundance in drought- and salt-stressed barley. Plant Mol Biol 2002; 48: 551-573

Paterson A.H., Bowers J.E., Paterson D.G., Estill J.C., Chapman B.A. Structure and evolution of cereal genomes. Curr Opin Genet Dev 2003; 13: 644-650

Paterson A.H., Lin Y.-R., Li Z., Schertz K.F., Doebley J.F., Pinson S.R.M., Liu S.-C., Stansel J.W., Irvine J.E. Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. Science 1995; 269: 1714-1718

Paterson A.H., Bowers J.E., Burow M.D., Draye X., Elsik C.G., Jiang C.X., Katsar C.S., Lan T.H., Lin Y.R., Ming R., *et al.* Comparitive genomics of plant chromosomes. Plant Cell 2000; 12: 1523-1539

Peng J.H., Ronin Y., Fahima T., Roder M.S., Li Y.C., Nevo E., Korol A. Domestication quantitative trait loci in *Triticum dicoccoides*, the progenitor of wheat. Proc Natl Acad Sci USA 2003; 100: 2489-2494

Phillips R.L., Vasil I.K. (eds.) *DNA-Based Markers in Plants*, Kluwer Academic Publishers, The Netherlands, 1994

Phillips R.L., Vasil I.K. (eds.) *DNA-Based Markers in Plants*, 2nd Edition, Kluwer Academic Publishers, The Netherlands, 2001

Potokina E., Sreenivasulu N., Altschmied L., Michalek W., Graner A. Differential gene expression during seed germination in barley (*Hordeum vulgare* L.). Funct Integr Genom 2002; 2:28-39

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

Rafalski A., Morgante M. Corn and humans: recombination and linkage disequilibrium in two genomes of similar size. Trends Genet 2004, 20: 103-111

Remington D.L., Thornsberry J.M., Matsuoka Y., Wilson L.M., Whitt S.R., Doebley J., Kresovich S., Goodman M.M., Buckler E.S. Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc Natl Acad Sci USA 2001; 98: 11479-11484

Ribaut J.-M., Bänziger M., Betran J., Jiang C., Edmeades G.O., Dreher K., Hoisington D. Use of molecular markers in plant breeding: drought tolerance improvement in tropical maize. In: M. S. Kang (ed.), *Quantitative Genetics, Genomics and Plant Breeding*, CAB International, USA, 2002; pp 85- 99

Ribaut J.-M., Hoisington D. Marker-assisted selection: new tools and startegies. Trends Plant Sci 1998; 3: 236-239

Richmond T., Somerville S. Chasing the dream: plant microarrays. Curr Opin Plant Biol 2000; 3:108-116

Salamini F., Özkan H., Brandolini A., Schäfer-Pegl R., Martin W. Genetics and geography of wild cereal domestication in the near East. Nature Rev Genet 2002; 3: 429-441

Salvi S., Tuberosa R., Chiapparino E., Maccaferri M., Veillet S., van Beuningen L., Isaac P., Edwards K., Phillips R.L. Toward positional cloning of *Vgt1*, a QTL controlling the transition from the vegetative to the reproductive phase in maize. Plant Mol Biol 2002; 48: 601-613.

Sandhu D., Gill K.S. Gene-containing regions of wheat and the other grass genomes. Plant Physiol 2002; 128: 803-811

SanMiguel P., Tikhonov A., Jin Y.K., Motchoulskaia N., Zakharov D., Melake-Berhan A., Springer P.S., Edwards K.J., Lee M., Avramova Z., *et al.* Nested retrotransposons in the intergenic regions of the maize genome. Science 1996; 274: 765-768

Sasaki T. Matsumoto T., Yamamoto K., Sakata K., Baba T., Katayose Y., Wu J., Niimura Y., Cheng Z., Nagamura Y., *et al.* The genome sequence and structure of rice chromosome 1. Nature 2002; 420: 312-316

Schena M. *DNA Microarrays: A Practical Approach*, Oxford University Press, New York, USA, 1999

Scholz S., Lorz H., Lutticke S.. Transposition of the maize transposable element Ac in barley (*Hordeum vulgare* L.). Mol Gen Genet 2001; 264: 653-661

Sharopova N., McMullen M.D., Schultz L., Schroeder S., Sanchez-Villeda H., Gardiner J., Bergstrom D., Houchins K., Melia-Hancock S., Musket T., *et al.* Development and mapping of SSR markers for maize. Plant Mol Biol 2002; 48: 463-481

Shirasu K., Lahaye T., Tan M.W., Zhou F.S., Azevedo C., Schulze-Lefert P. A novel class of eukaryotic zinc-binding proteins is required for disease resistance signaling in barley and development in *C. elegans*. Cell 1999; 99: 355-366

Song W.Y., Wang G.L., Chen L.L., Kim H.S., Pi L.Y., Holsten T., Gardner J., Wang B., Zhai W.X., Zhu L.H., *et al.* A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa21*. Science 1995; 270: 1804-1806

Sreenivasulu N., Altschmied L., Panitz R., Hahnel U., Michalek W., Weschke W., Wobus U. Identification of genes specifically expressed in maternal and filial tissues of barley caryopses: A cDNA array analysis. Mol Gen Genom 2002a; 266: 758-767

Sreenivasulu N., Kavikishor P.B., Varshney R.K., Altschmied L. Mining functional information from cereal genomes- the utility of expressed sequence tags (ESTs). Curr Sci 2002b; 83:965- 973

Stein N., Feuillet C., Wicker T., Schlagenhauf E., Keller B. 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.). Proc Nat Acad Sci USA 2000; 97: 13436-13441

TAGI, The Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 2000; 408: 796-815

Takahashi Y., Shomura A., Sasaki T., Yano M. *Hd6*, a rice quantitative trait locus involved in photoperiod sensitivity, encodes the alpha subunit of protein kinase CK2. Proc Nat Acad Sci USA 2001; 98: 7922-7927

Tao Q.Z., Chang Y.L., Wang J.Z., Chen H.M., Islam-Faridi M.N., Scheuring C., Wang B., Stelly D.M., Zhang H.B. Bacterial artificial chromosome-based physical map of the rice genome constructed by restriction fingerprint analysis. Genetics 2001; 158: 1711-1724

Tarchini R., Biddle P., Wineland R., Tingey S., Rafalski A. The complete sequence of 340 kb of DNA around the rice Adh1-Adh2 region reveals interrupted colinearity with maize chromosome 4. Plant Cell 2000; 12: 381-391

Thangavelu M., James A.B., Bankier A., Bryan G.J., Dear P.H., Waugh R. HAPPY mapping in plant genome: reconstruction and analysis of a high-resolution physical map of 1.9 Mpp region of *Arabidopsis thaliana* chromosome 4. Plant Biotech Jour 2003; 1:23-31

The International HapMap Consortium. The international HapMap project. Nature 2003; 426: 789-796

The Rice Chromosome 10 Sequencing Consortium: Yu Y., Rambo T., Currie J., Saski C., Kim H.R., Collura K., Thompson S., Simmons J., Yang T.-J., Nah G. *et al.* In-depth view of structure, activity, and evolution of rice chromosome 10. Science 2003; 300: 1566-1569

Thomas W.T.B. Molecular marker-assisted versus conventional selection in barley breeding. In: Slafer G.A., Molinacano J.L., Savin R., Araus J.L., Romagosa I. (eds.), *Barley Science: Recent Advances From Molecular Biology to Agronomy of Yield and Quality*, Food Products Press, Binghamton, NewYork, USA, 2002; pp177-204

Thomas W.T.B. Prospects for molecular breeding of barley. Ann Appl Biol 2003; 142:1-12

Thornsberry J.M., Goodman M.M., Doebley J., Kresovich S., Nielsen D., Buckler E.S. *Dwarf8* polymorphisms associate with variation in flowering time. Nature Genet 2001; 28: 286-289

Tikhonov A.P., SanMiguel P.J., Nakajima Y., Gorenstein N.M., Bennetzen J.L., Avramova Z. Colinearity and its exceptions in orthologous adh regions of maize and sorghum. Proc Natl Acad Sci USA 1999; 96: 7409-7414

Varshney R.K., Thiel T., Stein N., Langridge P., Graner A. *In silico* analysis on frequency and distribution of microsatellites in ESTs of some cereal species. Cell Mol Biol Lett 2002; 7: 537-546

Venter J.C., Adams M.D., Sutton G.G., Kerlavage A.R., Smith H.O., Hunkapiller M. Shotgun sequencing of the human genome. Science 1998; 280: 1540-1542

Vicient C.M., Suoniemi A., Anamthawat-Jónsson K., Tanskanen J., Beharav A., Nevo E., Schulman A.H. Retrotransposon *BARE*-1 and its role in genome evolution in the genus *Hordeum*. Plant Cell 1999b; 11: 1769-1784

Walbot V. Saturation mutagenesis using maize transposons. Curr Opin Plant Biol 2000; 3: 103-107

Wang Z.X., Yano M., Yamanouchi U., Iwamoto M., Monna L., Hayasaka H., Katayose Y., Sasaki T. The *Pib* gene for rice blast resistance belongs to the nucleotide binding and leucine-rich repeat class of plant disease resistance genes. Plant Jour 1999; 19: 55-64

Waugh R., Dear P.H., Powell W., Machray G.C. Physical education - new technologies for mapping plant genomes. Trends Plant Sci 2002; 7:521-523

Wei F., Gobelman-Werner K., Morroll S.M., Kurth J., Mao L., Wing R., Leister D., Schulze-Lefert P., Wise R.P. The *Mla* (powdery mildew) resistance cluster is associated with three NBS-LRR gene families and suppressed recombination within a 240-kb DNA interval on chromosome 5S (1HS) of barley. Genetics 1999; 153: 1929-1948

Wicker T., Matthews D.E., Keller B. TREP: a database for Triticeae repetitive elements. Trends Plant Sci 2002; 7: 561-562

Wise R.P. Disease resistance: what's brewing in barley genomics. Plant Disease 2000; 84: 1160-1170

Wu J.Z., Maehara T., Shimokawa T., Yamamoto S., Harada C., Takazaki Y., Ono N., Mukai Y., Koike K., Yazaki J., *et al.* A comprehensive rice transcript map containing 6591 e×pressed sequence tag sites. Plant Cell 2002; 14: 525-535

Yahiaoui N., Srichumpa P., Dudler R., Keller B. Genome analysis at different ploidy levels allows cloning of the powdery mildew resistance gene *Pm3b* from hexaploid wheat. Plant Jour 2003; in press

Yano M. Genetic and molecular dissection of naturally occurring variation. Curr Opin Plant Biol 2001; 4: 130-135

Yano M., Katayose Y., Ashikari M., Yamanouchi U., Monna L., Fuse T., Baba T., Yamamoto K., Umehara Y., Nagamura Y., *et al. Hd1*, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the Arabidopsis flowering time gene *CONSTANS*. Plant Cell 2000; 12: 2473-2483

Yim Y., Davis G., Duru N., Musket T., Linton E., Messing J., McMullen M., Soderlund C., Polacco M., Gardiner J., *et al.* Characterization of three maize bacterial artificial chromosome libraries toward anchoring of the physical map to the genetic map using high-density bacterial artificial chromosome filter hybridization. Plant Physiol 2002; 130:1686-1696

Yoshimura S., Yamanouchi U., Katayose Y., Toki S., Wang Z.-X., Kono I., Kurata N., Yano M., Iwata N., Sasaki T. Expression of *Xa1*, a bacterial blight-resistance gene in rice, is induced by bacterial inoculation. Proc Natl Acad Sci USA 1998; 95: 1663-1668

Yu J., Hu S., Wang J., Wang G., Li S.G., Wong K.S.G., Liu B., Deng Y., Dai L., Zhou Y., Zhang X., *et al.* A draft sequence of the rice genome (*Oryza sativa* L. sspl. *indica*). Science 2002; 296:79-92

Zhu T., Budworth P., Chen W., Provart N., Chang H.-S., Guimil S., Su W., Estes B., Zou G., Wang X. Transcriptional control of nutrient partitioning during rice grain filling. Plant Biotech Jour 2003; 1: 59 - 63

Zhu Z.G., Fu Y.P., Xiao H., Hu G.C., Si H.M., Yu Y.H., Sun Z.H. Ac/Ds transposition activity in transgenic rice population and DNA flanking sequence of Ds insertion sites. Acta Bot Sin 2003; 45: 102-107

Chapter 2

MOLECULAR MARKER SYSTEMS AND THEIR EVALUATION FOR CEREAL GENETICS

Daryl J. Somers

Agriculture and Agri-Food Canada – Cereal Research Centre, 195 Dafoe Rd, Winnipeg, Manitoba, R3T-2M9 Canada
E-mail: somersD@agr.gc.ca

1. INTRODUCTION

DNA-based molecular markers are the most powerful diagnostic tools to detect DNA polymorphism both at the level of specific loci and at the whole genome level. In the past, these DNA-based molecular markers were developed either from genomic DNA libraries (RFLPs and microsatellites) or from random PCR amplification of genomic DNA (RAPDs) or both (AFLP). More recently, however, the availability of genomic DNA and cDNA sequences (ESTs) in the public databases has made marker development more direct and sometimes also cost effective. The primary application of molecular markers is molecular characterization of germplasm or mapping of whole genomes. Consequently, in addition to DNA fingerprinting that has been done for germplasm of many crops, molecular genetic maps of whole genomes, based on allelic variation at individual marker loci, have been constructed in all major crops. This has been achieved through linkage analysis, utilizing specially designed mapping populations, which allow estimation of recombination between linked marker loci. Utilizing the data recorded on the same mapping population for the molecular marker genotypes and the phenotypic traits, linkage analysis also allows identification of molecular markers that are linked to genes/QTLs controlling important agronomic traits. This gene tagging is already being used for marker-aided selection (MAS) in plant breeding programmes, sometimes leading to desirable gene stacking or pyramiding. This is a common use of molecular markers and has the potential to make an important impact on the agriculture sector, at both the producer and consumer levels. Molecular

P.K. Gupta and R.K. Varshney (eds.), Cereal Genomics, 19-34.

markers have also been used for estimating genetic diversity and for fingerprinting the advanced lines, so that a plant breeder can design crosses, which may release increased desirable variation. This has been shown to be particularly useful in maize breeding, where lines from different heterotic pools can be identified and crossed in an attempt to maximize yield.

In recent years, due to the emergence of genomics as an important area of research, emphasis in many areas of molecular marker research and also in the field of application of molecular markers has shifted from the traditional concepts introduced above, to a broader field involving genomics research. Genomics research makes use of new technologies that enable highly parallel analyses of gene expression or allele characterization of whole genomes in minimal time and with higher precision. Fairly dense molecular genetic maps are available for most cereals, not only due to abundance of available molecular markers, but also due to several coordinated international efforts that are underway to generate these markers and to align maps based on these markers within and across cereal species. The DNA sequence information generated first via EST sequencing and then by whole genome sequencing, has also led to a rapid increase in our knowledge of allelic sequence variation. The ability to generate allelic DNA sequence variation or to mine these sequences from existing databases (Somers *et al.,* 2003a; Kanazin *et al.,* 2002) quickly lead to the discovery of single nucleotide polymorphisms (SNPs), which are clearly, the most abundant type of molecular markers that are also amenable to high throughput analysis. Thus, the direct access to enormous data on DNA sequences in cereal species has already accelerated the pace, at which allelic variation is not only being detected '*in silico*', but is also being mapped in reference populations.

It is heartening to note, that the pioneering work that went into marker development and accurate mapping subsequently proved useful for constructing physical maps of several cereal genomes, thus enabling direct gene discovery and map based cloning. Whole genome sequences, as well as ESTs developed through partial sequencing of cDNAs, provide a rich resource of molecular markers, but more importantly, they also provide direct information about sequence diversity between alleles. Improvements in cereal crop production and end use quality will perhaps come not simply from cloning genes, but mainly from our understanding of allelic variation. Molecular markers are already playing an important role in gene/allele discovery, and are thus becoming an essential component of current cereal genomics programs worldwide. This chapter will first deal with a variety of historical aspects of cereal molecular marker research; it will then focus on areas of newer genomics research and will finally deal with the applications of molecular markers in cereal genomics.

2. SOURCES AND TYPES OF MARKERS

In the initial phase of molecular marker research, genomic libraries were used to isolate low copy DNA fragments to be used as DNA probes. These DNA probes were hybridized back to restriction digests of genomic DNA to reveal allelic length variation. This is referred to as restriction fragment length polymorphism (RFLP). RFLP markers have been used in many species such as wheat, to develop genetic linkage maps (Chao *et al.,* 1989; see Gupta *et al.,* 1999 for references). These markers proved useful in comparative mapping studies because the DNA probes belonging to one species could be readily hybridized to genomic DNA digests of related species (Devos *et al.,* 1993). Gene probes, developed from cDNA are particularly useful for comparative mapping since there is sufficient homology in gene sequences to cross hybridize to related species. This line of research has led to alignments of the chromosomes from rice, wheat, barley and maize. Even today, RFLPs offer the best marker type to perform precise comparative genome mapping experiments. The only drawbacks of using RFLPs include the high cost of genotyping and the slow, low throughput nature of hybridization technology. Both these factors discourage plant breeders from adopting Southern analysis employed in RFLP as a molecular breeding tool.

Hybridization technique used for developing RFLP markers was later followed by PCR, a faster and less expensive technique. PCR provided a rapid development of many DNA amplification strategies that are all fundamentally similar. PCR-based DNA markers rely on sequence variation in annealing sites or DNA length differences between amplified products. The simplest of these is random amplification of polymorphic DNA (RAPD) (Williams *et al.,* 1990). RAPD has been widely used in cereals to measure and characterize genetic diversity, to create linkage maps and to tag genes controlling important traits. The most prevalent difficulty with the RAPD technique, however, is a lack of reproducibility and lack of locus specificity, particularly in polyploid species such as wheat. Amplified fragment length polymorphism (AFLP) (Vos *et al.,* 1995), which combines the merits of the both, the RFLP and the PCR techniques has been used widely in cereals for many types of genetic analyses, including genetic map construction. A key advantage of the AFLP technique is a higher degree of reproducibility.

A significant development in molecular marker technology during 1990s has been the introduction of microsatellites, a form of variable number of tandem repeats, in most cases containing two or three nucleotides repeated. Microsatellites have several advantages; they are PCR-based, locus specific (which is very important in polyploids) and are typically codominant. As a result, microsatellites have been used extensively to develop genetic maps in

a variety of cereal species including wheat, barley, rice and maize (Roder *et al.,* 1998; Macaulay, *et al.,* 2001;, Temnykh *et al.,* 2001; Sharopova *et al.,* 2002). Microsatellites are also a good molecular marker technology for plant breeding, given their high throughput, locus specificity and the high level of polymorphism which they detect. This is particularly important in crops such as wheat, where most markers detect very low levels of polymorphism (Bryan *et al.,* 1999). A negative aspect of microsatellites, however, is the cost of their development. DNA sequencing is required to discover the microsatellite and to design the locus specific PCR primers to amplify the alleles (Bryan *et al.,* 1997; Song *et al.,* 2002). DNA sequencing was not needed for marker development in the other techniques described above. Again, with the abundance of DNA sequences of cereals available either as ESTs or as whole genome sequences, '*in silico*' mining can be performed to find microsatellites, to design primers and to develop the markers (Kantety *et al.,* 2002; Varshney *et al*., 2002). Where DNA sequences are available in the public domain, marker development is far less expensive since the sequencing has already been done (for a review, see Gupta and Varsheny, 2000). Therefore, EST-based SSR markers are being developed in some cereal crops like barley (Thiel *et al*., 2003), wheat (Holton *et al.,* 2002), etc.

The latest of the DNA-based markers are the single nucleotide polymorphisms (SNPs). Although SNPs provide for a newer marker technology, they are not based on any new concept. Ever since DNA sequencing became feasible, alleles of one locus could be aligned and differences in a given sequence could be identified. Today, specialized bioinformatics tools using EST sequences available in public databases, has made SNP discovery more automated (Marth *et al.,* 1999). Discovery of SNPs in the Human genome is highly advanced with 1.8 million SNPs documented by early 2003. While SNP discovery in cereals is most advanced in maize (Tenaillon *et al.,* 2001), the resources for wheat, barley and rice are also now mature enough for large scale SNP discovery. SNPs have the potential to provide for high throughput and automation, as new detection platforms are developed. Currently, allele specific PCR, single base extension and array hybridization are only some of the methods for detecting SNPs, with varying input costs and levels of sophistication (Gupta *et. al.*, 2000). There are estimates in cereals that one SNP occurs every 100-600 nucleotides in coding DNA (Somers *et al.,* 2003a; Tenaillon *et al.;* 2001, Kanazin *et al.,* 2002).

Clearly, molecular marker technology over the past 20 years has evolved quite rapidly, and has been extensively used in cereal genetic studies. It is interesting to note that none of the early generation molecular marker

technologies has become obsolete. There are just more creative ways to use these techniques to understand cereal genetics.

3. NEED FOR MARKERS IN CEREAL GENOMICS

Cereal genomics can be split into two related fields namely, functional genomics and structural genomics. Functional genomics deals primarily with gene expression, transcriptional and translational control, comparative gene expression and highly parallel analyses of expressed genes. Structural genomics, on the other hand, deals with genetic mapping, comparative mapping, physical mapping, gene location, and the studies dealing with gene structure and allele variation. Molecular markers can be used for both functional and structural genomics.

Gene expression in bread wheat involves homoeologous loci, which are not easily distinguishable from each other by typical hybridization tests, since there is a high degree of sequence similarity among homoeoloci. Therefore, SNP markers are the most useful in distinguishing homoeologous transcripts from each other, due to the abundance and sensitivity of assays to single base changes. One of the prospects of EST sequencing and QTL analysis is to map ESTs or candidate gene sequences to genetic maps where trait information is available. This forms the basic linkage between plant form and function. SNPs can be used as molecular markers to precisely position an EST into a genetic map and thus gives the EST an added degree of relevance to a particular functional genomics experiment (Somers *et al.,* 2003a).

The application of molecular markers is more widespread in structural genomics than in functional genomics. For instance, genetic mapping that was discussed earlier in this chapter is the most basic form of structural genomics. Similarly, physical mapping relates to large-scale sequencing, production of large insert libraries using bacterial artificial chromosomes (BACs), BAC contig assembly, map-based gene cloning and '*in silico*' comparative mapping. In these studies, molecular markers that are already mapped genetically are used to derive information about organization and reference points to BAC clones and chromosome intervals. A BAC library is a collection of clones averaging 80-120kbp in length that may cover the genome 3-6 fold. Markers can be used to select clones, either by hybridization or PCR, which enables small sets of BAC clones to be ordered along the genetic map. In cases where there is a genetic interval densely populated with markers (i.e. 0.1 cM/marker), the markers can be used to select clones from the BAC library and used to assemble the BAC clones into contigs across this chromosome interval.

The other primary use of markers in structural genomics is to achieve physical alignments of chromosome segments derived from related species, such as rice, with wheat, barley or maize. Since the draft sequence of the rice genome is now available (Goff *et al.,* 2002; Yu *et al.,* 2002), rice RFLP markers can be assigned physical positions on the rice chromosomes. Rice RFLP markers that are also mapped to wheat enable alignment of the wheat chromosome map with the rice physical map. A remarkable resource summarizing much of this is available at "Gramene" http://www.gramene.org (Ware *et al.,* 2002). Wheat, barley and maize ESTs can also be used as markers for further chromosome alignment by knowing the genetic locations of ESTs and then working out their corresponding physical locations in the rice genome, facilitated by BLAST analysis (Altschul *et al.,* 1997). This resource of '*in silico*' mapping and genome alignment is still in the early stages of development, but is beginning to reveal the collinearity of wheat, barley and maize genomes with the rice genome. Recently, some laboratories in the US have established orthologous relationship between wheat and rice (Sorrells *et al*., 2003; see Chapter 19 by Sorrells in this book).

4. USE OF MARKERS IN CEREAL GENETICS

4.1 Gene Tagging

Gene tagging is a common application of molecular markers and is typically achieved by bulked segregant analysis (BSA) (Michelmore *et al.,* 1991) or by QTL analysis. In both cases, robust, reliable phenotypic data has a strong bearing on the success of the efforts to tag a gene controlling a component of a trait. BSA is very effective, where traits are Mendelian in nature and where PCR-based marker systems such as RAPD and AFLP are used. QTL analysis is, however, more appropriate, where the traits are quantitative in nature. In this case, full genetic maps are needed to scan the genome for associations of specific bins with the trait.

Most efforts on gene tagging in cereals, reported in the literature have been focused on disease resistance (see Friedt *et al*., 2003 for references). Examples include rice bacterial blight (Sanchez *et al.,* 2000), wheat powdery mildew (Huang *et al.;* 2000, Hartl *et al.,* 1999), wheat leaf rust (Huang and Gill 2001; Hussien *et al.,* 1998), wheat stripe rust (Bariana *et al.,* 2002), wheat hessian fly (Dweikat *et al.,* 1997), barley yellow mosaic virus (Saeki *et al.,* 1999), barley net blotch (Molnar *et al.,* 2000) and barley scald (Penner *et al.,* 1996). Many disease resistance traits may be quantitative in nature. A good example of this is resistance to *Fusarium graminiarum* in wheat which

causes Fusarium head blight (FHB). QTLs for FHB have been located on chromosomes 2D, 3B, 4B, 5A, 6B in wheat (Somers *et al.,* 2003b; Anderson *et al.,* 2001; Buerstmayr *et al.,* 2002). Seed quality, stress resistance and other agronomic traits are also typically quantitative traits, and as such gene tagging experiments for these triats often need QTL analysis. In the cereal species there are several examples of identifying chromosome intervals controlling seed quality including barley malt (Iguartua *et al.,* 2000), barley salt tolerance (Forster 2001), wheat bread making quality and protein content (Parker *et al.,* 1999; Chee *et al.,* 2001), and rice aluminum tolerance (Nguyen *et al.,* 2001). Once the molecular markers are available as gene tags, the information and dignostic technology can be incorporated into a molecular breeding strategy to accelerate variety production or to assemble complex genotypes, which would be otherwise difficult with conventional crossing involving usual selections.

4.2. Comparative Mapping

Comparative mapping also makes use of molecular markers and has been successfully used for structural genomics research in many cereals. Now that BAC libraries are available for wheat, rice, barley and maize, comparative mapping has come to the forefront of cereal genomics because of its utility in map-based cloning. Maps with high marker density and small chromosome bins is a prerequisite for map-based cloning. In such dense maps, markers from ancestrally related species known to map in the vicinity of a known chromosome interval can be used to screen a BAC library to select clones, which will facilitate in the assembly of BAC contigs. The subject of comparative mapping in cereals is well reviewed (Bennetzen and Ramakrishna 2002; Devos and Gale 2000; Devos and Gale 1997, Bennetzen 2002; Keller and Feuillet 2000), and relationships on the basis of comparative mapping have been drawn between wheat, barley, rice, sorghum, maize, rye, foxtail millet, sugar cane and oats. The ancestral and alien species of cereals provide further templates from which collinear gene relationships can be derived (Boyko *et al.,* 1999). This list is extensive and gives researchers a powerful resource to fine map genes in species using orthologous loci.

In cereals, initially, comparative mapping was performed with RFLP markers that cross hybridized to phylogenetically related species such as foxtail millet and rice (Devos *et al.,* 1998), wheat, rye and barley (Devos *et al.,* 1993), rice and sorghum (Ventelon *et al.,* 2001), wheat and rice (Sarma *et al.,* 1999), barley and rice (Han *et al.,* 1998) and barley and wheat (Salvo-Garrido *et al.,* 2001). More specific comparative mapping and fine scale mapping has been carried out for traits such as vernalization requirement, plant height and

photoperiod response. Since these traits have a significant impact on crop yield, an understanding of their biochemical basis can have a large economic impact. Comparative mapping can add to our knowledge of gene location and accelerate fine mapping of genes for map-based cloning. Studies on flowering time (*Ppd* genes) in wheat and barley suggest them to be in homoeologous positions on group 2 chromosomes. Major genes for vernalization requirement in barley (*Sh*2), wheat (*Vrn*1) and rye (*Sp*1) are also located in homoeologous regions (Laurie 1997). A barley dwarfing gene (*Dwf*2) was mapped to 4HS in barley and is syntenic with dwarfing genes (*Rht*-B1 and *Rht*-D1) in wheat (Ivandic *et al.,* 1999). In addition, the *Vrn*-A1 (vernalization) gene on wheat chromosome 5A was shown to be syntenic with the *Hd*-6 gene (heading date / photoperiod response) on chromosome 3 of rice (Kato *et al.,* 1998). Now that the rice genome sequence is available, studies relating to various traits and genes of maize, wheat and barley to rice bring researchers closer to gene discovery and also to an understanding of the biochemical basis of traits (see Chapter 5 by Paterson and Chapter 19 by Sorrells in this book).

The high throughput sequencing technologies of the genomics era have enabled comparative mapping and gene collinearity to be studied at the DNA sequence level. DNA probes can be hybridized to BAC libraries of different species, and this allows clones to be selected, sequenced and compared. Using a barley probe, this approach was used for selecting BAC clones from chromosome 5H of barley and chromosome 3 of rice (Dubcovsky *et al.,* 2001). A similar comparative study was undertaken for *Sh2* locus in maize and sorghum (Chen *et al.,* 1997). Such studies demonstrated a great deal of microcollinearity of gene sequences, gene spacer regions, gene structure (exons – introns) and repetitive elements (Dubcovsky *et al.,* 2001).

5. USE OF MARKERS IN MOLECULAR BREEDING OF CEREALS

5.1. Gene Pyramiding

Gene pyramiding is one of the most important applications of molecular markers since it is technically feasible using molecular markers and cannot be achieved easily through conventional methods of plant breeding. Marker-assisted selection can certainly help in achieving durable disease resistance and/or high seed quality traits by assembling the required genes in one accession. A few reports of pyramiding of genes for disease resistance are also available in cereals. For instance, Tabien *et al.,* (2000) mapped and

achieved pyramiding of four genes (*Pi-tq*5, *Pi-tq*, *Pi-tq*6, *Pi-lm*2) for rice blast (*Pyricularia grisea*) resistance located on chromosomes 2, 6,12 and 11 respectively. The lines carrying different combinations of these four genes were evaluated for blast resistance. Although, the genes do not operate in an additive fashion, the resistance to blast is more durable in three and four gene combinations, so that if one gene is defeated by an evolving pathogen, there are additional genes present to sustain blast resistance. A parallel study that involved pyramiding of three genes (*Pi*1, *Piz*-5, *Pi*ta) for blast resistance in rice located on chromosomes 11, 6, and 12 respectively, was conducted by Hittalmani *et al.,* (2000). In this case, additive effects for increased resistance were noted for two and three gene combinations. Based on chromosome locations, some of the genes in these two studies are likely to be the same. Similarly, pyramiding of three genes (*Xa5, Xa13, Xa21*) for bacterial blight (*Xanthomonas oryzae*) of rice resulted in a noted increase in the level of resistance and broadening of the resistance spectrum (Singh *et al.,* 2001). In this study, the bacterial blight resistance genes were brought together using PCR-based STS (sequence tagged site) markers for each of these genes.

More complex traits such as protein content, oil content, yield and resistance against diseases like FHB in wheat, may be the other targets for trait improvement through marker-assisted gene pyramiding. Part of the difficulty in these cases is the accurate phenotyping and precise genetic analysis of genes in isolation. There is a need to correctly tag the genes and then use the gene tags to select for the correct allele combinations in hybrid plants.

5.2. Recurrent Parent Recovery

In addition to the selection of desirable alleles that confer improved disease resistance or seed quality characteristics, molecular markers can also be used successfully to select for genomic constitution, and thus restore the recurrent parent background in a backcross breeding programme. There are real advantages in using markers to restore the background of the recurrent parent, since this process can accelerate the process of transfer of a desirable gene through backcross programme. Further, where two elite parents having unique traits are used in crosses, progeny can be characterized and selected for a desirable complex mixture of elite alleles. This establishes new recombinant types that would normally be difficult to score via visual selection during conventional plant breeding.

Basic genetics suggests that individuals in a BC1F1 population would possess on average 50% fixed recurrent parent alleles and 50% of the genetic loci would segregate. The individuals of a BC2F1 population similarly would

exhibit 75% fixed recurrent parent alleles and the remaining 25% loci will segregate. Experiments in our lab on bread wheat show that there are individuals in the BC1F1 population with 68-70% fixed recurrent parent alleles. Likewise, the BC2F1 population may contain individuals with 88-91% fixed recurrent parent alleles. Thus in two backcrosses, a sufficient amount of the recurrent parent alleles can be selected using molecular markers, so that the progeny stemming from these elite plants will be highly enriched for the recurrent parent plant type and quality. Examples of advanced backcross QTL selection include Pillen *et al.* (2003), Huang *et al.* (2003), Thomson *et al*. (2003).

Gene tagging, as above, requires two basic components that structural genomics supplies. These include either a good genomic fingerprint or a high-density genetic map of the crop (i.e. 1 marker / 2-3 cM) and a high throughput genotyping platform, such as capillary electrophoresis. The most widely used markers in large-scale breeding are microsatellite markers. Microsatellite-based maps are available in wheat (Roder *et al.,* 1998), barley (Ramsay *et al.,* 2000; Macaulay *et al.,* 2001; Thiel *et al.,* 2003)), rice (Temnykh *et al.,* 2001; McCouch *et al*., 2002) and maize (Sharopova *et al.,* 2002) (for details, consult Chapter 3 in this book by Varshney *et al.*). The marker density of the map determines the ability to select for a chromosome interval using the markers flanking the interval. Reasonable genome coverage is essential to ensure that selection across the genome for recurrent parent alleles is complete. Capillary electrophoresis of microsatellites is a well established technology, with its main benefit being speed and capacity to collect thousands of data points/day. Genotyping and plant selection should be done during the vegetative stage of growth, prior to flowering. This means that informed crosses can be made and that the natural development or biology of the plant is not delayed or interrupted. Although molecular breeding may proceed at a similar pace with conventional breeding, there are important genetic gains made with marker-assisted selection.

Combining gene pyramiding and the selection for the genome of recurrent parent through genome fingerprinting are providing new opportunities in cereal variety development. Plants can be selected to possess new combinations of disease resistance genes, and populations from elite selections can be enriched for disease resistance and high seed quality. The genomics era has brought with it both the development of genotyping equipment and large-scale marker developments. When combined, these advances are providing suitable molecular breeding tools for the cereal crops see Chapter 10 by Koebner in this book).

6. SNP MARKERS AND CEREAL GENOMICS

Single nucleotide polymorphisms are basic to all types of DNA poylmorphisms except the length polymorphism due to SSRs, rDNA repeat units and similar other types of DNA sequences. They are discovered by alignment of DNA sequences from alternate alleles whereby subtle DNA sequence variations are observed. DNA polymorphism in other types of markers such as RFLP, AFLP and RAPD are also derived from SNPs, although the allelic DNA sequences of these other marker types were not known at the time of their development. SNPs are thus not a new idea; only the accumulation of EST and genomic sequences permits discovery of SNPs, which behave genetically, like any other type of codominant markers.

SNPs in humans are now known to be abundant, due to the release of the Human genome sequence. The SNP Consortium, (http://snp.cshl.org/), a public and private sector funded group, has discovered over 1.8 million SNPs in the Human genome for biomedical research. The research in plant species, including cereals, is far behind this accomplishment. However it is recognized that there is benefit from coordinating the efforts of several labs. A public consortium (http://wheat.pw.usda.gov/ITMI/2002/WheatSNP.html) has been developed for wheat, whereby contributors who discover SNPs in wheat will upload the information to a public database.

SNP discovery methods and estimation of SNP frequency have been characterized in maize (Tenaillon *et al.,* 2001; Bhattramakki *et al.,* 2002), wheat (Somers *et al.,* 2003a) and barley (Kota *et al*., 2001; Kanazin *et al*., 2002). The frequency of SNPs in the genome is calculated by the statistics, $\theta = K/aL$ where K=number of SNPs, L=length in base pairs and a = number of genotypes examined. In maize, SNPs were reported at the rate of $\theta = 0.0028$ to 0.036 along chromosome 1 (Tenaillon *et al*., 2001), while in wheat $\theta = 0.0069$ was reported among 12 genotypes (Somers *et al.,* 2003a). In a report on five genotypes of barley, however, SNP frequency in terms of θ was not given (Kanazin *et al.,* 2002), but the data when used to calculate SNP frequency, gave a value of $\theta = 0.0011$. Therefore, SNPs in wheat appear to be more abundant than in barley and less abundant than in maize. These analyses are still preliminary and many more studies will be required to completely understand the frequency and distribution of SNPs in cereal genomes.

SNPs are codominant and typically biallelic. Currently, several PCR-based methods are available for SNP detection, the most popular being the single base extension followed by electrophoresis. SNPs discovered in ESTs can

also be used to develop locus specific diagnostic tests. Markers such as these can be used in genetic and physical mapping or to select BAC clones from a library to study the full-length genes. However, the work on SNPs in cereals is still in its early phase, but the future efforts for SNP work in cereals should include the following: (i) discovery of allele mutations related to function, (ii) development of non-PCR-based polymorphism detection methods such as array platforms, (iii) high throughput genotyping on a miniature scale, and (iv) lower genotyping costs.

7. SUMMARY AND OUTLOOK

Just as the human genome and Arabidopsis genome sequences are changing biomedical and plant sciences at a rapid pace, soon the whole genome sequence of rice (Goff *et al*., 2002; Yu *et al*., 2002) will change the breeding methods that will be used for improvement of cereal crops. A further understanding of nitrogen and water use efficiency, and of the synthesis of key biochemicals for food and feed uses, is certainly on the research horizon. The rice genome sequence will aid researchers in discovering genes controlling these traits listed above and many more. The manipulation and engineering of these genes, provide exciting new prospects for cereal research.

Molecular markers have already played an important role in the success achieved in cereal genomics by identifying the genes or unique alleles conferring desirable phenotypes. The marker-assisted selection can make use of markers ranging from RFLPs to SNPs. Molecular markers also allow us study the structure and organization of the cereal genome by aligning QTLs, and through physical mapping of BAC clones. This relationship, which is linked by molecular markers, will become stronger in the future, as more genome sequences become available, more allele differences are discovered and allele detection platforms are made more widely available.

8. REFERENCES

Altschul S.F., Madden T.L., Schaffer A.A., Zhang J., Zhang Z., Miller W., Lipman D.J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucl Acid Res 1997; 25:3389-3402.

Anderson J.A., Stack R.W., Liu S., Waldron B.L., Fjeld A.D., Coyne C., Moreno-Sevilla B., Mitchell Fetch J., Song Q.J., Cregan P.B., *et al.* DNA markers for *Fusarium* head blight resistance QTL in two wheat populations. Theor Appl Genet 2001; 102: 1164-1168.

Bariana H.S., Brown G.N., Ahmed N.U., Khatkar S., Conner R.L., Wellings C.R., Haley S., Sharp P.J., Laroche A. Characterization of *Triticum vavilovii*-derived stripe rust resistance using genetic, cytogenetic and molecular analysis and its marker-assisted selection. Theor Appl Genet 2002; 104:315-320

Bennetzen J. Opening the door to comparative plant biology. Science 2002, 296:60-63

Bennetzen J.L., Ramakrishna W. Numerous small rearrangements of gene content, order and orientation differentiate grass genomes. Plant Mol Biol 2002; 48:821-827

Bhattramakki D., Dolan M., Hanafey M., Wineland R., Vaske D., Register III J.C., Tingey S.V., Rafalski, A. Insertion-deletion polymorphisms in 3' regions of maize genes occur frequently and can be used as highly informative genetic markers. Plant Mol Biol 2002; 48:539-547

Boyko E.V., Gill K.S., Mickelson-Young L., Nasuda S., Raupp W.J., Yiegle J.N., Singh S., Hassawi D.S., Frity A.K., Namuth D. *et al.* A high-density genetic linkage map of *Aegilops tauschii*, the D-genome progenitor of bread wheat. Theor Appl Genet 1999; 99: 16-26

Bryan G.J., Collins A.J., Stephanson P., Orry A., Smith J.B., Gale M.D. Isolation and characterization of microsatellites from hexaploid bread wheat. Theor Appl Genet 1997; 94:557-563

Bryan G.J., Stephenson P., Collins A., Kirby J., Smith J.B., Gale M.D. Low levels of DNA variation among adapted genotypes of hexaploid wheat. Theor Appl Genet 1999; 99:192-198

Buerstmayr H., Lemmens M., Hartl L., Doldi L., Steiner B., Stierschneider M., Ruckenbauer, P. Molecular mapping of QTL for *Fusarium* head blight resistance in spring wheat. I. Resistance to fungal spread (SFI resistance). Theor Appl Genet 2002; 104:84-91

Chao S., Sharp P.J., Worland A.J., Koebner R.M.D., Gale, M.D. RFLP-based genetic maps of homoeologous group 7 chromosomes. Theor Appl Genet 1989; 78:495-504

Chee P.W., Elias E.M., Anderson J.A., Kianian S.F. Evaluation of high grain protein QTL from *Triticum turgidum* L. var. *dicoccoides* in a adapted durum wheat background. Crop Sci 2001; 41:295-301

Chen, M., Sanmiguel, P., De-Oliveira, A.C., Woo, S.S., Zhang, H., Wing, R.A., and Bennetzen, J.L. Microcollinearity in the Sh2-homologous regions of the maize, rice and sorghum genomes. Proc Natl Acad Sci USA 1997; 94:3431-3435

Devos K.M., Gale M.D. Comparative genetics in the grasses. Plant Mol Biol 1997; 35:3-15

Devos K.M., Gale M.D. Genome relationships: the grass model in current research. The Plant Cell 2000; 12:637-646

Devos K.M., Millan T., Gale M.D. Comparative RFLP maps of homoeologous group 2 chromosomes of wheat, rye, and barley. Theor Appl Genet 1993; 85:784-792

Devos K.M., Wang Z.M., Beales J., Sasaki T., Gale M.D. Comparative genetic maps of foxtail millet (*Seratia italica*) and rice (*Oryza sativa*). Theor Appl Genet 1998; 96:63-68

Dubcovsky J., Ramakrishna W., Sanmiguel P.J., Busso C.S., Yan L., Shilogg B.A., Bennetzen J.L. Comparative sequence analysis of collinear barley and rice bacterial artificial chromosomes. Plant Physiol 2001; 125:1342-1353

Dweikat I., Ohm H., Patterson F., Cambron S. Identification of RAPD markers for 11 Hessian fly resistance genes in wheat. Theor Appl Genet 1997; 94:419-423

Forster B.P. Mutation genetics of salt tolerance in barley: an assessment of Golden Promise and other semi-dwarf mutant. Euphytica 2001; 120:317-328

Friedt W., Werner K., Pellio B., Weiskorn C., Krämer M., Ordon F. Strategies of breeding for durable disease resistance in cereals. Prog Bot 2003; 64: 138-167

Goff S.A., Ricke D., Lan T.H., Presting G., Wang R.L., Dunn M., Glazebrook J., Sessions A., Oeller P., Varma H., *et al.* A draft sequence of the rice genome (*Oryza sativa* L. ssp *japonica*). Science 2002; 296: 92-100

Gupta P.K., Varshney, R.K. The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica 2000; 113: 163-185

Gupta P.K., Varshney R.K., Sharma P.C., Ramesh B. Molecular markers and their applications in wheat breeding. Plant Breed 1999; 118: 369-390

Gupta P. K., Roy J.K., Prasad M. Single nucleotide polymorphisms (SNPs): a new paradigm in molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. Curr Sci 2001; 80: 524-113

Han R., Klienhofs A., Ullrich S.E., Kilian A., Yano M., Sasaki T. Synteny with rice: analysis of barley malting quality QTLs and *rpg4* chromosome regions. Genome 1998; 41:373-380

Hartl L., Mohler V., Zeller F.J., Hsam S.L.K., Schweizer G. Identification of AFLP markers closely linked to the powdery mildew resistance genes *Pm1c* and *Pm4a* in common wheat (*Triticum aestivum* L.). Genome 1999; 42:322-329

Hittalmani S., Parco A., Mew T.V., Ziegler R.S., Huang N. Fine mapping and DNA marker-assisted pyramiding of the three major genes for blast resistance in rice. Theor Appl Genet 2000; 100:1121-1128

Huang L., Gill B.S. An RGA-like marker detects all known *Lr21* leaf rust resistance gene family members in *Aegilops tauschii* and wheat. Theor Appl Genet 2001; 103:1007-1013

Huang X.Q., Hsam S.L.K. Zeller F.J., Wenzel G., Mohler V. Molecular mapping of the wheat powdery mildew resistance gene *Pm24* and marker validation for molecular breeding. Theor Appl Genet 2000; 101:407-414

Huang X.Q., Coster H., Ganal M.W., Roder M.S. Advanced backcross QTL analysis for the identification of quantitative trait loci alleles from wild relatives of wheat (*Triticum aestivum* L.). Theor Appl Genet 2003; 106: 1379-1389

Hussien T., Bowden R.L., Gill B.S., Cox T.S. Chromosomal locations in common wheat of three new leaf rust resistance genes from *Triticum monococcum*. Euphytica 1998; 101:127-131

Igartua E., Edney M., Rossnagel B.G., Spaner D., Legge W.G., Scoles G.J., Eckstein P.E., Penner G.A., Tinker N.A., Briggs K.G., *et al.* Marker-based selection of QTL affecting grain and malt quality in two-row barley. Crop Sci 2000; 40:1426-1433

Ivandic V., Malyshev S., Korzun V., Graner A., Borner A. Comparative mapping of a gibberellic acid-insensative dwarfing gene (*Dwf2*) on chromosome 4HS in barely. Theor Appl genet 1999; 98:728-731

Kanazin V., Talbert H., See D., DeCamp P., Nevo E., Blake T. Discovery and assay of single-nucleotide polymorphisms in barley (*Hordeum vulgare*). Plant Mol Biol 2002; 48(5-6): 529-537

Kantety R.V., Rota M.L., Matthews D.E., Sorrells M.E. Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. Plant Mol Biol 2002; 48: 501-510

Kato K., Miura H., Sawada S. Comparative mapping of the wheat *Vrn*-A1 region with the rice *Hd*-6 region. Genome 1999; 42:204-209

Keller B., Feuillet C. Collinearity and gene density in grass genomes. Trends Plant Sci 2000; 5:246-251

Kota R., Varshney R.K., Thiel T., Dehmer K.-J., Graner A. Generation and comparison of EST-derived SSR and SNP markers in barley (*Hordeum vulgare* L.). Hereditas 2001; 135:145-151

Laurie D.A., Devos K.M. Trends in comparative genetics and their potential impacts on wheat and barley research. Plant Mol Biol 2002; 48:729-740

Laurie D.A. Comparative genetics of flowering time. Plant Mol Biol 1997, 35:167-177

Macaulay M., Ramsay L., Powell W., Waugh R. A representative, highly informative genotyping set of barley SSRs. Theor Appl Genet 2001; 102-801-809

Marth G.T., Korf I., Yandell M.D., Yeh R.T., Gu Z.J., Zakeri H., Stitziel N.O., Hillier L., Kwok P.Y., Gish W.R. A general approach to single-nucleotide polymorphism discovery. Nature Genet 1999; 23:452-456

McCouch S.R.,Teytelman L., Xu Y., Lobos K.B., Clare K., Walton K., Fu B., Maghirang R., Li Z., Xing Y., *et al.* Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.). DNA Res 2002; 9: 199-207

Michelmore R.W., Paran L., Kesseli R.V. Identification of markers to disease resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using separate populations. Proc Natl Acad Sci USA 1991; 88:9828-9832

Molnar S.J., James L.E., Kasha K.J. Inheritance and RAPD tagging of multiple genes for resistance to net blotch in barley. Genome 2000; 43:224-231

Nguyen V.T., Burow M.D., Nguyen H.T., Le B.T., Le T.D., Paterson, A.H. Molecular mapping of genes conferring aluminum tolerance in rice (*Oryza sativa* L.). Theor Appl Genet 2001; 102:1002-1010

Parker G.D., Chalmers K.J., Rathjen A.J., Langridge P. Mapping loci associated with milling yield in wheat (*Triticum aestivum* L.) Mol Breed 1999; 5:561-568

Penner, G.A., *et al.,* The genetic basis of scald resistance in western Canadian barley cultivars. Euphytica 1996; 92:367-374

Pillen K., Zacharias A., Léon J. Advanced backcross QTL analysis in barley (*Hordeum vulgare* L.). Theor Appl Genet 2003; 1253-1259

Ramsay L., Macaulay M., Ivanissevich D.S., MacLean K., Cardle L., Fuller J., Edwards K.J., Tuvesson S., Morgante M., Massari A., *et al.* Simple sequence repeat-based linkage map of barley. Genetics 2000; 156: 1997-2005

Röder M.S., Korzun V., Wendehake K., Plaschke J., Tixier M., Leroy P., Ganal M.W. A microsatellite map of wheat. Genetics 1998b; 149: 2007-2023

Saeki K., Miyazaki C., Hirota N., Saito A., Ito K., Konishi T. RFLP mapping of BaYMV resistance gene *rym3* in barley (*Hordeum vulgare*). Theor Appl Genet 1999; 99:727-732

Salvo-Garrido H., Laurie D.A., Jaffe B., Snape J.W. An RFLP map of diploid *Hordeum bulbosum* L. and comparison with maps of barley (*H. vulgare* L.) and wheat (*Triticum aestivum* L.) Theor Appl Genet 2001; 103:869-880

Sanchez A.C., Brar D.S., Huang N., Li Z., Khush G.S. Sequence tagged site marker-assisted selection for three bacterial blight resistance genes in rice. Crop Sci 2000; 40:792-797

Sarma R.N., Fish L., Gill B.S., Snape J.W. Physical characterization of the homoeologous group 5 chromosomes of wheat in terms of rice linkage blocks, and physical mapping of some important genes. Genome 2000; 43: 191-198

Sharopova N., McMullen M.D., Schultz L., Schroeder S., Sanchez-Villeda H., Gardiner J., Bergstrom D., Houchins K., Melia-Hancock S., Musket T. *et al.* Development and mapping of SSR markers for maize. Plant Mol Biol 2002; 48: 463-481

Singh S., Sidhu J.S., Huang N., Vikal Y., Li Z., Brar D.S., Dhaliwal H.S., Khush G.S. Pyramiding three bacterial blight resistance genes (*xa5, xa13* and *xa21*) using marker-assisted selection into indica rice cultivar PR106. Theor Appl Genet 2001; 102:1011-1015

Somers D.J., Fedak G., Savard M. Molecular mapping of novel genes controlling Fusarium head blight resistance and deoxynivalenol accumulation in spring wheat. Genome 2003a; 49:555-564

Somers D.J., Kirkpatrick R., Moniwa M., Walsh A. Mining single nucleotide polymorphisms from hexaploid wheat ESTs. Genome 2003b; 49:431-437

Song Q.J., Fickus E.W., Cregan P.B. Characterization of trinucleotide SSR motifs in wheat. Theor Appl Genet 2002; 104:286-293

Tabien R.E., Li A., Paterson A.H., Marchetti M.A., Stansel J.W., Pinson S.R.M. Mapping four major rice blast resistance genes from 'Lemont' and 'Teqing' and evaluation of their combinatorial effect for field resistance. Theor Appl Genet 2000; 101:1215-1225

Temnykh S., DeClerck G., Lukashova A., Lipovich L., Cartinhour S., McCouch S. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. Genome Res 2001; 11:1441-1452

Tenaillon M.I., Swakins M.C., Long A.D., Gaut R.L., Doebley J.F., Gaut B.S. Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. Mays L.). Proc Natl Acad Sci USA 2001; 98: 9161-9166

Thiel T., Michalek W., Varshney R.K., Graner A. Exploiting EST databases for the development of cDNA derived microsatellite markers in barley (*Hordeum vulgare* L.). Theor Appl Genet 2003; 106: 411-422

Thomson M.J., Tai T.H., McClung A.M., Lai X.H., Hinga M.E., Lobos K.B., Xu Y., Martinez C.P., McCouch S.R. Mapping quantitative trait loci for yield, yield components and morphological triats in an advanced back cross population between *Oryza rufipogon* and the *Oryza sativa* cultivar Jefferson. Theor Appl Genet 2003, DOI 10.1007/s00122-003-1270-8

Varshney R. K., Thiel T., Stein N., Langridge P., Graner A. *In silico* analysis on frequency and distribution of microsatellites in ESTs of some cereal species. Cell and Mol Biol Lett 2002; 7**:** 537-546

Ventelon M., Deu M., Garsmeur O., Doligez A., Ghesquiere A., Lorieux M., Rami J.F., Glazmann J.C., Grivet L. A direct comparison between the genetic maps of sorghum and rice. Theor Appl Genet 2001; 102-379-386

Vos P., Hogers R., Bleeker M., Reijans M., van de Lee T., Hornes M., Frijters A., Pot J., Peleman J., Kuiper M., ***et al.,*** AFLP : a new technique for DNA fingerprinting. Nucl Acid Res 1995; 23:4407-4414.

Ware D., Jaiswal P., Ni J.J., Pan X., Chang K., Clark K., Teytelman L., Schmidt S., Zhao W., Cartinhour S., *et al.,* Gramene: a resource for comparative grass genomics. Nucl Acid Res 2002; 30:103-105

Williams J.G.K., Kubelik A.R., Livak K.J., Rafalski J.A., Tingey S.V. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucl Acid Res 1990; 18:6531-6535

Yu J., Hu S., Wang J., Wang G., Li S.G., Wong K.S.G., Liu B., Deng Y., Dai L., Zhou Y., Zhang X., *et al.* A draft sequence of the rice genome (*Oryza sativa* L. sspl. *indica*). Science 2002; 296:79-92

Chapter 3

MOLECULAR MAPS IN CEREALS: METHODOLOGY AND PROGRESS

Rajeev K. Varshney[1], Viktor Korzun[2] and Andreas Börner[1,*]

[1]*Institut for Plant Genetics and Crop Plant Research (IPK), Corrensstrasse 3, D-06466 Gatersleben, Germany* and [2]*Lochow-Petkus GmbH, PF 1197, D-29296 Bergen, Germany*
[*]Author for correspondence: boerner@ipk-gatersleben.de

1. INTRODUCTION

Cereals provide for our major food crops, and therefore have been a subject of detailed genetic and cytogenetic studies during major part of the last century. These studies led to the preparation of linkage maps, which were also assigned to individual chromosomes, thus leading to the construction of chromosome maps in all major cereals. In some cases, the availability of cytogenetic stocks (e.g. deletion stocks in bread wheat) also allowed construction of physical maps. In the past, a major limitation in the construction of genetic maps has been the non-availability of mutants for majority of individual genes, so that only handful of genes could be mapped. However, during 1980s, the availability of molecular markers and the high level of DNA polymorphism, which they detect, led to renewed emphasis on genetic and physical mapping of genomes in many plant/animal systems, and cereals were no exception. Consequently, not only genetic and physical maps were constructed for all major cereal genomes, but these maps were also put to a variety of uses, so that crop improvement programmes are now undergoing a paradigm shift, making use of genes and technologies, hitherto not available to plant breeders. With the advent of genomics, the physical maps have also been found useful for high quality whole genome sequencing (Sasaki and Burr, 2000). For construction of these molecular maps, a variety of molecular markers have been used, which have received detailed treatment elsewhere (for molecular markers, consult in this book, Chapter 2 by Somers).

P.K. Gupta and R.K. Varshney (eds.), Cereal Genomics, 35-82.

The molecular markers that have been used for construction of molecular maps in cereals can be broadly classified in three groups, the first generation markers, the RFLPs (restriction fragment length polymorphisms) and RAPDs (randomly amplified polymorphic DNA), the second generation markers, the SSRs (simple sequence repeats or microsatellites) and AFLPs (amplified fragment length polymorphisms) and the third generation markers, the SNPs (single nucleotide polymorphisms) and InDels (insertion-deletion) (for a reveiew see Gupta *et al.,* 2002b). In addition to above, some other classes of molecular markers (i.e. derivatives of RFLPs, SSRs, AFLPs) such as STSs (sequence tagged sites), ISSRs (inter simple sequence repeats), SAMPL (selective amplification of microsatellite polymorphic loci), etc. have also been used. More recently, EST (expressed sequence tag)-based markers (EST-SSRs and EST-SNPs) are also being developed in all major cereals (see Sreenivasulu *et al.,* 2002). All these marker types except the recently developed SNPs have been utilized for the construction of molecular maps, and efforts are being made to construct SNP maps also in all major cereals. In this chapter, we first briefly describe the methods involved in the construction of these molecular maps, and then discuss the present status of these maps (including transcript maps) and the future prospects of developing high-density maps, which are needed both for map based cloning and marker-assisted selection (MAS). In addition, towards the end of this chapter, we also describe some special issues of applying molecular techniques to genome analysis and molecular breeding of cereal species. The various uses of these maps will be dealt with elsewhere in this book.

2. MOLECULAR GENETIC MAPS

Molecular markers detect both sequence polymorphisms (e.g. SNPs, resulting in RFLPs, RAPDs, AFLPs, etc.) and length polymorphisms (polymorphisms due to length variation of a sequence, as in SSRs and sometimes in RFLPs also), which are ubiquitous and abundant in all living organisms. These polymorphisms would segregate, majority of times, in a Mendelian manner, so that the conventional basis of linkage and recombination can be used for constructing these maps like the classical maps prepared during the middle of the last century. A major advantage of molecular mapping is the possibility of analyzing a large numbers of DNA-markers in a single mapping population. However, the systematic construction of these maps requires a mapping population and software, which facilitates the construction of map utilizing the large amount of data that is generated using the molecular genotyping of the mapping population.

2.1. Mapping Populations

Molecular markers are used for constructing genetic maps by recording co-segregation of markers in defined populations. Several types of mapping populations can be derived from crosses involving any two diverse parents. For instance, an F_2 population or backcross population can be derived from F_1 plants through selfing or backcrossing them to one of the parents; recombinant inbred lines (RILs) can be derived by single seeds descent for at least five or more generations, and doubled haploids (DHs) can be derived from haploid obtained from F_1 plants through anther/egg cell/ovule culture or distant hybridization. The simplest of these mapping populations are the F_2 populations or the backcross (BC) populations. For the majority of cereal species, populations such as these are easy to construct although sterility in the F_1 hybrid may limit some combinations of parents, particularly in wide crosses. A major drawback of using F_2 and BC populations is that they are ephemeral, that is, seed derived from selfing these individuals will not breed true (Young, 2001). The RILs and DHs, on the other hand, are immortal and can be permanently maintained and evaluated in repeated experiments. RILs have an additional advantage of being the product of several meioses, so that each RIL contains a different combination of linkage blocks from the original parents. However, generation of RILs is time-consuming, and some regions of genome tend to stay heterozygous longer than expected from theory (Burr and Burr, 1991). Therefore, in several mapping projects, DHs were preferred, because they can be used for linkage mapping with many of the same advantages of RILs and take less time in production (Heun *et al.,* 1991). In addition to the above, some other mapping populations such as recombinant inbred substitution lines (RISLs), recurrent intermated populations, etc. were also used for increasing the efficiency and genetic resolution of genome mapping in cereals (Araki *et al.,*1999; Rousset *et al.,* 2001;Sharopova *et al.,*2002). For instance in maize, an intermated population has been generated from a common population (B73 × Mo17) after five generations of intermating. Genotyping of this population before and after intermating with the same set of 190 RFLP loci resulted in nearly a four-fold increase in the genetic map distance and increased the potential for improved genetic resolution in 91% of the intervals evaluated (Lee *et al.,* 2002).

In cereals, sometimes only a few markers could be mapped with a specific segregating population (as described above) due to low levels of polymorphism between the parents of the mapping population. In wheat, to overcome this problem, either synthetics, created by combining tetraploids (A and B genomes) with *Aegilops tauschii* (D genome) were used in crosses, or else mapping of individual genomes was done at the diploid level, so that

mapping populations were constructed using the diploid progenitors *Aegilops tauschii* – (D genome) (Boyko *et al.,* 1999) and *Triticum monococcum* – (A genome) (Dubcovsky *et al.,* 1996). In some cases, wild species were included in crosses with cultivated species for preparing mapping populations. For instance, in barley *Hordeum spontaneum* was crossed with *Hordeum vulgare* (Ramsay *et al.,* 2000; Chang *et al.,* 2001) and in case of rice, *Oryza glaberrima* or *Oryza glumaepatula* was crossed with *Oryza sativa* (Lorieux *et al.,* 2000; Brondani *et al.,* 2001). Use of such strategies enhanced the level of polymorphism of markers thus facilitating the mapping of a much larger number of markers.

2.2. Computer Programmes

In principle, linkage mapping with DNA markers does not differ from mapping with classical genetic markers, so that genetic distances between DNA markers are based on frequencies of genetic crossing over, and are represented on the genetic map in centiMorgans (cM). However, the number of markers, to be analyzed in a single mapping population used for mapping, can reach several thousands, thus necessitating the use of computer programmes; many such programmes have been developed and used in the past for constructing genetic maps in a variety of plant genomes.

The most widely used mapping software is 'Mapmaker' (Lander *et al.,* 1987), which is based on the concept of a LOD score, the "log of the odds-ratio" (Morton, 1955). Mapmaker performs multipoint analysis of many linked loci, which is essential to sort out the many different possible marker orders. In the same species, several maps can be prepared using different mapping populations. In such cases it is useful to relate the maps derived from different mapping populations, to produce an integrated or consensus map. The computer programme 'JoinMap' is often used for this purpose (Stam, 1993). 'Map Manager' is another programme which helps to keep track of markers data in a population of interest (Manly and Olson, 1999).

2.3. Whole Genome Genetic Maps in Cereals

Over the last two decades, using a variety of molecular markers high-density molecular linkage maps have been developed for all major cereal species. A summary of these maps is presented in Table 1. Phillips and

Table 1. A list of some important genetic (including transcript) maps* available in cereals

Map type /crop	Population used for mapping	Number of loci mapped	Genetic map length (cM)	Reference
RFLP maps				
Barley	DHs (Proctor × Nudinka)	154	1,091	Heun *et al.* (1991)
Barley	DHs (Igri × Franka; *Hordeum vulgare* ssp. Vada × *H. vulgare* ssp. *spontaneum*line 1B-87)	226	1,453	Graner *et al.* (1991)
Barley	DHs (Steptoe × Morex)	295	1,250	Kleinhofs et al (1993)
Barley	DHs (Harrington × TR306)	898	1,060	Kasha *et al.* (1994)
Barley	F2s (Ko A × Mokusekko)	222	1,389	Miyazaki *et al.* (2000)
Maize	F2s (CO159 × T× 303)	215	-	Gardiner *et al.* (1993)
Maize	F2s (CO159 × T× 303)	92	-	Chao *et al.* (1994)
Maize	Intermated RILs (B73 × M017)	180	5,917	Lee *et al.* (2002)
Oat	F2s/ F3s (*Avena atlantica* (M66/3) × *A.. hirtula* (Cc7050 - CAV4490)	192	614	O'Donoughe *et al.* (1992)
Oat (2×)	F2s (*Avena strigosa* Shreb × *A. wiestii* Sleud)	208	2,416	Rayapaty *et al.* (1994)
Oat	RILs (*Avena byzantina* cv Kanota × *A. sativa* cv Ogle)	561	1,482	O'Donoughe *et al.* (1995)
Oat (2×)	F2s (*Avena strigosa* CI3815 × *A. wiestii* CI1994)	181	880	Kremer *et al.* (2001)
Rice	BC lines (*Oryza sativa* × *O. longistaminata*)	726	1,491	Causse *et al.* (1994)
Rice	F2s/ F3s (*indica* var. IR24 × *japonica* marker stocks)	83	-	Yoshimura *et al.* (1997)
Rice-wild	F2s (*Oryza sativa* var. Johnson × *Zizania palustris* L.)	121	1,805	Kennard *et al.* (2000)
Rye	F2s (DS2 × R×L10)	~50	-	Devos *et al.* (1993a,b)
Rye	F2s (P87 × P105)	88	660	Korzun *et al.* (1998)
Sorghum	F2s (BSC35 × BT× 631)	71	633	Ragab *et al.* (1994)

Table 1. Continued

Sorghum	F2s (*Sorghum bicolor* × *S. propinquum*)	276	1,445	Chittenden *et al.* (1994)
Sorghum	F2s (*Sorghum bicolor* ssp. bicolor IS3620 × BT× 623*)*	190	1,789	Xu *et al.* (1994)
Sorghum	RILs (*Sorghum bicolor* BT× 623 × IS3620C*)*	323	1,347	Peng *et al.* (1999)
Wheeat (D-genome)	*T. tauschii* (TA1691 var. meyeri × TA1704 var. typica)	152	1,554	Gill *et al.* (1991)
Aegilops tauschii	F2s [*Aegilops tauschii* var. *meyeri* (TA1691) × *Ae. tauschii* var. *typica*(TA1704)]	546	-	Boyko *et al.* (1999)
Wheat (Group 1)	ITMI RILs (W7984 × Opata85)	98	146 to 344	Van Deynze *et al.* (1995)
Wheat (Group 2)	F2/F3s (Chinese Spring × SyntheticTimgalen)	114	-	Devos *et al.* (1993b)
Wheat (Group 2)	ITMI RILs (W7984 × Opata85)	173	~ 600	Nelson *et al.* (1995b)
Wheat (Group 3)	F2/F3s (Chinese Spring × SyntheticTimgalen)	~60	-	Devos *et al.* (1992) Devos and Gale (1993)
Wheat (Group 3)	ITMI RILs (W7984 × Opata85)	160	~ 660	Nelson *et al.* (1995c)
Wheat (Group 4)	ITMI RILs (W7984 × Opata85)	98	-	Nelson *et al.* (1995a)
Wheat (Group 5)	F2/F3s (Chinese Spring × SyntheticTimgalen)	~50	-	Xie *et al.* (1993)
Wheat (Group 5)	ITMI RILs (W7984 × Opata85)	118	-	Nelson *et al.* (1995a)
Wheat (Group 6)	ITMI RILs (W7984 × Opata85)	154	516	Marino *et al.* (1996)
Wheat (Group 6)	F2/F3s (Chinese Spring × Synthetic)	62	317	Jia *et al.* (1996)
Wheat (Group 7)	ITMI RILs (W7984 × Opata85)	109	-	Nelson *et al.* (1995a)
Wheat	F2s (*Triticum aestivum* var. Chinese Spring × *Triticum spelta* var. duha	197	-	Liu and Tsunewaki (1991)
Wheat	DHs (Chinese Spring × Courtot)	264	1,772	Cadalen *et al.* (1997)
Wheat- durum	RILs (*T. durum* var. Messapia × *T. turgidium* var. MG4343)	245	-	Blanco *et al.* (1998)

Table 1. Continued

SSR maps				
Barley	DHs (*Hordeum vulgare* var. Lina × *H. spontaneum* Canada Park)	242	1,173	Ramsay *et al.* (2000)
Barley	F2s (Lerche × BGRC41936), DHs (Igri × Franka)	57	840	Pillen *et al.* (2000)
Barley	Consensus map- DHs (Igri × Franka; Steptoe × Morex; OWB Dom × OWB Rec)	76	-	Thiel *et al.* (2003)
Barley	Consensus map – DHs (Igri × Franka; Steptoe × Morex)	127	-	Li *et al.* (2003)
Maize	Intermated RILs (B73 × Mo17)	978	4,906	Sharopova *et al.* (2002)
Rice	DHs (*indica* var. IR64 × *japonica* var. Azucena)	120	-	McCouch *et al.* (1997)
Rice	DHs (IR64 × Azucena, ZYQ × JX), RILs (Milyang 23 × Gihobyeo)	121	~ 1,900	Chen *et al.* (1997)
Rice	DHs (IR64 × Azucena), RILs (Milyang 23 × Gihobyeo; Lemont × Teqing)	312	1,822	Temnykh *et al.* (2000)
Rye	Consensus map- F2s (P87 × P105; N6 × N2; N7× N2; N7 × N6)	99	-	Khelestkina *et al.* (2004)
Wheat	ITMI RILs (W7984 × Opata85)	279	-	Roder *et al.* (1998b)
Wheat	F2s (Chinese Spring × Synthetic)	53		Stephenson *et al.* (1998)
Wheat	ITMI RILs (W7984 × Opata85)	55		Pestsova *et al.* (2000)
Wheat	DHs	172	-	Harker *et al.* (2001)
Wheat	ITMI RILs (W7984 × Opata85)	65	-	Gupta *et al.* (2002a)
Wheat	4 mapping populations (W7984 × Opata85, Courtot × Chinese Spring, Eureka × Renan; Arche × Recital)	533	-	Gandon *et al.* (2002)
Wheat	ITMI RILs (W7984 × Opata85) Chromosomal assignment by using nulli-tetrasomic lines	144 73	-	Song *et al.* (2002)
Wheat- durum	RILs (*Triticum durum* var. Messapia × *T. turgidium* var. MG4343)	79	-	Korzun *et al.* (1999)

Table 1. Continued

AFLP maps				
Barley	DHs (Proctor × Nudinka)	118	1,096	Becker *et al.* (1995)
Barley	RILs (L94 × Vada)	566	1,062	Qi *et al.* (1998)
Barley	DHs (Proctor × Nudinka)	511	2,673	Castiglioni *et al.* (1998)
Maize	RILs (B73 × Mo17)	1539	1,178	Vuylsteke *et al.* (1999)
Maize	F2s (D32 × D145)	1355	1,376	Vuylsteke *et al.* (1999)
Maize	F2s (B73 × A7)	246	2,057	Castiglioni *et al.* (1999)
Rice	DHs (IR64 × Azucena)	208	3,058	Maheswaran *et al.* (1997)
Rye	F2s (synthetic I0.1 lines)	71	215 (total 1100)	Saal and Wricke (2002)
Bread wheat	DHs (Garnet × Saunders)	426	-	Penner *et al.* (1998)
Wheat	ITMI RILs (W7984 × Opata85)	140	-	Hazen *et al.* (2002)
Composite maps				
Aegilops tauschii	F2s (*Aegilops tauschii* var. *meyeri* (TA1691) × *Ae. tauschii* var. *typical* (TA1704) (marker loci- defense related genes, REMAPs/ SREMAPs, IRAPs, SSRs, ISSRs, RFLPs)	732	-	Boyko *et al.* (2002)
Barley	Consensus map from 7 maps (marker loci- mainly RFLPs)	587	1,087	Langridge *et al.* (1995)
Barley	Consensus map from 4 maps (marker loci- mainly RFLPs)	880	-	Qi *et al.* (1996)
Barley	F2s (*Hordeum chilense*) (marker loci- RAPDs, SSRs, RFLPs, SCARs, STS, etc.)	123	694	Hernandez *et al.* (2001)
Barley	DHs (OWBDom × OWB Rec) (marker loci- RFLPs, RAPDs, SSRs, AFLPs)	~ 720	1,387	Costa *et al.* (2001)

Table 1. Continued

Barley	RILs (Azumamugi × Kato Nakate Gold) (marker loci AFLPs, STSs, etc.)	272	926	Mano *et al.* (2001)
Barley	DHs (*Hordeum vulgare* var. Lina × *H. spontaneum* var. Canada Park) (marker loci- IMPs)	88	-	Chang *et al.* (2001)
Maize	F2/F5-6s (I0 × F2, I252 × F2) (marker loci- ESTs and RFLPs)	275	1,765	Causse *et al.* (1996)
Maize	F2 (T×303 × CO159) (marker loci- ESTs , RFLPs, SSRs, etc.)	1736	1,727	Davis *et al.* (1999)
Maize	RILs (B73 × Mo17) (marker loci- MITE- HBr)	213	1,092	Casa *et al.* (2000)
Maize	F2 (T1 × T2)	~ 310	-	Marsan *et al.* (2001)
Maize	Several mapping populations (marker loci- RFLPs and SSRs)	> 1800	-	http://www.maizemap.org/ maps.htm
Oat (6×)	RILs (Kanota × Ogle; Clintland64 × IL86-5698) (marker loci- AFLPs and RFLPs)	300	2,351	Jin *et al.* (2000)
Oat (6×)	RILs (Ogle × TAM O-301) (marker loci- RFLPs, AFLPs, RAPDs, STSs, etc.)	441	2,049	Portyanko *et al.* (2001)
Rice	BCs (*Oryza sativa* × *O. longistaminata*) (marker loci- cDNA/RFLPs, SSRs, etc.)	726	1,491	Causse *et al.* (1994)
Rice	F2s (Nipponbare × Kasalath) (marker loci- ESTs, RFLPs, RAPDs, etc.)	1383	1,575	Kurata *et al.* (1994)
Rice	F2 population (Nipponbare × Kasalath) (marker loci- ESTs, RFLPs, RAPDs, etc.)	2275	1,522	Harushima *et al.* (1998)
Rice	RILs (Milyang 23 × Gihobyeo) (marker loci- AFLPs, RFLPs, SSRs, etc.)	~530	1,814	Cho *et al.* (1998)

Table 1. Continued

Rice	BCs (*Oryza sativa* × *O. glaberrima*) (marker loci- SSRs, STSs, AFLPs, RAPDs, etc.)	129	1,923	Lorieux *et al.* (2000)
Rice	BCs (*Oryza glumaepatula* RS-16 × *O. sativa* BG-90-2) (marker loci- SSRs, STSs)	162	1,500	Brondani *et al.* (2001)
Rice	DHs (IR64 × Azucena), BCs (*Oryza sativa* × *O. longistaminata*) (marker loci- RFLPs, SSRs)	630	1,491	McCouch (2001)
Rice	F2s (Nipponbare × Kasalath) (marker loci- RFLPs, RAPDs, STSs, etc.)	3267	-	http://rgp.dna.affrc.go.jp/publicdata/geneticmap2000/index.html
Rice	F2s (Nipponbare × Kasalath) (marker loci- STSs, CAPS, etc.)	332	-	http://rgp.dna.affrc.go.jp/publicdata/caps/index.html
Rye	F2s (E × R) (marker loci- RFLPs, RAPDs)	99		Loarce *et al.* (1996)
Rye	F2s (synthetic I0.1 lines I-line × a genebank accession) (marker loci- RFLPs, RAPDs)	92	760	Senft and Wricke (1996)
Rye	consensus map from 13 mapping populations (marker loci- mainly RFLPs)	415	-	Börner and Korzun (1998)
Rye	F2s (P87 × P105) (marker loci- RFLPs, SSRs, etc.)	183	1,063	Korzun *et al.* (2001)
Rye	F2s (UC-90 × E-line, King II × Imperial) (marker loci- RFLPs, SSRs, etc.)	184	727	Ma *et al.* (2001)
Rye	F2s (DS2 × R×L10) (marker loci- RFLPs, RAPDs, etc.)	282	1,140	Masojć *et al.* (2001)
Sorghum	RILs (IS2807 × 379; IS2807 × 249) (marker loci- RFLPs, AFLPs)	443	1,899	Boivin *et al.* (1999)

Table 1. Continued

Sorghum	RILs (BT×623 × IS3620C) (marker loci- RFLPs, SSRs)	470	1,406	Bhattramakki *et al.* (2000)
Sorghum	RILs (B35 × T×7000) (marker loci- RFLPs, SSRs)	214	1,200	Subudhi and Nguyen (2000)
Sorghum	RILs (BT×623 × IS3620C) (marker loci- AFLPs, SSRs, RFLPs, etc.)	2926	1,713	Menz *et al.* (2002)
Wheat-durum	F2s/ F3s (*T. monococcum* ssp. *monococcum* DV92 × *T. monococcum* ssp. *Aegilopoides* C3116) (marker loci- mainly RFLPs)	335	714	Dubcovsky *et al.* (1996)
Wheat- einkorn	F2s (*T. monococcum* × *T. boeoticum* ssp. *boeoticum*) (marker loci- RFLPs, RAPDs, ISSRs)	81	-	Kojima *et al.* (1998)
Wheat-durum	RILs [*T. durum* (Messapia) × *T. turgidium* (MG4343)] (marker loci- AFLPs, RFLPs)	88	2,063 (total)	Lotti *et al.* (2000)
Wheat-durm	RILs (Jennah Khetifa × Cham1) (marker loci- RFLPs, SSRs, AFLPs, etc.)	206	3,598	Nachit *et al.* (2001)
Wheat	RILs (*Triticum aestivum* L. var. Forno × *T. spelta* L. var. Oberkulmer) (marker loci- RFLPs, SSRs)	230	2,469	Messmer *et al.* (1999)
Wheat	DHs (Cranbook × Halbred, CD87 × Katepwa, Sunco × Tasman) (marker loci RFLPs, SSRs, AFLPs,etc.)	355 to 902	-	Chalmers *et al.* (2001)
Wheat	DHs (Courtot × Chinese Spring) (marker loci- RFLP, SSRs, AFLPs)	659	3,685	Sourdille *et al.* (2003)
Wheat	F5s (Arina × Forno) (marker loci- RFLPs, SSRs)	396	3,086	Paillard *et al.* (2003)

*only maps comprising >50 loci are listed. Details and updated version of these maps are available at Graingenes website http://wheat.pw.usda.gov/ggpages/map_summary.html

Vasil (1994, 2001) also compiled details of many maps and this information is available on-line at GrainGenes (http://wheat.pw.usda.gov /ggpages/maps), which is regularly updated.

It is evident from Table 1, that genetic maps were prepared in the past mainly by using co-dominant RFLP markers in different cereals like wheat (see Gupta *et al.,* 1999), barley (see Varshney *et al.,* 2004), maize (http://www.maizemap.org/) and rice (http://rgp.dna.affrc.go.jp/Public data.html). RFLP analysis, however, is time consuming, labour intensive and is too slow for rapid evaluation of large segregating populations used in commercial breeding programmes (Gale *et al.,* 1995). Subsequently, other marker systems such as RAPDs, because of the ease of analysis, were also used for mapping (Giese *et al.,* 1994; Harushima *et al.,* 1998; Masojć *et al.,* 2001). However, RAPDs were not found suitable for preparation of genome-wide dense molecular map, since they exhibited low level of reproducibility between laboratories due to variation in PCR conditions and/or due to the use of different models of thermal cyclers (Devos and Gale, 1992; Penner *et al.,* 1993). Like RAPDs, AFLPs are also assumed to represent a dominant marker system, but these were found to be superior, due to high multiplex ratio (number of different genetic loci that may be analysed per primer pair and per gel lane), so that they were included in genetic maps in many cereals (Maheswaran *et al.,* 1997; Qi *et al.,* 1998; Castiglioni *et al.,* 1998; Vuylsteke *et al.,* 1999). During 1990s, microsatellites became the markers of choice due to a variety of attributes including their multiallelic nature, co-dominant inheritance, relative abundance and extensive genome coverage (Powell *et al.,* 1996). In the past, it was expensive and cumbersome to generate microsatellites, even though they were generated for many plant species including cereals (see Gupta and Varshney 2000; Table 1). In cereals a large number of microsatellites have been developed and mapped. For instance in wheat, several microsatellite maps are already available (Röder *et al.,* 1998b; Pestsova *et al.,* 2000; Varshney *et al.,* 2000; Gupta *et al.,* 2002) and an integrated map with ~1000 SSR loci will become available soon (D. Somers, Canada, personal communication). Similarly more than 1800 mapped microsatellite loci are available in maize (Sharopova *et al.,* 2002; http://www.agron.missouri.edu/ssr.html). In rice, a set of >2700 SSRs are available in public domain, as bioinformatics-based approach facilitated mapping of 2240 SSR loci after utilizing the draft sequence of rice generated by Monsanto (McCouch *et al.,* 2002). However, in barley, only ~700 SSR loci have been mapped (Ramsay *et al.,* 2000; Pillen *et al.,* 2000; Li *et al.,* 2003; Varshney *et al.,* unpublished). Over the last five years, emphasis has shifted to SNP markers, which are biallelic in nature and are abundant in any genome. Efforts have been initiated to develop SNP maps in barley (Kota *et al.,* 2001), maize (Batley *et al.,* 2003), rice (Nasu *et al.,* 2002), wheat

(http://wheat.pw.usda.gov/ITMI/2002/WheatSNP.html). It is believed that dense SNP maps will be available soon especially as the cost SNP assays continues to come down.

Due to availability of different marker assays, 'integrated' or 'composite' maps involving more than one type of molecular marker (particularly the SSRs, AFLPs, InDels) have been prepared (see Table 1). In some cases, molecular marker maps have included mapped genes for economic traits also. In most cereals especially barley, wheat, rice and maize, a large number of markers have been mapped in different mapping populations. Comparisons among certain regions of chromosomes mapped with common markers in different populations indicate that the order of molecular markers on the linkage maps is similar, although the distances may differ. Consequently, the construction of 'consensus maps' becomes possible by using common markers as anchors and extrapolating the positions of markers mapped between the anchors. For instance, a consensus genetic linkage map for rye chromosome 7R could be generated from seven different genetic maps (Börner and Korzun, 1998; Fig. 1). Similarly in barley, availability of common markers, mapped in different mapping population, allowed several groups to construct consensus maps (see Varshney *et al.,* 2004; Table 1). These consensus maps display higher marker densities than their individual components, which make them highly useful resources. On the other hand, the reliability of consensus maps may decrease over distances of a few centiMorgans, or where marker densities are high and the number of common markers is low. To increase the genetic resolution, Kleinhofs and Graner (2001) divided the barley genome in approximately 10 cM intervals ("BINs"). Each BIN is defined by its two flanking markers, which have been anchored in the Steptoe/Morex and the Igri/Franka maps. Such BINs, each BIN encompassing a 20 cM interval, are already available in maize (Gardiner *et al.,* 1993; Coe *et al.,* 2001). BIN maps readily allow the placement of markers mapped in different mapping populations. Although their genetic resolution is limited, they accommodate the information from a large number of maps. Thus availability of BIN maps facilitates identification of a large number of markers for a given chromosomal region.

2.4. Transcript Genetic Maps

Due to current emphasis on functional genomics in cereals, large-scale EST sequencing projects have generated a large amount of sequence data (Sreenivasulu *et al.,* 2002; Rudd, 2003). Since each EST corresponds to an

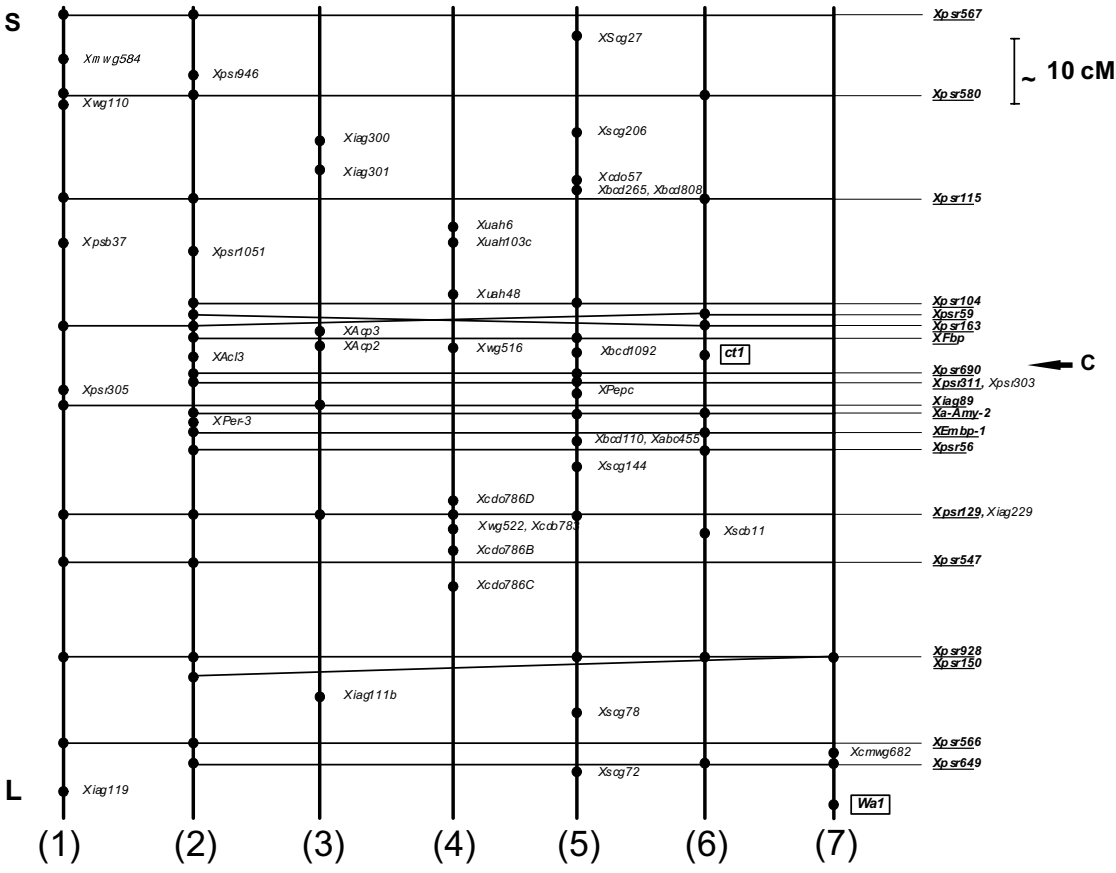


Figure 1. Preparation of a consensus linkage map of chromosomes 7R. This map was constructed (Börner and Korzun 1998) by using the following basic maps: (1) Korzun *et al.* (1998), (2) Devos *et al.* 1993a, (3) Senft and Wricke (1996), (4) Loarce *et al.* (1996), (5) Wanous *et al.* 1995, (6) Plaschke *et al.* (1995), (7) Korzun *et al.* (1997a). Mapped loci are marked with a point. The horizontal lines connect common loci used as anchor markers which are underlined. The map positions of unique loci were extrapolated. Genetic distances (roughly estimated) are given in centimorgans (cM). The gene loci are boxed. c = estimated centromere position, S = short arm, L = long arm

mRNA, these ESTs are being mapped, and will be integrated in genetic maps. Such genetic maps are termed 'functional map'/'transcript map' or 'gene map' (Schuler *et al.,* 1996). For placing ESTs (transcripts/genes) onto genetic map, ESTs can be converted into different marker assays like RFLPs, STSs, CAPSs (cleaved amplified polymorphic sequences), SSRs or SNPs. For instance, an EST can be amplified by using genomic DNA as a template with the help of PCR primers designed from this EST. The PCR products obtained thus can either be used as RFLP probes in Southern hybridization (Smilde *et al.,* 2002) or may be directly tested for length or sequence polymorphism in parents of a mapping population (Gilpin *et al.,* 1997). Sometimes, PCR products can also be digested with a set of restriction enzymes to test restriction polymorphism in parents of mapping populations for mapping ESTs as CAPS.

ESTs have also been used for developing EST-SSRs or even EST-SNPs, if ESTs for the same region are available from two or more genotypes (see Sreenivasulu *et al.,* 2002). Many software packages or algorithms are available for mining SSRs or SNPs in ESTs (Table 2) and corresponding PCR primers may be designed from the EST sequences (Kota *et al.,* 2001; Varshney *et al.,* 2002; Batley *et al.,* 2003). Thus, EST-derived SSRs or SNPs are a free by-product of EST sequencing projects. Mapping of ESTs via these marker assays is important, since QTLs or genes for different disease resistance or other agronomic traits associated with these ESTs may provide the 'candidate genes' for the trait in question. Potential of these 'candidate genes' can be further assessed in RCSLs (recombinant chromosome substitution lines) that carry variants of the 'candidate gene'. Thus direct gene markers for different traits may be generated which will be of great value in marker-assisted breeding (see later), although it does not conclusively prove function (Thomas, 2003). The above functional maps are also very useful for comparative mapping studies (see Varshney *et al.,* 2004).

'Functional maps' are already available in some cereals like rice (Harushima *et al.,* 1998) and maize (Davis *et al.,* 1999). Recently, in barley more than 1000 ESTs have been placed on the genetic map (A. Graner, Germany,

Table 2. Some algorithms/ programmes for mining SSRs and SNPs in ESTs

Programme	Source
SSR	
*MI*cro*SA*tellite	http://pgrc.ipk-gatersleben.de/misa
SSRIT	http://www.gramene.org/gramene/searches/ssrtool
SPUTNIK	http://espressosoftware.com:8080/esd/pages/sputnik.jsp
SNP	
SniPpER	http://mips.gsf.de/proj/sputnik
AutoSNP	http://www.cerealsdb.uk.net/discover.htm
Jalview	http://www.ebi.ac.uk/~michele/jalview/

personal communication) In wheat, though deletion-based physical mapping by using ESTs is in progress (discussed later), some efforts have been initiated for the genetic mapping of ESTs in the form of SSRs also (Holton *et al.,* 2002; M. Sorrells, USA, personal communication). In other cereal species also, ESTs have been screened for presence of SSRs (Hackauf and Wehling, 2002; Kantety *et al.,* 2002; Varshney *et al.,* 2002; Gao *et al.,* 2003; Gupta *et al.*, 2003). Mapping of SSR-ESTs (SSR containing ESTs) is also in progress in rye (Khelestkina *et al.*, 2004; B. Hackauf and P. Wehling, Germany, personal communication).

3. PHYSICAL MAPS

Physical maps of whole genomes are based on physical distances between genes or molecular markers measured either in terms of base pairs (megabasepairs = 10^6 base pairs) or in terms of relative physical lengths of chromosome segments. For instance, the distance of a gene/marker from the centromere may be represented as a fraction of the whole arm. While genetic maps are based on recombination frequencies, physical maps rely on direct size estimates, whether measured at the chromosome level under the microscope or else measured in terms of DNA sequence, if complete sequence of the chromosome or a part thereof is available. Physical maps provide virtually unlimited numbers of DNA markers from any chromosomal region for gene tagging and manipulation. They provide an framework for studies in genome molecular structure, organization and evolution, gene regulation, and gene interaction. Physical maps, therefore, are central tools to every type of genetic and molecular enquiry and manipulation including genome analysis, gene isolation and eventually crop

improvement. The following methods can be utilized for preparation of physical maps.

3.1. Physical Maps based on FISH

Physical maps can be generated through *in situ* hybridization (ISH), where chromosome sites that are homologous to a known labelled DNA probe can be directly visualized under the microscope. The technique initially proved useful for DNA probes that were at least a few kilobases in length. Several improvements were made to make the technique suitable for smaller DNA fragments (reviewed by Jiang and Gill, 1994; Maluszynska, 2002). For instance, fluorescence *in situ* hybridization (FISH), including DNA fibre FISH, was successfully utilized for physical mapping of centromeric and other small DNA sequences in rice (Dong *et al.,* 1998; Cheng Z.K. *et al.,* 2002).

ISH was used in many studies in cereals with different objectives. A comparison was made between the physical distances and genetic distances (between adjacent markers) in hexaploid wheat using ISH with 21 RFLP probes from linkage groups 5 and 6 (Zhang X.Q. *et al.,* 2000). The linear order and linkage relationships between DNA probes on these physical maps were generally the same as those on the RFLP-based genetic maps, but there was a significant difference between the genetic or recombinational distances on a linkage map and the physical distances obtained using ISH. The results also showed that the available linkage map did not completely cover the physical length of all the chromosomes. Similarly FISH mapping has been conducted in some other cereals using randomly selected or RFLP marker-anchored BAC (bacterial artificial chromosome) clones (Jiang *et al.,* 1995; Zwick *et al.,* 1998; Cheng *et al.,* 2001a,b). In rice, for cytological characterization of the genome and for identification of each chromosome arm, a set of 24 chromosome arm-specific BACs was used (Cheng *et al.,* 2001a). A standardized rice karyotype was also constructed which was anchored by centromere-specific and chromosome arm-specific cytological landmarks. This karyotype fully matched to the rice genetic map.

The potential of fibre FISH was successfully used to determine the size of seven segmental physical gaps, measuring 30 to192kb, and two telomere gaps on rice chromosome 10, measuring 80 and 30kb (The Rice Chromosome 10 Sequencing Consortium, 2003). Some details of physical mapping, using ISH/ FISH technology, are summarized in Table 3.

Table 3. Some examples of physical mapping in cereals using ISH/FISH

Cereal specie	Probes used for ISH/FISH	Target region	Reference
Aegilops	pSc119.2, pAs1, PSR907	Wheat-alien breakpoints (BPs) along the 3 BS and 3 DS arms	Biagetti *et al.* (1999)
Barley	BAC clones	Telomere	Lapitan *et al.* (1997)
Barley	Germin-like cDNAs with 26 BAC clones	Chromosomes 2H, 3H, 4H, 7H	Druka *et al.* (2002)
Rice	Telomeres and telomere- associated satellites	Chromosomes 9, 11	Wu and Tanksley (1993)
Rice	14 RFLPs	Chromosomes 7, 8, 11, 12	Song and Gustafson (1995)
Rice	24 chromosomal arm specific BAC clones (containing 24 RFLP markers)	Cytological characterization of rice genome	Cheng *et al.* (2001a)
Rice	Chromosome 10 specific 18 BAC clones	Chromosome 10	Cheng *et al.* (2001b)
Maize	4 markers (umc105a, csu145a, Cent C, pZm4-21)	Chromosome 9	Sadder and Weber (2002)
Oats (*6x*)	*Lrk10*-like receptor kinase sequences	Linkage groups 4, 12, 5, 6, 13	Cheng D. W. *et al.* (2002)
Sorghum	BAC clones contatining markers	Chromosome 1	Islam-Faridi *et al.* (2002)
Sorghum	22 BAC clones (encompassing 10 linkage groups)	Integrated karyotyping of Sorghum	Kim *et al.* (2002)
Tritordium*	Glu-1 loci	Chromosome arms 1AL, 1 BL, 1H(ch)L	Cabrera *et al.* (2002)
Wheat	47 RFLPs	*In situ* hybridization	Chen and Gustafson (1995)
Wheat	Rice markers	Homoeologous group 5 chromsomomes	Sarma *et al.* (2000)
Wheat	*Glu-1* loci	Homoeologous group 1 long arms (1 AL, 1 BL and 1 DL)	Cabrera *et al.* (2002)
Wheat	HSP70 gene homologue	Chromosome 1 B	Francki *et al.* (2002)

*Tritordeum- an amphiploid between *Triticum turgidum* cv. *durum* and *Hordeum chilense*

3.2. Physical Maps based on Deletion Stocks

Cytogenetic stocks can also be used for generating physical maps by locating genetically mapped DNA markers to specific chromosomal segments. Different types of cytogenetic stocks, are available for this purpose, including B-A translocations in maize (Weber and Helentjaris, 1989) and maize chromosome additions to oat genome (Riera-Lizarazu *et al.,* 2000), deletion stocks in wheat (Endo and Gill, 1996) and chromosomal translocation stocks in barley (Künzel *et al.,* 2000). These stocks were extensively used for physical mapping of the genomes of these cereals, in particular wheat and barley, genomes.

In bread wheat availability of a set of more than 400 deletion stocks facilitated preparation of physical maps for all the seven homoeologous groups (reviewed by Gupta *et al.,* 1999; P.K. Gupta, India, personal communication). It was shown that the deletion lines from three chromosomes of each of the seven homoeologous groups could be pooled together, so that each of the seven consensus chromosomes representing seven homoeologous groups can be divided into approximately 62 different 'physical bins', each bin with an average size of 40 Mb. By assuming a uniform distribution of recombination, these bins each represents a segment of 10 cM on the genetic map, so that the average ratio of physical to genetic distance becomes 4 Mb/cM (Lagudah *et al.,* 2001). A consortium of 13 laboratories in USA funded by National Science Foundation, USA, is engaged in assigning 10,000 unique ESTs to physical locations on chromosomes by using deletion lines (http://wheat.pw.usda.gov/NSF/progress_mapping.html, Qi *et al.,* 2003). When physical mapping data were used to assess organizational and evolutionary aspects of the wheat genome, it was found that recombination has played a central role in the evolution of wheat genome structure. The gradients of recombination rates along chromosome arms promoted more rapid rates of genome evolution in distal, high-recombination regions (hot spots of recombination) than in the low recombination proximal, regions (Akhunov *et al.,* 2003). In another project in France, a total of 725 microsatellite loci were assigned to 94 breakpoints in a homozygous (88 distal delitions, 6 interstitial) and 5 in a heterozygous state representing 159 delition bins with an average of 4.97 SSR/bin (Sourdille *et al.,* 2004). Assignment of ESTs and genetically mapped SSRs to deltion bins in above studies will be useful not only for deletion stock verifications but also for allocating associated QTLs to deletion bins as numerous ESTs that could be potential candidate genes have been assigned.

Table 4. A summary of physical mapping in wheat and barley using various cytogenetic stocks

Cereal species	Marker loci mapped	Cytogenetic stocks used	Reference
Barley (whole genome)	301 STSs	240 TLs	Künzel *et al.* (2000)
Barley (chromosome 7H)	28 STSs, 17 AFLPs	22 TLs	Serizawa *et al.* (2001)
Barley (chromosome 3H)	24 SSRs	14 TLs	Künzel and Waugh (2002)
Wheat (Homoeologous group 1)	19 RFLP	18 DLs	Kota *et al.* (1993)
Wheat (Homoeologous group 1)	50 RFLPs	56 DLs	Gill *et al.* (1996a)
Wheat (Homoeologous group 2)	30 RFLPs	21 DLs	Delaney *et al.* (1995a)
Wheat (Homoeologous group 2)	43 SSRs	25 DLs	Roder *et al.* (1998a)
Wheat (Homoeologous group 3)	29 RFLPs	25 DLs	Delaney *et al.* (1995b)
Wheat (Homoeologous group 4)	40 RFLPs	39 DLs	Mickelson-Young *et al.* (1995)
Wheat (Homoeologous group 5)	155 RFLPs	65 DLs	Gill *et al.* (1996b)
Wheat (Homoelogous group 5)	245 RFLPs, 3 SSRs	36 DLs	Faris *et al.* (2000)
Wheeat (short arm of homoeologous group 5)	100 RFLPs	17 DLs	Qi and Gill (2001)
Wheat (chromosome 5A)	22 RFLPs	19 DLs	Ogihara *et al.* (1994)
Wheat (Homoeologous group 6)	24 RFLPs	26 DLs	Gill *et al.* (1993)
Wheat (Homoeologous group 6)	210 RFLPs	45 DLs	Weng *et al.* (2000)
Wheat (Homoeologous group 6-short arm)	82 RFLPs	14 DLs	Weng and Lazar (2002)
Wheat (Homoeologous group 7)	16 RFLPs	41 DLs	Werner *et al.* (1992)
Wheat (Homoeologous group 7)	91 RFLPs, 6 RAPDs	54 DLs	Hohmann *et al.* (1995)
Wheat (chromosomes 6B, 2D and 7D)	16 SSRs	13 DLs	Varshney *et al.* (2001)
Wheat (1 BS)	24 AFLPs	8 DLs	Zhang H.N. *et al.* (2000)
Wheat (chromosome 4DL)	61 AFLPs, 2 SSRs, 2 RFLPs	8 DLs	Milla and Gustafson (2001)
Wheat (chromosome arm 1BS)	22 expressed sequences	DLs	Sandhu *et al.* (2002)
Wheat (whole genome)	7,697 unique ESTs	101 DLs	http://wheat.pw.usda.gov/NSF/progress_mapping.htm ; Qi *et al.* (2003)
Wheat (whole genome)	725 SSRs	159 DLs	Sourdille *et al.* (2004)

TLs= translocation lines ; DLs= delition lines

In barley, translocation breakpoints were used for the preparation of a physical map for all the seven chromosomes (Künzel *et al.,* 2000). Deletion-based physical mapping has also been conducted in barley for some chromosomes including 7H (Serizawa *et al.,* 2001). The status and future prospects on physical mapping of the barley genome has been discussed in a recent review (Varshney *et al.,* 2004), and available information on physical mapping in wheat and barley is summarized in Table 4. Based on physical mapping of wheat and barley by using cytogenetic stocks, it has been speculated that the Triticeae genomes contain gene-rich regions (Sandhu and Gill 2002; for detail see Chapter 12 by K.S. Gill in this book).

3.3. Physical Maps Based on Contigs

The availability of genome wide DNA-contigs and their physical mapping has been a prerequisite for high quality sequencing of the genomes of model organisms, Arabidopsis and rice (TAGI 2000; Sasaki and Burr, 2000). DNA contigs can be assembled and physical maps prepared (through fingerprints of the BACs) using large insert DNA clones, such as yeast artificial chromosomes (YACs) and bacterial artificial chromosomes (BACs). The above contig assembly is dependent on a high density genetic map and/or high-quality large insert DNA libraries. Physical maps generated through contig assembly are then used to find the minimum tilling path for sequencing.

As a prerequisite for whole genome sequencing (WGS), physical maps using BACs, YACs and contigs have been prepared in several crops including some of the cereals. Among cereals, YAC libraries were prepared initially in maize (Edwards *et al.,* 1992), barley (Kleine *et al.,* 1993) and rice (Umehara *et al.,*1995). These libraries were used for a number of studies but their general use has been limited by the high frequency of chimeric and unstable clones. Therefore, BAC libraries became popular due to ease of handling, relative simplicity in their development and a low frequency of chimeric clones. BAC libraries in many cereals have been constructed using Texas A & M GENEfinder Genomic Resources (http://hbz.tamu.edu/bacindex.html) and the BAC/EST Resource Centre at Clemson University Genomics Institute (http://www.genome.clemson.edu/groups/bac/). A gene-enriched BAC library has also been prepared in maize for cloning allele-specific fragments (Fu and Dooner, 2000). In hexaploid wheat, a library in Chinese Spring has been constructed using a newly developed transformation-competent artificial chromosome (TAC) vector, pYLTAC17 (Liu *et al.,*2000). These libraries are useful resources for physical mapping, positional cloning, WGS, genomic structural analysis and comparisons of

specific regions in different cereal species (for details see Chapter 11 by Stein and Graner, Chapter 13 by Yu and Wing).

In an Chapter 13 of this book, the methodology for preparation of contig-based physical map is described by Yu and Wing. The physical maps that have been prepared using YAC and BAC clones are summarized in Table 5. In some cases BAC libraries have been screened with genetically mapped ESTs or EST-derived markers, so that the physical maps can be compared with EST maps or transcript maps (discussed above). For example, for the preparation of a physical map, the CUGI (Clemson University Genomics Institute) collected the fingerprint data of two rice (Nipponbare) BAC libraries (20 fold coverage) (Soderlund *et al.,* 2000) and assembled contigs from 127,459 BAC end sequences (Mao *et al.,* 2000). With the availability of Monsanto working draft of the rice genome (Barry, 2001) and fingerprinted contig map from CUGI, a comprehensive physical map of the entire rice genome was prepared (Chen *et al.,* 2002). Using genetically mapped markers, most of the rice genome (~90.6%) was anchored through overgo hybridization, DNA gel blot hybridization and *in silico* anchoring. This physical map consists of 66,384 fingerprinted BAC clones (including 2,278 singletons) representing 20 fold coverage of the genome (http://www.genome.clemson.edu/projects/rice/fpc/integration). Simultaneously in an EST-project at RGRP (Rice Genome Research Programme), Japan, specific primers were designed for 6,713 unique (non-redundant) ESTs derived from 19 cDNA libraries. Subsequently, these primers were screened against 4,387 YAC clones and a comprehensive YAC-based rice transcript map was prepared. This map contains 6,591 EST sites and covers 80.8% of the genome (Wu *et al.,* 2002). In another recent study, 28,000 cDNA clones were physically mapped onto a *japonica* rice (Kikuchi *et al.,* 2003). As part of the International Rice Genome Sequencing Project (IRGSP), a fine physical map of the rice (*Oryza sativa japonica* Nipponbare) chromosome 4 has been prepared using 114 BAC clones from a taxonomically related subspecies *Oryza sativa indica* Guangluai 4 with 182 RFLP and 407 EST markers (Zhao *et al.,* 2002). In another recent study, rice sequence data from 2,251 ordered BAC/PAC clones was compared with 4,485 wheat ESTs. This study suggested that numerous translocations will complicate the use of rice as a model for cross-species transfer of information (Sorrells *et al.,* 2003).

Physical mapping by anchoring BAC clones with markers in cereals other than rice is also in progress. For instance, in maize 2,036 Mb of the 2500 Mb has already been covered (release 1/27/03) (http://www.maizemap.org /iMapDB/iMap.html; http://www.genome.arizona.edu/fpc/maize/; Coe *et*

Table 5. Contig-based physical mapping in some cereals

Cereal species	Approach	Clones (library) used	Markers used	Genome coverage and contigs	Reference
Rice (chromosome 1)	Screening YAC library with markers	476 YAC clones found positive	182 markers	284 YACs defined 69 contigs, coverage 60% of the chromosome length	Wang *et al.* (1996)
Rice	BAC fingerprinting	20, 682 BAC clones used	565 (RFLPs, SSRs, cDNAs and anchor probes)	120 kb resolution contig map, 631 contigs, genome covered 398 Mb (92%)	Hong *et al.* (1997)
Rice	Screening YAC library with markers	7,000 YAC clones used; 2,443 YAC clones found positive	1,285 (RFLPs and RAPDs)	222 Mb (52%)	Kurata *et al.* (1997)
Rice (chromosome arms 11 S and 12 S)	Screening YAC library with markers	7,000 YAC clones used; 38 YAC clones were identified as positive clones	46 genetic markers	Chromosome arm 11S- 2.09 Mb/ 2.51 Mb; chromosome arm 12S- 2.29 Mb/ 2.48 Mb	Wu *et al.* (1998)
Rice	*In silico* anchoring	80,143 BAC end sequences (54.2 Mb, 11.8% of rice genome)	2,152 DNA markers (spanning sequence length of 0.78 Mb) used and 418 markers were anchored to BAC clones	0.09 Mb (11.5% of total marker length)	Yuan *et al.* (2000)
Rice	BAC fingerprinting	21,087 BAC clones used	16 DNA markers associated with 2 or more contigs were used for analysis	298 BAC contigs; genome covered 419 MB (95%)	Tao *et al.* (2001)

Table 5. Continued

Rice	Screening YAC library with DNA markers	7,606 YAC clones used; 1,892 YAC clones identified as positive	1,439 DNA markers	297 YAC contigs and 142 YAC islands; genome covered 270 Mb (63%)	Saji *et al.* (2001)
Rice (Chromosome 9)	Screening BAC library	6 BAC ends, 1 YAC end	3 RFLPs	6.8 cM interval	Kamolsukyunyong *et al.* (2001)
Rice	Screening YAC library with ESTs	4,387 YAC clones	6,591 ESTs	384 YAC contigs; genome covered 347.3 Mb (80.8%)	Wu *et al.* (2002)
Rice	BAC fingerprinting (overgo hybridization, DNA gel blot hybridization, *in silico* anchoring)	65,287 BAC clones	1,704 markers (3,199 probes)	362.9 Mb (90.6%)	Chen *et al.* (2002)
Rice (Chromosome 4)	Screening BAC library with RFLPs and *in silico* anchoring of ESTs with BAC-end sequence database	566 BAC clones identified positive; 13,000 BAC-end sequences were used for *in silico* anchoring	182 RFLPs, 407 ESTs	11 contigs with 34.5 Mb (94% of estimated chromosome size)	Zhao *et al.* (2002)
Sorghum	Screening of BAC pools	2,400 BACs	32 different AFLP primer combinations	3,366 contigs, each containing an average of 5 BACs	Klein *et al.* (2000)
Sorghum	Screening of BAC pools with RFLP probes	38,016 BAC clones were used; 550 BAC clones were identified as positive	156 probes (160 loci)	103 contigs containing an average of 1.6 markers and 5.3 BACs)	Draye *et al.* (2001)

al., 2002; Cone *et al.,* 2002; Yim *et al.,* 2002). Efforts are also underway in sorghum (http://www.genome.arizona.edu/fpc/sorghum/; Klein *et al.,*2000; Draye *et al.,* 2001), and in the D-genome of wheat (http://wheat.pw.usda.gov/PhysicalMapping/). However, physical mapping of the complete hexaploid wheat genome using large insert libraries has yet to be undertaken. In barley some preliminary work has been conducted in this direction after anchoring BAC clones by using EST-derived SSR markers (Varshney *et al.,* unpublished results). *al.,* 2000). With the availability of Monsanto working draft of the rice genome (Barry, 2001) and fingerprinted contig map from CUGI, a comprehensive physical map of the entire rice genome was prepared (Chen *et al.,* 2002). Using genetically mapped markers, most of the rice genome (~90.6%) was anchored through overgo hybridization, DNA gel blot hybridization and *in silico* anchoring. This physical map consists of 66,384 fingerprinted BAC clones (including 2,278 singletons) representing 20 fold coverage of the genome (http://www.genome.clemson.edu/projects/rice/fpc/integration). In parallel, in an EST-project at RGRP (Rice Genome Research Programme), Japan, specific primers were designed for 6,713 unique (non-redundant) ESTs derived from 19 cDNA libraries. Subsequently, these primers were screened against 4,387 YAC clones and a comprehensive YAC-based rice transcript map was prepared. This map contains 6,591 EST sites and covers 80.8% of the genome (Wu *et al.,* 2002). In another recent study, 28,000 cDNA clones were physically mapped onto a japonica rice (Kikuchi *et al.,* 2003). As part of the International Rice Genome Sequencing Project (IRGSP), a fine physical map of the rice (*Oryza sativa japonica* Nipponbare) chromosome 4 has been prepared using 114 BAC clones from a taxonomically related subspecies *Oryza sativa indica* Guangluai 4 with 182 RFLP and 407 EST markers (Zhao *et al.,* 2002). In another recent study, rice sequence data from 2,251 ordered BAC/PAC clones was compared with 4,485 wheat ESTs. This study suggested that numerous translocations will complicate the use of rice as a model for cross-species transfer of information (Sorrells *et al.,* 2003).

Physical mapping by anchoring BAC clones with markers in cereals other than rice is also in progress. For instance, in maize 2,036 Mb of the 2500 Mb has already been covered (release 1/27/03) (http://www.maizemap.org/iMapDB/iMap.html; http://www.genome.arizona.edu/fpc/maize/; Coe *et al.,* 2002; Cone *et al.,* 2002; Yim *et al.,* 2002). Efforts are also underway in sorghum (http://www.genome.arizona.edu/fpc/sorghum/; Klein *et al.,*2000; Draye *et al.,* 2001), and in the D-genome of wheat (http://wheat.pw.usda.gov/PhysicalMapping/). However, physical mapping of the complete hexaploid wheat genome using large insert libraries has yet to be undertaken. In barley some preliminary work has been conducted in this direction after anchoring BAC clones by using EST-derived SSR markers.

Anchoring of genetically mapped SSR markers to BAC clones also gave a clue about the presence of gene-rich regions in barley genome (Varshney *et al.,* unpublished results).

3.4. Novel Strategies

Although preparation of contig-based physical maps is underway in larger genomes such as maize and diploid progenitors of hexaploid wheat, full genome contig physical maps could not be developed in barley or hexaploid wheat. As an alternative, efforts are underway to establish subgenomic physical maps from radiation hybrid (RH) panels (Cox *et al.,* 1990) or by "HAPPY" mapping (Dear and Cook, 1989). These methods do not rely on the availability of BAC-contigs or cloned DNA fragments and may be suitable for the high throughput mapping of PCR-based markers even in the absence of polymorphism (Waugh *et al.,* 2002).

3.4.1. Radiation Hybrid (RH) Mapping

In human genetics, RH panels have been constructed by fusing irradiated human cells containing highly fragmented chromosomes in their nuclei with intact hamster cells. Hamster cells selected for the presence of a selectable marker from the human genome, each contains random fragments from human chromosomes. In this way a population of hamster cell lines can be developed, which contain fragments of human chromosomes. The size of the fragments determines the physical resolution, which is a function of the intensity of irradiation used. Mapping is performed on the basis of the presence or absence of PCR-amplicons (Cox *et al.,* 1990). Similar efforts have also been initiated in some cereals. For instance, transgenic barley protoplasts harbouring the bar transgene as a selectable marker was fused with tobacco protoplasts to produce radiation hybrid panels (Wardrop *et al.,* 2002). In maize, individual chromosome additions to hexaploid oat (M9, maize chromosome 9 addition line) were irradiated and a panel of 100 informative M9RHs (maize chromosome 9 radiation hybrids), with an average of 3 breaks per chromosome were prepared (Riera-Lizarazu *et al.,* 2000). This allowed mapping with a resolution of 0.5 to 1.0 Mb. RH mapping of one scs^{ae} (species cytoplasm specific) gene in durum wheat is also in progress (Hossain *et al.,* 2002; http://cropandsoil.oregonstate.edu/cgb/projects.html).

3.4.2. HAPPY Mapping

An in *vitro* version of RH mapping is popularly described as HAPPY mapping. In contrast to RH mapping, HAPPY mapping does not require any cell fusion. It is based on the preparation of a series of small aliquots from genomic DNA. Each aliquot contains less than the amount of the haploid genome, hence the term HAPPY (Haploid genome; polymerase chain reaction) mapping. The DNA is sheared either in solution or by irradiation and size fractionated. Presence of physically linked DNA segments can be identified by their co-amplification in a given aliquot. The resolution of the procedure depends on the size of the DNA fragments that are used to prepare the aliquots (Dear and Cook, 1989; for a review see Waugh *et al.,* 2002). HAPPY mapping may be superior to RH mapping, as it does not suffer from problems due to cloning artefacts, or effects of chromosome structure. Using this approach, Thangavelu *et al.* (2003) successfully constructed a high resolution physical map of 1.9 Mbp region around the FCA locus within the genome of of *Arabidopsis thaliana*, and concluded that even in large genomes like that of barley, HAPPY mapping can facilitate the construction of high resolution local physical gene maps, if not the complete genome maps.

4. USES OF MOLECULAR MAPS

Both genetic and physical maps find a variety of uses not only in breeding but also in genomics research. Since several of these uses are discussed in other chapters of this book, only a very brief account will be included in this chapter. Molecular genetic maps have been extensively used for comparative genomic studies, throwing light on genome organization in grasses in general and in cereal crops in particular. The molecular genetic maps are also used for the identification of quantitative trait loci (QTLs) for a number of morphological, physiological and economic traits in several cereals. The QTLs not only help in marker-assisted selection for cereal breeding, but also facilitate the study of changes that the cereal genomes have undergone during breeding and selection. QTL analysis along with transcript maps may also be used for the identification of candidate genes for specific QTLs. Physical maps provide a large number of DNA markers from any chromosomal region for gene isolation. They also provide a framework not only for studies on structure, organization and evolution of the genome, but also for studies on gene regulation and gene interaction (Akhunov *et al.,* 2003; Sorrells *et al.,* 2003). Thus, physical and genetic maps are central to research involving genome sequencing and analysis, gene isolation, and crop

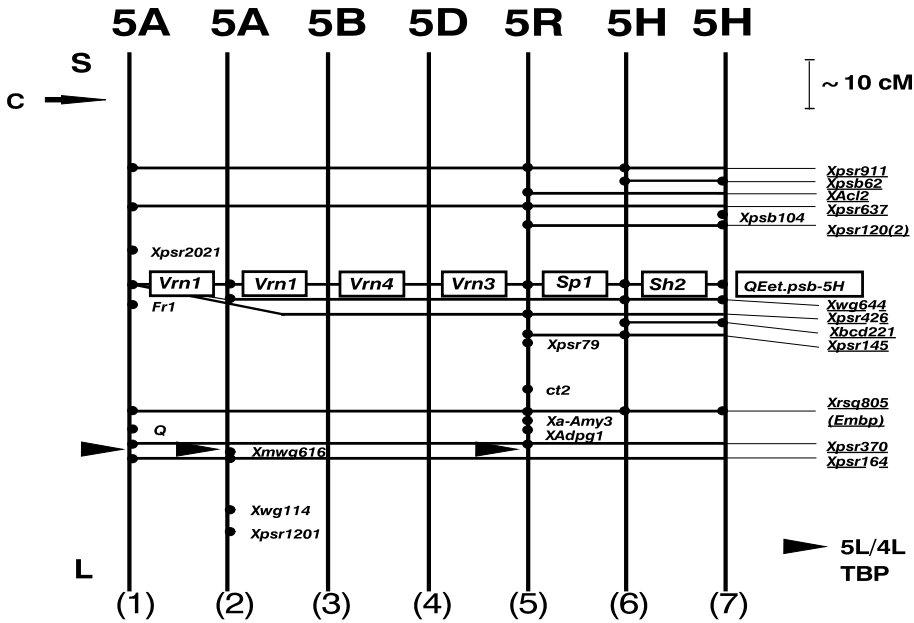


Figure 2: Comparative location of genes determining vernalisation response on chromosomes 5A, 5B and 5D of wheat, 5R of rye and 5H of barley as published by Börner *et al.* (1999). The following basic maps were used: (1) Galiba *et al.* (1995), (2) Korzun *et al.* (1997b), (3, 4) McIntosh *et al.* (1998), (5) Plaschke *et al.* (1993), (6) Laurie *et al.* (1995), (7) Bezant *et al.* (1996). Mapped loci are marked with a point. The connecting lines between chromosomes indicate common loci which are underlined. Genetic distances (roughly estimated) are given in centimorgans (cM). The gene loci are boxed. c = estimated centromere position, S = short arm, L = long arm, TPB = translocation break point.

improvement. 'Functional maps' will also prove very useful for comparative mapping and genomics (see Varshney *et al.,* 2004).

4.1. Collinearity and Synteny

RFLP probes allowed cross-species hybridization within the tribe Triticeae, and allowed comparisons among specific regions of homoeologous chromosomes (Devos *et al.,* 1992). Cross-hybridization resolved substantial conservation of the linear order of not only molecular marker loci, but also of gene loci. In these comparisons, although extensive interchromosomal translocations were detected between species, colinearity was retained within the translocated chromosome segments (Devos *et al.,* 1993a). For instance, genes determining vernalization response have not only been identified in linkage groups belonging to all the three homoeologous

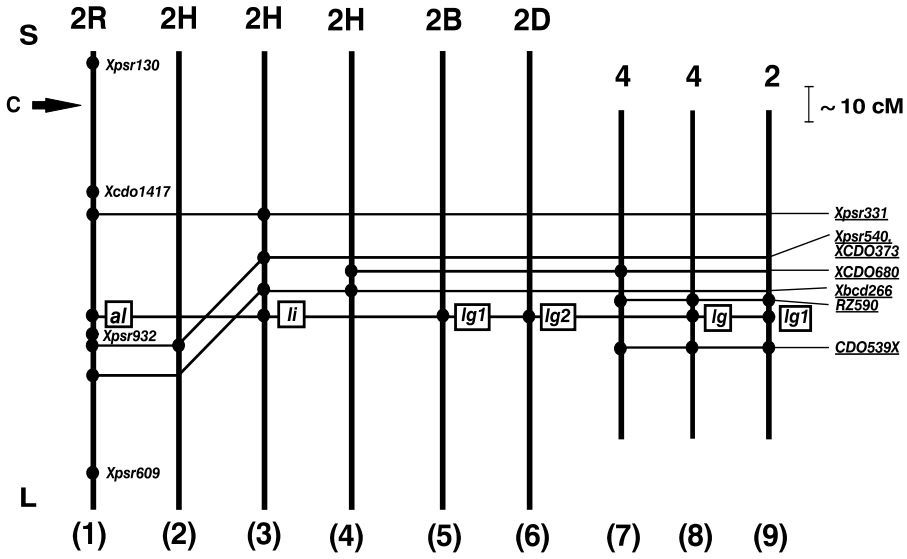


Figure 3. Comparative location of genes determining absence of ligules on chromosomes 2R of rye, 2H of barley, 2B and 2D of wheat, 4 of rice and 2 of maize as published by Börner *et al.* (1999). The following basic maps were used: (1) Korzun *et al.* (1997a), (2) Laurie *et al.* (1993), (3) Pratchett and Laurie (1994), (4) Heun *et al.* (1991), (5,6) McIntosh *et al.* (1998), (7) Causse *et al.* (1994), (8,9) Ahn and Tanksley (1993). Mapped loci are marked with a point. The connecting lines between chromosomes indicate common loci which are underlined. Genetic distances (roughly estimated) are given in centimorgans (cM). The gene loci are boxed. c = estimated centromere position, S = short arm, L = long arm.

chromosomes of group 5 of wheat (all the 3 genomes) ,but were also identified in the syntenous segments of the corresponding barley and rye chromosomes (Fig. 2). Colinearity was also described between genomes of species belonging to different tribes within the Poaceae (Gale and Devos, 1998). For instance, linkage groups 2R of rye, 2H of barley, 2B and 2D of wheat, 4 of rice and 2 of maize are synteneous, so that the genes determining absence of ligules are located in corresponding chromosome segments in all these species (Fig. 3). This suggested that the information available in maps of one cereal species could be transferred to the map of other species. For instance molecular markers mapped in wheat and barley can be integrated to genetic map of rye. Furthermore, detailed information available for the relatively small genome of rice, can be applied to larger genomes of wheat, barley and rye (Gale and Devos, 1998), although a recent study suggested that the presence of numerous translocations between wheat and rice genomes may complicate the use of rice as a model (Sorrells *et al.,* 2003). More details on synteny and comparative mapping have been described elsewhere in this book (Chapter 5 by Paterson).

4.2. Linkage Disequilibrium and Association Mapping

The above technique of molecular mapping requires a mapping population. The mapping population used for this purpose is the products of just a few cycles of recombinations, limiting the resolution of genetic maps, and often is not representative of germplasm that is being actively used in breeding programs. To oversome these problems, association mapping, based on linkage disequilibrium (LD) is being used for cereal genomics research. LD is the non-random association of markers in a populationand can provide high resolution maps of markers and genes. Association mapping based on LD may also help to resolve QTLs for specific traits (Lai *et al.,* 1994; Buckler and Thornsberry, 2002). LD depends on the evolutionary or selection history, and as a result only gene/marker with tight linkage will be detected (see Wall and Pritchard, 2003). However it is not the case in inbred species, such as wheat or barley, where large linkage blocks (often almost entire chromosome arms) have been maintained over long histories of selection Because of the narrow population structure in many crop plants due to the breeding history, association mapping has not been conducted in many plant systems (for review see Jannick and Walsh, 2002; Flint-Garcia *et al.,* 2003). In cereals so far reports on LD are available only in maize (Remington *et al.,* 2002; Ching *et al.,* 2002), wheat (Paull *et al.,* 1994; 1998) and rice (Garris *et al.*, 2003). Association mapping based on LD has also been demonstrated in maize for *Dwarf8* gene involved in flowering time (Thornsberry *et al.,* 2002) and yellow endosperm colour (Palaisa *et al.,* 2003). Efforts are underway in other cereals like barley (A. Graner, Germany, personal communication), wheat (P. Langridge, Australia, personal communication). Such high resolution mapping of traits/QTLs to the level of individual genes will provide a new possibility for studying the molecular and biochemical basis of quantitative traits variation and will help to identify specific targets for crop improvement. Though LD-based approaches hold great promise for speeding up the fine mapping, conventional linkage mapping will continue to be useful particularly when trying to mendelize QTLs and assessing the effect of a QTL in isolation (Rafalski and Morgante, 2004). In some studies, the utility of an approach involving the use of conventional linkage mapping along with LD has also been recommended for the construction of molecular maps, and for QTL analysis (Nordborg *et al.,* 2002; Zhu *et al.,* 2003). Keeping in view the importance of LD in crop plants in particular cereals, SCRI, Dundee (UK) has organized a workshop on *Gametic Phase Disequilibrium Mapping in Crop Plants* (http://wheat.pw.usda.gov/ggpages/calendar/ SCRI_2004.html) in Australia, recently.

4.3. Marker- Assisted Selection (MAS) for Crop Improvement

In a large number of studies, molecular markers have been used as tools to identify molecular markers associated with major genes and QTLs for agronomically important genes. Among cereals, in wheat alone, molecular markers have been identified for as many as ~40 traits of economic importance (see Gupta *et al.,* 1999 for a review). Similarly in barley, a large number of QTLs and genes for disease resistance, grain quality and physiological traits have been identified; these were compiled by Pat Hayes (Canada) and colleagues (http://www.barleyworld.org/NABGMP/qtlsum.htm). Details on identification of genes and QTLs for biotic and abiotic stresses in cereals are available in Chapter 8 by Jahoor *et al.* and Chapter 9 by Tuberosa and Salvi, respectively in this book.

As an example, some important studies on identification of genes and QTLs with molecular markers in rye are shown in Table 6. Availability of markers associated with these genes offers the possibility to apply marker-assisted selection (MAS) of desirable plants at the juvenile stage from an early generation. For simply inherited traits, PCR-based markers, which require each a small amount of DNA, is becoming very popular for screening large segregating populations. Unfavourable alleles can be eliminated or greatly reduced during the early stages of plant development through marker-assisted selection, focusing the selection in the field on reduced numbers of plants.

Although some examples of utilization of MAS are available in cereals like maize and barley, the promise of MAS at large scale in crop breeding still remains to be realized. The main reasons for this delay are the insufficient number of quality markers (with respect to their predictive and diagnostic value), inadequate experimental design, high costs and complexity of quantitative traits (Koebner and Summers, 2003; Chapter 10 by Koebner in this book). Only close interactions between breeders and biotechnologists will accelerate the effective implementation of MAS in cereal breeding programmes.

6. SUMMARY AND OUTLOOK

Molecular maps are now available for all cereals, and for some cereals such as rice and maize, high density maps are also available. The availability

Table 6. Utilization of molecular markers and genetic maps in identification of gene and QTLs in rye

Traits	Gene/ QTL	Marker type(s)	Location (chromo-some)	Reference
Morphological/Physiological traits				
Reduced plant height (Compactum)	*ct1*	RFLP	7R	Plaschke *et al.* (1995)
	ct2	RFLP	5R	Plaschke *et al.* (1993)
Reduced plant height	*Ddw1*	RFLP	5R	Korzun *et al.* (1996)
Spring growth habit (Vernalisation response)	*Sp1*	RFLP	5R	Plaschke *et al.* (1993)
Flowering time	*QTL*	RFLP	2R,5R, 7R	Börner *et al.* (2000)
Florets per spike	*QTL*	RFLP	6R	Börner *et al.* (2000)
Self fertility	*S*	RFLP, RAPD	1R	Senft and Wricke (1996) Voylokov *et al.* (1998)
	Z	RFLP, RAPD	2R	Senft and Wricke (1996) Voylokov *et al.* (1998)
	S5	RFLP	5R	Voylokov *et al.* (1998)
Fertility restoration	*Rfg1*	RFLP,	4R	Börner *et al.* (1998)
		RAPD,	4R	Miedaner *et al.* (2000)
		RAPD, CAPS AFLP, SCAR	4R	Stracke *et al.* (2003)
Biotic/Abiotic stress response				
Reaction to leaf rust	*Lr-a*	RFLP	6R	Ruge *et al.* (1999)
	Lr-c	RFLP, SSR	1R	Ruge *et al.* (1999)
	Lr-g	RFLP	1R	Ruge *et al.* (1999)
Reaction to powdery mildew	*Pm*	RFLP	1R	Wricke *et al.* (1996)
Resistance against cereal cyst nematode	*CreR*	RFLP, RAPD	6R	Taylor *et al.* (1998)
Aluminium tolerance	*Alt1*	RAPD, SCAR	6R	Gallego *et al.* (1998)
	Alt3	AFLP	4R	Miftahudin and Gustafson (2001)
Quality				
Secalins	*Sec2*	RFLP	2R	Malyshev *et al.* (1998)
	Sec5	RFLP	2R	Malyshev *et al.* (1998)
Waxy endosperm	*Wx*	RFLP	4R	Korzun *et al.* (1997a)

of efficient and cost effective markers will certainly be used in future for improving the available maps of other cereals also. Availability of transcript and functional maps in cereals and comparative genomics of grasses as a

whole will also facilitate transfer of markers from the major cereals to minor species including rye, sorghum, oats and millets. Similarly, physical maps of wheat and barley based on cytogenetic stocks, and thoseof rice, maize and sorghum based on BACs, will be used as a resource for future cereal breeding. Although progress in the construction of contig-based physical maps in wheat and barley is slow due to their large genomes, deletion stocks in wheat and novel approaches such as HAPPY mapping in barley (for local physical maps) are already being used for high resolution mapping in these crops. All these maps offer an opportunity both for understanding the genome organization leading to their use for crop improvement programmes. Advances in bioinformatics will also facilitate integration of information from these maps into genome sequences and gene expression profiles. In the not too distant a future, all this information should be ready on-line to address issues of plant breeding with an ultimate objective of crop improvement.

7. ACKNOWLEDGEMENTS

Authors are grateful to Professor P. K. Gupta, Ch Charan Singh University, Meerut, India and Professor P. Langridge, University of Adelaide, Australia for reviewing, useful suggestions and improving the manuscript of the article.

8. REFERENCES

Ahn S.N., Tanksley S.D. Comparative linkage maps of the rice and maize genomes. Proc Natl Acad Sci USA 1993, 90: 7980-7984

Akhunov E.D., Goodyear A.W., Geng S., Qi L.-L., Echalier B., Gill B.S., Miftahudin, Gustafson J.P., Lazo G., Chao S.M. *et al.* The organization and rate of evolution of wheat genomes are correlated with recombination rates along chromosome arms. Genome Res 2003, 13: 753-763

Araki E., Miura H., Sawada S. Identification of genetic loci affecting amylose content and agronomic traits on chromosome 4A of wheat Theor Appl Genet 1999; 98: 977-984

Barry G.F. The use of the Monsanto draft rice genome sequence in research. Plant Physiol 2001; 125: 1164-1165

Batley J., Barker G.L.A., O'Sullivan H.D., Edwards K.J., Edwards D. Mining for single nucleotide polymorphisms and insertions/deletions in maize expressed sequence tag data. Plant Physiol 2003; 132: 84 - 91

Becker J., Vos P., Kupier M., Salamini F., Heun M. Combined mapping of AFLP and RFLP markers in barley. Mol Gen Genet 1995; 249: 65-73

Bezant J., Laurie D., Pratchett N., Chojecki J., Kearsey M. Marker regression mapping of QTL controlling flowering time and plant height in a spring barley (*Hordeum vulgare* L.) cross. Heredity 1996; 77: 64-73

Bhattramakki D., Dong J.M., Chhabra A.K., Hart G.E. An integrated SSR and RFLP linkage map of *Sorghum bicolor* (L.) Moench. Genome 2000; 43: 988-1002

Biagetti M., Vitellozzi F., Ceoloni C. Physical mapping of wheat-Aegilops longissima breakpoints in mildew-resistant recombinant lines using FISH with highly repeated and low-copy DNA probes. Genome 1999; 42: 1013-1019

Blanco A., Bellomo M.P., Cenci A., Degiovanni C., Dovidio R., Iacono E., Laddomada, B., Pagnotta, M.A., Porceddu E., Sciancalepore A. *et al.* A genetic linkage map of durum wheat. Theor Appl Genet 1998; 97: 721-728

Boivin K., Deu M., Rami J.F., Trouche G., Hamon P. Towards a saturated sorghum map using RFLP and AFLP markers. Theor Appl Genet 1999; 98: 320-328

Börner A., Korzun V. A consensus linkage map of rye (*Secale cereale* L.) including 375 RFLPs, 25 isozymes and 14 gene loci. Theor Appl Genet 1998; 97: 1279-1288

Börner A., Korzun V., Malyshev S., Ivandic V., Graner A. Molecular mapping of two dwarfing genes differing in their GA response on chromosome 2H of barley. Theor Appl Genet 1999; 99: 670-675

Börner A., Korzun V., Voylokov A.V., Worland A.J., Weber W.E. Genetic mapping of quantitative trait loci in rye (*Secale cereale* L.). Euphytica 2000; 116: 203-209

Börner A., Polley A., Korzun V., Melz G. Genetics and molecular mapping of a male fertility restoration locus (*Rfg1*) in rye (*Secale cereale* L.). Theor Appl Genet 1998; 97: 99-102

Boyko E.V., Gill K.S., Mickelson-Young L., Nasuda S., Raupp W.J., Yiegle J.N., Singh S., Hassawi D.S., Frity A.K., Namuth D. *et al.* A high-density genetic linkage map of *Aegilops tauschii*, the D-genome progenitor of bread wheat. Theor Appl Genet 1999; 99: 16-26

Boyko E., Kalendar R., Korzun V., Fellers J., Korol A., Schulman A.H., Gill B.S. A high-density cytogenetic map of the *Aegilops tauschii* genome incorporating retrotransposons and defense-related genes: insights into cereal chromosome structure and function. Plant Mol Biol 2002; 48: 767-790

Brondani C., Pereira R., Brondani V., Rangel P.H.N., Ferreira M.E. Development and mapping of *Oryza glumaepatula*-derived microsatellite markers in the interspecific cross *Oryza glumaepatula* × *O. sativa*. Hereditas 2001; 134: 59-71

Buckler E.S., Thornsberry J. Plant moleculer diversity and applications to genomics. Curr Opin Plant Biol 2002; 5: 107-111

Burr B., Burr F.A. Recombinant inbreds for molecular mapping in maize. Trends Genet 1991; 7:55-60

Cabrera A., Martin A., Barro F. In-situ comparative mapping (ISCM) of Glu-1 loci in Triticum and Hordeum. Chrom Res 2002; 10: 49-54

Cadalen T., Boeuf C., Bernard S., Bernard M. An intervarietal molecular marker map in *Triticum aestivum* L. M. Thell. and comparison with a map from wide cross. Theor Appl Genet 1997; 94: 367-377

Casa A.M., Brouwer C., Nagel A., Wang L.J., Zhang Q., Kresovich S., Wessler S.R. The MITE family Heartbreaker (Hbr): molecular markers in maize. Proc Natl Acad Sci USA 2000; 97: 10083-10089

Castiglioni P., Ajmone-Marsan P., van Wijk R., Motto M. AFLP markers in a molecular linkage map of maize: codominant scoring and linkage group distribution. Theor Appl Genet 1999; 99: 425-431

Castiglioni P., Pozzi C., Heun M., Terzi V., Muller K.J., Rohde W., Salamini F. An AFLP-based procedure for the efficient mapping of mutations and DNA probes in barley. Genetics 1998; 149: 2039-2056

Causse M.A., Fulton T.M., Cho Y.G., Ahn S.N., Chungwongse J., Wu K., Xiao J., Yu Z., Ronald P.C., Harrington S.E. *et al.* Saturated molecular map of the rice genome based on an interspecific backcross population. Genetics 1994; 138: 1251-1274

Causse M., Santoni S., Damerval C., Maurice A., Charcosset A., Deatrick J., deVienne D. A composite map of expressed sequences in maize. Genome 1996; 39: 418-432

Chalmers K.J., Campbell A.W., Kretschmer J., Karakousis A., Henschke P.H., Pierens S., Harker N., Pallota M., Cornish G.B., Shariflou M.R. *et al.* Construction of three linkage maps in bread wheat, *Triticum aestivum*. Aust Jour Agric Res 2001; 52: 1089-1119

Chang R.Y., O'Donoughue L.S., Bureau T.E. Inter-MITE polymorphism (IMP): a high throughput tansposon-based genome mapping and fingerprinting approach. Theor Appl Genet 2001; 102: 773-781

Chao S., Baysdorfer C., Herediadiaz O., Musket T., Xu G., Coe E.H. RFLP mapping of partially sequenced leaf cDNA clones in maize. Theor Appl Genet 1994; 88: 717-721

Chen J.M, Gustafson J.P. Physical mapping of restriction-fragment-length-polymorphisms (RFLPs) in homoeologous group-7 chromosomes of wheat by in-situ hybridisation. Heredity 1995; 75: 225-233

Chen M., Presting G., Barbazuk W.B., Goicoechea J.L., Blackmon B., Fang G., Kim H., Frisch D., Yu Y., Sun S., *et al.* An integrated physical and genetic map of the rice genome. Plant Cell 2002; 14: 537- 545

Chen X., Temnykh S., Xu Y., Cho Y.G., McCouch S.R. Development of a microsatellite framework map providing genome-wide coverage in rice (*Oryza sativa* L.). Theor Appl Genet 1997; 95: 553-567

Cheng D.W., Armstrong K.C., Tinker N., Wight C.P., He S., Lybaert A., Fedak G., Molnar S.J. Genetic and physical mapping of Lrk10-like receptor kinase sequences in hexaploid oat (*Avena sativa* L.). Genome 2002; 45: 100-109

Cheng Z.K., Buell C.R., Wing R.A., Gu M. Jiang J. Toward a cytological characterization of the rice genome. Genome Res 2001a; 11: 2133-2141

Cheng Z.K., Buell C.R., Wing R.A., Jiang J. Resolution of fluorescence in-situ hybridization mapping on rice mitotic prometaphase chromosomes, meiotic pachytene chromosomes and extended DNA fibers. Chromosome Res 2002; 10: 379-387

Cheng Z.K., Presting G.G., Buell C.R., Wing R.A., Jiang J.M. High-resolution pachytene chromosome mapping of bacterial artificial chromosomes anchored by genetic markers reveals the centromere location and the distribution of genetic recombination along chromosome 10 of rice. Genetics 2001b; 157: 1749-1757

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

Chittenden L.M., Schertz K.F., Lin Y.R., Wing R.A., Paterson A.H.. A detailed RFLP map of *Sorghum bicolor* × *S. Propinquum*, suitable for high-density mapping, suggests ancestral duplication of *Sorghum* chromosomes as chromosomal segments. Theor Appl Genet 1994; 87: 925-933

Cho Y.G., McCouch S.R., Kuiper M., Kang M.R., Pot J., Groenen J.T.M., Eun M.Y. Integrated map of AFLP, SSLP and RFLP markers using a recombinant inbred population of rice (*Oryza sativa* L.) Theor Appl Genet 1998; 97: 370-380

Coe E., Cone K., McMullen M., Chen S., Davis G., Gardiner J., Liscum E., Polacco M., Paterson A., Sanchez-Villeda H., Soderlund C., Wing R. Access to the maize genome: an integrated physical and genetic map. Plant Physiol 2002; 128: 9-12

Coe E.H., Polacco M., Davis G., McMullen M. Maize molecular maps: markers, bins, and database. In: R.L. Phillips, I.K. Vasil (eds.) *DNA-Based Markers in Plants, 2nd Edition*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2001, pp 255-284

Cone K., McMullen M., Bi I.V., Davis G., Yim Y., Gardiner J., Polacco M., Sanchez-Villeda H., Fang Z., Schroeder S. *et al.* Genetic, physical, and informatics resources for maize, on the road to an integrated map. Plant Physiol 2002; 130: 1598-1605

Costa J.M., Corey A., Hayes P.M., Jobet C., Kleinhofs A., Kopisch-Obusch A., Kramer S.F., Kudrna D., Li M., Riera-Lizarazu O., *et al.* Molecular mapping of the Oregon Wolfe

Barleys: a phenotypically polymorphic doubled-haploid population. Theor Appl Genet 2001; 103: 415-424

Cox D.R., Burmeister M., Price E.R., Kim S., Mayers R.M. Radiation hybrid mapping- a somatic-cell genetic method for constructing high-resolution maps of mammalian chromosomes. Science 1990; 250: 245-250

Davis G.L., McMullen M.D., Baysdorfer C., Musket T., Grant D. Staebell M., Xu G., Polacco M., Koster L., Melia-Hancock S., *et al.* A maize map standard with sequenced core markers, grass genome reference points and 932 expressed sequence tagged sites (ESTs) in a 1736-locus map. Genetics 1999; 152: 1137- 1172

Dear P.H., Cook R.R. HAPPY mapping- a proposal for linkage mapping the human genome. Nucl Acids Res 1989; 17: 6795-6807

Delaney D.E., Nasuda S., Endo T.R., Gill B.S., Hulbert S.H.. Cytologically based physical maps of the group-2 chromosomes of wheat. Theor Appl Genet 1995a; 91: 568-573

Delaney D.E., Nasuda S., Endo T.R., Gill B.S., Hulbert S.H.. Cytologically based physical maps of the group-3 chromosomes of wheat. Theor Appl Genet 1995b; 91: 780-782

Devos K.M., Atkinson M.D., Chinoy C.N., Francis H.A., Harcourt R.L., Koebner R.M.D., Liu C.J., Masojć P., Xie D.X., Gale M.D. Chromosomal rearrangement in the rye genome relative to that of wheat. Theor Appl Genet 1993a; 85: 673-680

Devos K.M., Atkinson M.D., Chinoy C.N., Liu C.J., Gale M.D. RFLP-based genetic map of the homoeologous group 3 chromosomes of wheat and rye. Theor Appl Genet 1992; 83: 931-939.

Devos K.M., Gale M.D. The use of random amplified polymorphic DNA markers in wheat. Theor Appl Genet 1992; 84: 567-572

Devos K.M., Gale M.D. Extended genetic maps of the homoeologous group-3 chromosomes of wheat, rye and barley. Theor Appl Genet 1993; 85: 649-652

Devos KM, Millan T, Gale MD. Comparative RFLP maps of the homoeologous group-2 chromosomes of wheat, rye and barley. Theor Appl Genet 1993b; 85: 784-792

Dong F.G., Miller J.T., Jackson S.A., Wang G.L., Ronald P.C., Jiang J.M. Rice (*Oryza sativa*) centromeric regions consist of complex DNA. Proc Natl Acad Sci USA 1998; 95: 8135-8140

Draye X., Lin Y.-R., Qian X.-Y., Bowers J.E., Burrow G.B., Morrell P.L., Peterson D.G., Presting G.G., Ren S-X., Wing R.A. *et al.* Toward integration of comparative genetic physical, diversity, and cytomolecular maps for grasses and grains, using the sorghum genome as a foundation. Plant Physiol 2001; 125:1325-1341

Druka A., Kudrna D., Kannangara C.G., von Wettstein D., Kleinhofs A. Physical and genetic mapping of barley (*Hordeum vulgare*) germin-like cDNAs. Proc Natl Acad Sci USA 2002; 99: 850-855

Dubcovsky J., Luo M.C., Zhong G.Y., Bransteitter R., Desai A., Kilian A., Kleinhofs A., Dvorak J. Genetic map of diploid wheat, *Triticum monococcum* L., and its comparison with maps of *Hordeum vulgare* L. Genetics 1996; 143: 983-999

Edwards K.J., Thompson H., Edwards D., deSaizieu A., Sparks C., Thompson J.C., Greenland A.J., Eyers M., Schuh W. Construction and characterization of a yeast artificial chromosome library containing three haploid maize genome equivalents. Plant Mol Biol 1992; 19: 299-308

Endo T.R., Gill B.S. The deletion stocks of common wheat. Jour Heredity 1996; 87: 295-307

Faris J.D., Haen K.M., Gill B.S. Saturation mapping of a gene-rich recombination hot spot region in wheat. Genetics 2000; 154: 823-835

Flint-Garcia S.A., Thornsberry J.M., Buckler E.S. IV. Structure of linkage disequilbrium in plants. Annu Rev Plant Biol 2003; 54: 357-374

Francki M.G., Berzonsky W.A., Ohm H.W., Anderson J.M. Physical location of a HSP70 gene homologue on the centromere of chromosome 1B of wheat (*Triticum aestivum* L.). Theor Appl Genet 2002; 104: 184-191

Fu H.H., Dooner H.K. A gene-enriched BAC library for cloning large allele-specific fragments from maize: isolation of a 240-kb contig of the bronze region. Genome Res 2000; 10: 866-873

Gale M.D, Atkinson M.D, Chinoy C.N, Harcourt R., Jia J., Li Q.Y., Devos K.M. Genetic maps of hexaploid wheat. In: S. Chen (ed.), *Proc 8th Intern Wheat Genet Symp*, China Agricultural Scientech Press, Beijing, China, 1995, pp 29-40

Gale M.D., Devos K.M. Comparative genetics in the grasses. Proc Natl Acad Sci USA 1998; 95: 1971-1974

Galiba G., Quarrie S.A., Sutka J., Morgunov A., Snape J.W. RFLP mapping of the vernalisation (*Vrn1*) and frost resistance (*Fr1*) genes on chromosome 5A of wheat. Theor Appl Genet 1995; 90: 1174-1179

Gallego F.J., Calles B., Benito C. Molecular markers linked to the aluminum tolerance gene *Alt1* in rye (*Secale cereale* L.). Theor Appl Genet 1998; 97: 1104-1109

Gandon B., Chiquet V., Guyomarc'h H., Baron C., Sourdille P., Specel S., Foisset N., Murigneu× A., Dufour P., Bernard M. Development of microsatellite markers for wheat genetic mapping improvement. In: *Plant, Animal & Microbe Genomes X Conf*, San Diego, CA, USA, Jan 12-16, 2002 (http://www.intl-pag.org/pag/10/abstracts/PAGX_P187.html)

Gao L.F., Tang J.F., Li H.W., Jia J.Z. Analysis of microsatellites in major crops assessed by computational and experimental approaches. Mol Breed 2003; 12: 245-261

Gardiner J.M., Coe E.H., Melia- Hancock S., Hoisington D.A., Chao S. Development of a core RFLP map using an immortalized F2 population of maize. Genetics 1993; 134: 1001-1012

Garris A.J., McCouch S.R., Kresovich S. Population structure and its effect on haplotypes diversity and linkage disequilibrium surrounding the *xa5* locus of rice (*Oryza sativa* L.). Genetics 2003; 165: 759-769

Giese H., Holm-Jensen A.G., Mathiassen H., Kjær B., Rasmussen S.K., Bay H., Jensen J. Distribution of RAPD markers on a linkage map of barley. Hereditas 1994; 120: 267-273

Gill K.S., Gill B.S., Boyko E.V. Identification and high density mapping of gene- rich regions in chromosome group 5 of wheat. Genetics 1996b; 143: 1001-1012

Gill K.S., Gill B.S., Endo T.R. A chromosome region-specific mapping strategy reveals gene-rich telomeric ends in wheat. Chromosoma 1993; 102: 374-381

Gill K.S., Gill B.S., Endo T.R., Taylor T. Identification and high-density mapping of gene-rich regions in chromosome group 1 of wheat. Genetics 1996a; 144: 1883-1891

Gill K.S., Lubbers E.L., Gill B.S., Raupp W.J., Cox T.S. A genetic linkage map of *Triticum tauschii* (DD) and its relationship to the D genome of bread wheat (AABBDD). Genome 1991; 34: 362-374

Gilpin B.J., McCallum J.A., Frew G.M., Timmerman-Vaughan G.M. A linkage map of the pea (*Pisum sativum* L.) genome containing cloned sequences of known functions and expressed sequence tags (ESTs). Theor Appl Genet 1997; 95: 1289-1299

Graner A., Jahoor A., Schondelmaier J., Siedler H., Pillen K., Fischbeck G., Wenzel G., Herrmann R.G. Construction of an RFLP map of barley. Theor Appl Genet 1991; 83: 250-256

Gupta P.K., Balyan H.S., Edwards K.J., Isaac P., Korzun V., Roder M., Gautier M.F., Joudrier P., Schlatter A.R., Dubcovsky J., *et al.* Genetic mapping of 66 new microsatellite (SSR) loci in bread wheat Theor Appl Genet 2002a; 105: 413-422

Gupta P.K., Rustgi S., Sharma S., Singh R., Kumar N., Balyan H.S. Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. Mol Gen Genom 2003; 270: 315-323

Gupta P.K., Varshney, R.K. The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica 2000; 113: 163-185

Gupta P.K., Varshney R.K., Prasad M. Molecular markers: principles and methodology. In: S.M. Jain, D.S. Brar, B.S. Ahloowalia (eds.), *Molecular Techniques in Crop Improvement*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2002b; pp 9- 54

Gupta P.K., Varshney R.K., Sharma P.C., Ramesh B. Molecular markers and their applications in wheat breeding. Plant Breed 1999; 118: 369-390

Hackauf B., Wehling P. Identification of microsatellite polymorphisms in an expressed portion of the rye genome. Plant Breed 2002a; 121: 17-25

Harker N., Rampling L.R., Shariflou M.R., Hayden M.J., Holton T.A., Morell M.K., Sharp P.J., Henry R.J., Edwards K.J. Microsatellites as markers for Australian wheat improvement. Aust Jour Agric Res 2001; 52: 1121-1130

Harushima Y., Yano M., Shomura A., Sato M., Shimono T., Kuboki Y., Yamamoto T., Lin S.Y., Antonio B.A., Parco A. *et al.* A high density rice genetic linkage map with 2275 markers using a single F2 population. Genetics 1998; 148: 479-494

Hazen S.P., Leroy P., Ward R.W. AFLP in *Triticum aestivum* L.: patterns of genetic diversity and genome distribution. Euphytica 2002; 125: 89-102

Hernandez P., Dorado G., Prieto P., Gimenez M.J., Ramirez M.C., Laurie D.A., Snape J.W., Martin A. A core genetic map of *Hordeum chilense* and comparisons with maps of barley (*Hordeum vulgare*) and wheat (*Triticum aestivum*). Theor Appl Genet 2001; 102: 1259-1264

Heun M., Kennedy A.E., Anderson J.A., Lapitan N.L.V., Sorrells, M.E., Tanksley S.D. Construction of a restriction fragment length polymorphism map for barley (*Hordeum vulgare*). Genome 1991; 34: 437-447

Hohmann U., Graner A., Endo T.R., Gill B.S., Herrmann R.G. Comparison of wheat physical maps with barley linkage maps for group 7 chromosomes. Theor Appl Genet 1995; 91: 618-626

Holton T.A., Christopher J.T., McClure L., Harker N., Henry R.J. Identification and mapping of polymorphic SSR markers from e×pressed gene sequences of barley and wheat. Mol Breed 2002; 9: 63-71

Hong G.F., Qian Y.M., Yu S.L., Hu X., Zhu J., Tao W.H., Li W., Su C., Zhao H.Y., Qiu L.F., *et al*. A 120 kilobase resolution contig map of the rice genome. DNA Seq 1997; 7: 319-335

Hossain G.K., Riera-Lizararu O., Vales I.M. , Maan S.S., Kalavacharla V., Kianian S.F. Molecular characterization and localization of an *scs* gene in wheat using aneuploids and radiation hybrid mapping. In: *Plant, Animal & Microbe Genomes X Conf*, San Diego, CA, USA, January 12-16, 2002, www.intl-pag.org/10/abstracts/PAGX_ P347.html

Islam-Faridi M.N., Childs K.L., Klein P.E., Hodnett G., Menz M.A., Klein R.R., Rooney W.L., Mullet J.E., Stelly D.M., Price H.J. A molecular cytogenetic map of sorghum chromosome 1: Fluorescence in situ hybridization analysis with mapped bacterial artificial chromosomes. Genetics 2002; 161: 345-353

Jannink J.-L., Walsh B. Association mapping in plant populations. In: M.S. Kang (ed.), *Quantitative Genetics, Genomics and Plant Breeding*, CAB International, Wallingford, USA, 2002; pp 59- 68

Jia J., Devos K.M., Chao S., Miller T.E., Reader S.M., Gale M.D. RFLP-based maps of the homoeologous group-6 chromosomes of wheat and their application in the tagging of *Pm12,* a powdery mildew resistance gene transferred from *Aegilops speltoides* to wheat Theor Appl Genet 1996; 92: 559-565

Jiang J.M., Gill B.S. Nonisotopic in-situ hybridization and plant genome mapping - the first 10 years. Genome 1994; 37: 717-725

Jiang J.M., Gill B.S., Wang G.L., Ronald P.C., Ward D.C. Metaphase and interphase fluorescence in-situ hybridization mapping of the rice genome with bacterial artificial chromosomes. Proc Natl Acad Sci USA 1995; 92: 4487-4491

Jin H., Domier L.L., Shen X.J., Kolb F.L. Combined AFLP and RFLP mapping in two hexaploid oat recombinant inbred populations. Genome 2000; 43: 94-101

Kamolsukyunyong W., Ruanjaichon V., Siangliw M., Kawasaki S., Sasaki T., Vanavichit A., Tragoonrung S. Mapping of quantitative trait locus related to submergence tolerance in rice with aid of chromosome walking. DNA Res 2001; 8: 163-171

Kantety R.V., Rota M.L., Matthews D.E., Sorrells M.E. Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. Plant Mol Biol 2002; 48: 501-510

Kasha K.J., Kleinhofs A., the North American Barley Genome Mapping Project. Mapping of the barley cross Harrington × TR306. Barley Genet Newsl 1994; 23:65-69

Kennard W.C., Phillips R.L., Porter R.A., Grombacher A.W. A comparative map of wild rice (*Zizania palustris* L. 2n=2×=30). Theor Appl Genet 2000; 101: 677-684

Khlestkina E.K., Than M.H.M., Pestsova E.G., Röder M.S., Malyshev S.V., Korzun V., Börner A. Mapping of 99 microsatellite loci in rye (*Secale cereale* L.) including 39 expressed sequence tags. Theor Appl Genet 2004; communicated

Kikuchi S., Satoh K., Nagata T., Kawagashira N., Doi K., Kishimoto N., Yazaki J., Ishikawa M., Yamada H., Ooka H., *et al.* Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice. Science 2003; 301:376-379

Kim J.S., Childs K.L., Islam-Faridi M.N., Menz M.A., Klein R.R., Klein P.E., Price H.J., Mullet J.E., Stelly D.M. Integrated karyotyping of sorghum by *in situ* hybridization of landed BACs. Genome 2002; 45: 402-412

Klein P.E., Klein R.R., Cartinhour S.W., Ulanch P.E., Dong J., Obert J.A., Morishge D.T., Schlueter S.D., Childs K.L., Ale M., *et al.* A high throughput AFLP based method for constructing integrated genetic and physical maps: progress toward a sorghum genome map. Genome Res 2000; 10: 789-807

Kleine M., Michalek W., Diefenthal H., Dargatz H., Jung C. Construction of a barley (*Hordeum vulagre* L.) YAC library and isolation of a Hor1-specific clone. Mol Gen Genet 1993; 240: 265-272

Kleinhofs A., Graner A. An integrated map of the barley genome. In: R.L. Phillips, I.K. Vasil (eds.) *DNA-Based Markers in Plants*, 2nd Edition, Kluwer Academic Publishers, Dordrecht, The Netherlands 2001; pp. 187-200

Kleinhofs A., Kilian A., Saghai Maroof M.A., Biyashev R.M., Hayes P.M., Chen F.Q., Lapitan N., Fenwick A., Blake T., Kanasin V., *et al.* A molecular, isozyme and morphological map of the barley (*Hordeum vulgare*) genome. Theor Appl Genet 1993; 86: 705-712

Koebner R., Summers R. 21st century wheat breeding: selection in plots or detection in plates? Trends Biotech 2003; 21: 59-63

Kojima T., Nagaoka T., Noda K., Ogihara Y. Genetic linkage map of ISSR and RAPD markers in einkorn wheat in relation to that of RFLP markers. Theor Appl Genet 1998; 96: 37-45

Korzun V., Malyshev S., Kartel N., Westermann T., Weber W.E., Börner A. A genetic linkage map of rye (*Secale cereale* L.). Theor Appl Genet 1998; 96: 203-208

Korzun V., Malyshev S., Voylokov A., Börner A. RFLP based mapping of three mutant loci in rye (*Secale cereale* L.) and their relation to homoeologous loci within the *Gramineae*. Theor Appl Genet 1997a; 95: 468-473

Korzun V., Malyshev S., Voylokov A.V., Börner A. A genetic map of rye (*Secale cereale* L.) combining RFLP, isozyme, microsatellite and gene loci. Theor Appl Genet 2001; 102: 709-717

Korzun V., Melz G., Börner A. RFLP mapping of the dwarfing (*Ddw1*) and hairy peduncle (*Hp*) genes on chromosome 5 of rye (*Secale cereale* L.). Theor Appl Genet 1996; 92: 1073-1077

Korzun V., Röder M.S., Wendehake K., Pasqualone A., Lotti C., Ganal M.W., Blanco A. Integration of dinucleotide microsatellites from hexaploid bread wheat into a genetic linkage map of durum wheat . Theor Appl Genet 1999; 98: 1202-1207

Korzun V., Röder M., Worland A.J., Börner A. Mapping of the dwarfing (*Rht12*) and vernalisation response (*Vrn1*) genes in wheat by using RFLP and microsatellite markers, Plant Breed 1997b; 116: 227-232

Kota R.S., Gill K.S., Gill B.S., Endo T.R. A cytogenetically based physical map of chromosome-1B in common wheat. Genome 1993; 36: 548-554

Kota R., Varshney R.K., Thiel T., Dehmer K.-J., Graner A. Generation and comparison of EST-derived SSR and SNP markers in barley (*Hordeum vulgare* L.). Hereditas 2001; 135:145-151

Kremer C.A., Lee M., Holland J.B. Restriction fragment length polymorphism based linkage map of a diploid *Avena* recombinant inbred line population. Genome 2001; 44: 192-204

Künzel G., Korzun L., Meister A. Cytologically integrated physical restriction fragment length polymorphism maps for the barley genome based on translocation breakpoints. Genetics 2000; 154: 397-412

Künzel G., Waugh R. Integration of microsatellite markers into the translocation-based physical RFLP map of barley chromosome 3H. Theor Appl Genet 2002; 105: 660-665

Kurata N., Umehara Y., Tanoue H., Sasaki T. Physical mapping of the rice genome with YAC clones. Plant Mol Biol 1997; 35: 101-113

Kurata, N., Nagamura, Y., Yamamoto, K., Harushima, Y., Sue, N., Wu, J., Antonio, B.A., Shomura, A., Shimizu, T., Lin, *et al.* A 300 kolobase interval genetic map of rice including 883 expressed sequences. Nature Genet 1994; 8: 365-372

Lagudah E.S., Dubcovsky J., Powell W. Wheat genomics. Plant Phsyiol Biochem 2001; 39: 335-344

Lai C., Lyman R.F., Long A.D., Langley C.H., Mackay F.C. Naturally occuring variation in bristle number and DNA polymorphisms at the scabrous locus of *Drosophila melanogaster*. Science 1994; 226: 1697-1702

Lander E.S., Green P., Abrahamson J., Barlow A., Daly M.J., Lincoln S.E., Newburg L. Mapmaker: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1987; 1: 174-181

Langridge P., Karakousis A., Collins N., Kretschmer J., Manning S. A consensus linkage map of barley. Mol Breed 1995; 1: 389-395

Lapitan N.L.V., Brown S.E., Kennard W., Stephens J.L., Knudson D.L. FISH physical mapping with barley BAC clones. Plant Jour 1997; 11: 149-156

Laurie D.A., Pratchett N., Bezant J.H., Snape J.W. RFLP mapping of five major genes and eight quantitative trait loci controlling flowering time in a winter × spring barley (*Hordeum vulgare* L.) cross. Genome 1995; 38: 575-585

Laurie D.A., Pratchett N., Devos K.M., Leitch I.J., Gale M.D. The distribution of RFLP markers on chromosome 2(2H) of barley in relation to the physical and genetical location of 5S rDNA. Theor Appl Genet 1993; 87: 177-183

Lee M., Sharopova N., Beavis W.D., Grant D., Katt M., Blair D., Hallauer A. E×panding the genetic map of maize with the intermated B73 × Mo17 (IBM) population. Plant Mol Biol 2002; 48: 453-461

Li J.Z., Sjakste T.G., Röder M.S., Ganal M.W. Development and genetic mapping of 127 new microsatellite markers in barley. Theor Appl Genet 2003; 107: 1021-1027

Liu Y.G., Nagaki K., Fujita M., Kawaura K., Uozumi M., Ogihara Y. Development of an efficient maintenance and screening system for large-insert genomic DNA libraries of he×aploid wheat in a transformation-competent artificial chromosome (TAC) vector. Plant Jour 2000; 23: 687-695

Liu Y.G., Tsunewaki K. Restriction fragment length polymorphism (RFLP) analysis in wheat. II. Linkage maps of the RFLP sites in common wheat. Jpn Jour Genet 1991; 66: 617-633

Loarce Y., Hueros G., Ferrer E. A molecular linkage map of rye. Theor Appl Genet 1996; 93: 1112-1118

Lorieux M., Ndjiondjop M.N., Ghesquiere A. A first interspecific *Oryza sativa* × *Oryza glaberrima* microsatellite-based genetic linkage map. Theor Appl Genet 2000; 100: 593-601

Lotti C., Salvi S., Pasqualone A., Tuberosa R., Blanco A. Integration of AFLP markers into an RFLP-based map of durum wheat. Plant Breed 2000; 119: 393-401

Ma X.F., Ross K., Gustafson J.P. Physical mapping of restriction fragment length polymorphism (RFLP) markers in homoeologous groups 1 and 3 chromosomes of wheat by in situ hybridization. Genome 2001; 44: 401-412

Ma X.F., Wanous M.K., Houchins K., Rodriquez Mila M.A., Goicoechea P.G., Wang Z., Xie M., Gustafson J.P. Molecular linkage mapping in rye (*Secale cereale* L.). Theor Appl Genet 2001; 102: 517-523

Maheswaran, M., Subudhiu, P.K., Nandi, S.S., Xu, J.C., Parco, A., Yang, D.C., and Huang N. Polymorphism, segregation and distribution of AFLP markers in a doubled haploid rice population. Theor Appl Genet 1997; 94: 39-45

Maluszynska J. *In situ* hybridization in plants- methods and application. In: S.M. Jain, D.S. Brar, B.S. Ahloowalia (eds.), *Molecular Techniques in Crop Improvement*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2002; pp 299-326

Malyshev S., Khmyl T.O., Zabenkova K.I., Voylokov A.V., Korzun V.N., Kartel N.A. RFLP-based mapping of the *Sec-2* and *Sec-5* loci encoding 75K *y*-secalins of rye. Plant Breed 1998; 117: 329-333

Manly K., Olson J. Overview of QTL mapping software and introduction to Map Manager QT. Mammal Genome 1999; 10: 327-334

Mano Y., Kawasaki S., Takaiwa F., Komatsuda T. Construction of a genetic map of barley (*Hordeum vulagre* L.) cross Azumamungi × Kanto Nakate Gold using a simple and efficient amplified fragment length polymorphism system. Genome 2001; 44: 284-292

Mao L., Wood T.C., Yu Y., Budiman M.A., Tomkins J., Woo S., Sasinowski M., Presting G., Frisch D., Goff S. *et al.* Rice transposable elements: A survey of 73,000 sequence-tagged-connectors. Genome Res 2000; 7: 1072-1084

Marino C. L., Nelson J.C., Lu Y.H., Sorrells M.E., Leroy P., Tuleen N.A., Lopes C.R., G.E., G.E. Hart Molecular genetic maps of the group 6 chromosomes of hexaploid wheat *Triticum aestivum* L. em. Thell.). Genome 1996; 39, 359-366

Marsan P.A., Gorni C., Chitto A., Redaelli R., van Vijk R., Stam P., Motto M. Identification of QTLs for grain yield and grain-related traits of maize (*Zea mays* L.) using an AFLP map, different testers, and cofactor analysis. Theor Appl Genet 2001; 102: 230-243

Masojć P., Myśków B., Milczarski P. Extending a RFLP-based genetic map of rye using random amplified polymorphic DNA (RAPD) and isozyme markers. Theor Appl Genet 2001; 102:1273-1279

McCouch S.R. Rice molecular map. In: R.L. Phillips, I.K. Vasil (eds.) *DNA-Based Markers in Plants*, 2nd Edition, Kluwer Academic Publishers, Dordrecht, The Netherlands 2001; pp 337-345

McCouch S.R., Chen X., Panaud O., Temnykh S., Xu Y., Cho Y.G., Huang N., Ishii T. Blair M. Microsatellite marker development, mapping and applications in rice genetics and breeding. Plant Mol Biol 1997; 35: 89-99

McCouch S.R.,Teytelman L., Xu Y., Lobos K.B., Clare K., Walton K., Fu B., Maghirang R., Li Z., Xing Y., *et al.* Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.). DNA Res 2002; 9: 199-207

McIntosh R.A., Hart G.E., Devos K.M., Gale M.D., Rogers W.J. Catalogue of gene symbols for wheat. In: Slinkard A.E. (ed.), *Proc 9th Int Wheat Genet Symp*, Vol 5, University Extension Press, University of Saskatchewan, Canada, 1998; pp 1-236

Menz M.A., Klein R.R., Mullet J.E., Obert J.A., Unruh N.C. Klein P.E. A high- density map of *Sorghum bicolour* (L.) Moench based on 2926 AFLP, RFLP and SSR markers. Plant Mol Biol 2002; 48: 483-499

Messmer M.M., Keller M., Zanetti S., Keller B. Genetic linkage map of wheat × spelt cross. Theor Appl Genet 1999; 98: 1163-1170

Mickelson-Young L., Endo T.R., Gill B.S. A cytogenetic ladder-map of the wheat homoeologous group-4 chromosomes. Theor Appl Genet 1995; 90: 1007-1011

Miedaner T., Glass C., Dreyer F., Wilde P., Wortmann Geiger H.H. Mapping of genes for male fertility restoration in "Pampa" CMS winter rye (*Secale cereale* L.) Theor Appl Genet 2000; 101: 1226-1233

Miftahudin, Gustafson, J.P. Molecular markers tightly linked to the Aluminum tolerance gene *Alt3* in rye (*Secale cereale* L.). In: *Proc EUCARPIA Rye Meeting,* Radzikow, Poland 2001; pp 341-343

Milla M.A.R., Gustafson J.P. Genetic and physical characterization of chromosome 4DL in wheat. Genome 2001; 44: 883-892

Miyazaki C., Osanai E., Saeki K., Hirota N., Ito K., Ukai Y., Konishi T., Saito A. Construction of a barley RFLP linkage map using an F_2 population derived from a cross between Ko A and Mokusekko 3. Barley Genet Newsl 2000; 30: 41-43

Morton N.E. Sequential tests for the detection of linkage Amer Jour Hum Genet 7: 277-318. 1955

Nachit M., Elouafi I., Pagnotta M.A., El Saleh A., Iacono E., Labhili M., Asbati A., Azrak M., Hazzam H., Benscher D. *et al.,* Molecular linkage map for an intraspecific recombinant inbred population of durum wheat (*Triticum turgidum* L. var. *durum*). Theor Appl Genet 2001; 102: 177-186

Nasu S., Suzuki J., Ohta R., Hasegawa K., Yui R., Kitazawa N., Monna L., Minobe Y. Search for and analysis of single nucleotide polymorphisms (SNPs) in rice (*Oryza sativa*, *Oryza rufipogon*) and establishment of SNP markers. DNA Res 2002; 9: 163-171

Nelson J.C., Sorrells M.E., Van Deynze A.E., Lu Y.H., Atkinson M., Bernard M., Leroy P., Faris J.D., Anderson J.A. Molecular mapping of wheat: Major genes and rearrangements in homoeologous groups 4, 5, and 7. Genetics 1995a; 141: 721-731

Nelson J.C., Van Deynze A.E., Autrique E., Sorrells M.E., Lu Y. H., Merlino M., Atinkson M., Leroy P. Molecular mapping of wheat. Homoeologous group 2. Genome 1995b; 38: 516-524

Nelson J.C., Van Deynze A.E., Autrique E., Sorrells M.E., Lu Y.H., Negre M, Atinkson M., Leroy P. Molecular mapping of wheat. Homoeologous group 3. Genome 1995c; 38: 525-533

Nordborg M., Borevitz J.O., Bergelson J., Berry C.C., Chory J., Hagenblad J., Kreitman M., Maloof J.N., Noyes T., Oefner P.J., *et al.* The extent of linkage disequilibrium in *Arabidopsis thaliana.* Nature Genet 2002; 30: 190-193

O' Donoughue L.S., Kianian S.F., Raypati P.J., Penner G.A., Sorrells M.E., Tanksley S.D., Phillips R.L., Rines H.W., Lee M., Fedak G., *et al.* A molecular map of cultivated oat. Genome 1995; 38: 368-380

O' Donoughue L.S., Wang Z., Roder M., Kneen B., Leggett M., Sorrells M.E., Tanksley S.D. An RFLP- based linkage map of oats based on cross between two diploid taxa (*Avena atlantica* × *A. hirtula*). Genome 1992; 35: 765-771

Ogihara Y., Hasegawa K., Tsujimoto H. 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.* Mol Gen Genet 1994; 244: 253-259

Paillard S., Schnurbusch T., Winzeler M., Messmer M., Sourdille P., Abderhalden O., Keller B. Schachermayr G. An integrative genetic linkage map of winter wheat (*Triticum aestivum* L.). Theor Appl Genet 2003; 107: 1235-1242

Palaisa K.A., Morgante M., Williams M., Rafalski A. Contrasting effects of selection on sequence diversity and linkage disequilibrium at two phytoene synthase loci. Plant Cell 2003; 15: 1795-1806

Paull J.G., Chalmers K.J., Karakousis A., Kretschmer J.M., Manning S., Langridge P. Genetic diversity in Australian wheat varieties and breeding material based on RFLP data. Theor Appl Genet 1998; 96: 435-446

Paull J.G., Pallotta M.A., Langridge P. The T.T. RFLP markers associated with *Sr22* and recombination between chromosome 7A of bread wheat and the diploid species *Triticum boeoticum*. Theor Appl Genet 1994; 89: 1039-1045

Peng Y., Schertz K.F., Cartinhour S., Hart G.E. Comparative genome mapping of *Sorghum bicolor* (L.) Moench using an RFLP map constructed in a population of recombinant inbred lines. Plant Breed 1999; 118: 225-235

Penner G.A., Bush A., Wise R., Kim W., Dormier L., Kasha K., Laroche A., Scoles G., Molnar S., Fedak G. Reproducibility of random amplified polymorphic DNA (RAPD) analysis among laboratories. PCR Meth Appl 1993; 2: 341-345

Penner G.A., Zirino M., Kruger S., Townley-Smith F. Accelerated recurrent parent selection in wheat with microsatellite markers. In: A. E. Slinkard (ed.), *Proc 9th Intern Wheat Genet Symp* Vol 1, University E×tension Press, University of Saskatchewan, Saskatoon, Canada, 1998; pp 131-134

Pestsova E., Ganal M.W. Röder M.S. Isolation and mapping of microsatellite markers specific for the D genome of bread wheat. Genome 2000; 43: 689-697

Pestsova E., Ganal M.W., Roder M.S. Isolation and mapping of microsatellite markers specific for the D genome of bread wheat. Genome 2000; 43: 689-697

Phillips R.L., Vasil I.K. (eds.) *DNA-Based Markers in Plants*, Kluwer Academic Publishers, The Netherlands, 1994

Phillips R.L., Vasil I.K. (eds.) *DNA-Based Markers in Plants*, 2nd Edition, Kluwer Academic Publishers, The Netherlands, 2001

Pillen K., Binder A., Kreuzkam B., Ramsay L., Waugh R., Forster J., Leon J. Mapping new EMBL-derived barley microsatellites and their use in differentiating German barley cultivars. Theor Appl Genet 2000; 101: 652-660

Plaschke J., Börner A., Xie D.X., Koebner R.M.D., Schlegel R., Gale M.D. RFLP-mapping of genes affecting plant height and growth habit in rye. Theor Appl Genet 1993; 85: 1049-1054

Plaschke J., Korzun V., Koebner R.M.D., Börner A. Mapping of the GA_3-insensitive dwarfing gene *ct1* on chromosome 7R in rye. Plant Breed 1995; 114: 113-116

Portyanko V.A., Hoffman D.L., Lee M., Holland J.B. A linkage map of hexaploid oat based on grass anchor DNA clones and its relationship to other oat maps. Genome 2001; 44: 249-265

Powell W., Machray G.C., Provan J. Polymorphism revealed by simple sequence repeats. Trends Plant Sci 1996; 1: 215-222

Pratchett N., Laurie D.A. Genetic map location of the barley developmental mutant *liguleless* in relation to RFLP markers. Hereditas 1994; 120: 35-39

Qi L.-L., Echalier B., Friebe B., Gill BS. Molecular characterization of a set of wheat deletion stocks for use in chromosome bin mapping of ESTs. Funct Integr Genom 2003; 3: 39-55

Qi L.-L., Gill B.S. 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. Theor Appl Genet 2001; 103: 998-1006

Qi X., Stam P., Lindhout P. Comparison and integration of four barley genetic maps. Genome 1996; 39: 379-394

Qi X., Stam P., Lindout P. Use of locus-specific AFLP markers to construct a high- density molecular map in barley. Theor Appl Genet 1998; 96: 376-384

Rafalski A., Morgante M. Corn and humans: recombination and linkage disequilibrium in two genomes of similar size. Trends Genet 2004, 20: 103-111

Ragab R.A., Dronavalli S., Saghai Maroof M.A., Yu Y.G. Construction of a sorghum RFLP linkage map using sorghum and maize DNA probes. Genome 1994; 37: 590-594

Ramsay L., Macaulay M., Ivanissevich D.S., MacLean K., Cardle L., Fuller J., Edwards K.J., Tuvesson S., Morgante M., Massari A., *et al.* Simple sequence repeat-based linkage map of barley. Genetics 2000; 156: 1997-2005

Rayapati P.J., Gregory J.W., Lee M., Wise R.P. A linkage map of diploid Avena based on RFLP lici and locus conferring resistance to nine isolates of *Puccinia coronata* var. *avenae*. Theor Appl Genet 1994; 89: 831-837

Remington D.L., Thornsberry J.M., Matsuoka Y., Wilson L.M., Whitt S.R., Doebley J., Kresovich S., Goodman M.M., Buckler E.S. Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc Natl Acad Sci USA 2001; 98: 11479-11484

Riera-Lizarazu O., Vales M.I., Ananiev E.V., Rines H.W., Phillips R.L. Production and characterization of maize chromosome 9 radiation hybrids derived from an oat-maize addition line. Genetics 2000; 156: 327-339

Röder M.S., Korzun V., Gill B.S., Ganal M.W. The physical mapping of microsatellite markers in wheat. Genome 1998a; 41: 278-283

Röder M.S., Korzun V., Wendehake K., Plaschke J., Ti×ier M., Leroy P., Ganal M.W. A microsatellite map of wheat. Genetics 1998b; 149: 2007-2023

Rousset M., Brabant P., Kota R.S., Dubcovsky J., Dvorak J. Use of recombinant substitution lines for gene mapping and QTL analysis of bread making quality in wheat. Euphytica 2001; 119: 81-87

Rudd S. Expressed sequence tags: alternative or complement to whole genome sequences? Trends Plant Sci 2003; 87: 321-329

Ruge B. Roux S.R., Linz A., Wehling P. Erschließung und molekulare Charakterisierung von Resistenzen gegen Braunrost. Vortr Pflanzenzüchtg 1999; 46:169-176

Saal B., Wricke G. Clustering of amplified fragment length polymorphism markers in a linkage map of rye. Plant Breed 2002; 121: 117-123

Sadder M.T., Weber G.Comparison between genetic and physical maps in *Zea mays* L. of molecular markers linked to resistance against Diatraea spp. Theor Appl Genet 2002; 104: 908-915

Saji S., Umehara Y., Antonio B.A., Yamane H., Tanoue H., Baba T., Aoki H., Ishige N., Wu J., Koike K. *et al.* A physical map with yeast artificial chromosome (YAC) clones covering 63% of the 12 rice chromosomes. Genome 2001; 44: 32-37

Sandhu D., Gill K.S. Gene-containing regions of wheat and the other grass genomes. Plant Physiol 2002; 128: 803-811

Sandhu D., Sidhu D., Gill K.S. Identification of expressed sequence markers for a major gene-rich region of wheat chromosome group 1 using RNA fingerprinting-differential display. Crop Sci 2002; 42: 1285-1290

Sarma R.N., Fish L., Gill B.S., Snape J.W. Physical characterization of the homoeologous Group 5 chromosomes of wheat in terms of rice linkage blocks, and physical mapping of some important genes. Genome 2000; 43: 191-198

Sasaki T., Burr B. International rice genome sequencing project: the effort to completely sequence the rice genome. Curr Opin Plant Biol 2000; 3: 138-141

Schuler G.D., Boguski M.S., Stewart E.A., Stein L.D., Gyapay G., Rice K., White R.E., Rodriguez-Tome P., Aggarwal A., Bajorek E., Bentolila S., Birren B.B., Butler B. *et al.* A gene map of the human genome. Science 1996; 274: 540-546

Senft P., Wricke G. An extended genetic map of rye (*Secale cereale* L.). Plant Breed 1996; 115: 508-510

Serizawa N., Nasuda S., Shi F., Endo T.R., Prodanovic S., Schubert I., Kuenzel G. Deletion-based physical mapping of barley chromosome 7H. Theor Appl Genet 2001; 103: 827-834

Sharopova N., McMullen M.D., Schultz L., Schroeder S., Sanchez-Villeda H., Gardiner J., Bergstrom D., Houchins K., Melia-Hancock S., Musket T. *et al.* Development and mapping of SSR markers for maize. Plant Mol Biol 2002; 48: 463-481

Smilde D.W., Haluskova J., Sasaki T., Graner A. New evidence for the synteny of rice chromosome 1 and barley chromosome 3H from rice expressed sequence tags. Genome 2001; 44: 361-367

Soderlund C.S., Dunham H.A., French L. Contigs built with fingerprints, markers, and FPC V4.7. Genome Res 2000; 10: 1772-1787

Song Y.C., Gustafson J.P. The physical location of 14 RFLP markers in rice (*Oryza sativa* L.). Theor Appl Genet 1995; 90: 113-119

Song Q.J., Shi J.R., Singh S., Fickus E.W., Fernalld R., Gill B.S., Cregan P.B., Ward R. Development and mapping of wheat microsatellite markers. In: *Plant, Animal & Microbe Genomes X Conf*, San Diego, CA, USA, Jan 12-16, 2002, http://www.intl-pag.org/pag/10/abstracts/PAGX_P371.html

Sorrells M.E., La Rota M., Bermudez-Kandianis C.E., Greene R.A., Kantety R., Munkvold J.D., Miftahudin, Mahmoud A., Ma X., Gustafson P.J., *et al.* Comparative DNA sequence analysis of wheat and rice genomes. Genome Res 2003; 13:1818-1827

Sourdille P., Cadalen T., Guyomarc'h H., Snape J.W., Perretant M.R., Charmet G., Boeuf C., Bernard S. An update of the Courtot × Chinese Spring intervarietal molecular marker linkage map for the QTL detection of agronomic traits in wheat. Theor Appl Genet 2003; 106: 530-538

Sourdille P., Singh S., Cadalen T., Brown-Guedira G.L., Gay G., Qi L., Gill B.S., Dufour P., Murigneux A., Bernard M. Microsatellite-based delition bin system for the establishment of genetic-physical map relationships in wheat (*Triticum aestivum* L.). Funct Integr Genom 2004; 4:12-25

Sreenivasulu N., Kavikishor P.B., Varshney R.K., Altschmied L. Mining functional information from cereal genomes- the utility of expressed sequence tags (ESTs). Curr Sci 2002; 83:965- 973

Stam P. Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. Plant Jour 1993; 3: 739-74.

Stephenson P., Bryan G., Kirby J., Collins A., Devos K.M., Busso C., Gale M.D. Fifty new microsatellite loci for the wheat genetic map. Theor Appl Genet 1998; 97: 946-949

Stracke S., Schilling A.G., Glass C., Weiss E., Miedaner T., Geiger H.H. Development of PCR-based markers linked to dominant genes for male -fertility restoration in Pampa CMS of rye. Theor Appl Genet 2003; 106: 1184-1190

Subudhi P.K., Nguyen H.T. Linkage group alignment of sorghum RFLP maps using a RIL mapping population. Genome 2000; 43: 240-249

TAGI, The Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 2000; 408: 796-815

Tao Q.Z., Chang Y.L., Wang J.Z., Chen H.M., Islam-Faridi M.N., Scheuring C., Wang B., Stelly D.M., Zhang H.B. Bacterial artificial chromosome-based physical map of the rice genome constructed by restriction fingerprint analysis. Genetics 2001; 158: 1711-1724

Taylor C., Shepherd K.W., Langridge P. A molecular genetic map of the long arm of chromosomes 6R of rye incorporating the cereal cyst nematode resistance gene, *CreR*. Theor Appl Genet 1998; 97: 1000-1012

Thangavelu M., James A.B., Bankier A., Bryan G.J., Dear P.H., Waugh R. HAPPY mapping in plant genome: reconstruction and analysis of a high-resolution physical map of 1.9 Mpp region of *Arabidopsis thaliana* chromosome 4. Plant Biotech Jour 2003; 1:23-31

Temnykh S., Park W.D., Ayres N., Cartinhour S., Hauck N., Lipovich L., Cho Y.G., Ishii T., McCouch S.R. Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa* L.). Theor Appl Genet 2000; 100: 697-712

The Rice Chromosome 10 Sequencing Consortium: Yu Y., Rambo T., Currie J., Saski C., Kim H.R., Collura K., Thompson S., Simmons J., Yang T.-J., Nah G., *et al.* In-depth view of structure, activity, and evolution of rice chromosome 10. Science 2003; 300: 1566-1569

Thiel T., Michalek W., Varshney R.K., Graner A. Exploiting EST databases for the developement and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theor Appl Genet 2003; 106: 411-422

Thomas W.T.B. Prospects for molecular breeding of barley. Ann Appl Biol 2003; 142: 1-12

Thornsberry J.M., Goodman M.M., Doebley J., Kresovich S., Nielsen D., Buckler E.S. *Dwarf8* polymorphisms associate with variation in flowering time. Nature Genet 2001; 28: 286-289

Umehara Y., Inagaki A., Tanoue H., Yasukouchi Y., Nagamura Y., Saji S., Otsuki Y., Fujimura N., Kurata N., Minobe Y. Construction and characterization of rice YAC library for physical mapping. Mol Breed 1995; 1: 79-89

Yuan, Q., Liang F., Hsiao J., Zismann V., Benito M.I., Quackenbush J., Wing R., and Buell R. Anchoring of rice BAC clones to the rice genetic map in silico. Nucl Acids Res 2000; 28: 3636-3641

Van Deynze A.E., Dubcovsky J., Gill K.S., Nelson J.C. Sorrells M.E., Dvorak J. Gill B.S., Lagudah E.S., McCouch S.R., Appels R. Molecular-genetic maps for group 1 chromosomes of Triticeae species and their relation to chromosomes in rice and oat. Genome 1995, 38: 45-59

Varshney R.K., Kumar A., Balyan H.S., Roy J.K., Prasad M., Gupta P.K. Characterization of microsatellites and development of chromosome specific STMS markers in bread wheat. Plant Mol Biol Rep 2000; 18: 5-16

Varshney R.K, Prasad M., Graner A. Molecular marker maps of barley: a resource for intra- and interspecific genomics a. In: G. Wenzel, L. Horst (eds.), *Molecular Markers in Improvement of Agriculture and Forestry*, Spinger Verlag, Germany, 2004; in press

Varshney R.K., Prasad M., Roy J.K., Roder M.S., Balyan H.S., Gupta P.K Integrated physical maps of 2DL, 6BS and 7DL carrying loci for grain protein content and pre-harvest sprouting tolerance in bread wheat Cereal Res Commun 2001; 29: 33-40

Varshney R.K., Thiel T., Stein N., Langridge P., Graner A. *In silico* analysis on frequency and distribution of microsatellites in ESTs of some cereal species. Cell Mol Biol Lett 2002; 7: 537-546

Voylokov A.V., Korzun V., Börner A. Mapping of three self-fertility mutations in rye (*Secale cereale* L.) by using RFLP, isozyme and morphological markers. Theor Appl Genet 1998; 97: 147-153

Vuylsteke M., Mank R., Antonise R., Bastiaans E., Senior M.L., Stuber C.W., Melchinger A.E., Lubberstedt T., Xia X.C., Stam P. *et al.,* Two high-density AFLP (R) linkage maps of *Zea mays* L.: analysis of distribution of AFLP markers. Theor Appl Genet 1999; 99: 921-935

Wall J.D., Pritchard J.K. Haplotype blocks and linkage disequilibrium in the human genome. Nature Rev Genet 2003; 4: 587- 597

Wang Z.X., Idonuma A., Umehara Y., van Houten W., Ashikawa I., Minobe Y., Kurata N., Sasaki T. Physical mapping of rice chromosome 1 with yeast artificial chromosomes (YACs). DNA Res 1996; 3:291-296

Wanous M.K., Goicoechea P.G., Gustafson J.P. RFLP maps of rye chromosomes 6R and 7R including terminal C-bands. Genome 1995; 38: 999-1004

Wardrop J., Snape J.W., Powell W., Machray G.C. Constructing plant radiation hybrid panels. Plant Jour 2002; 31:223-228

Waugh R., Dear P.H., Powell W., Machray G.C. Physical education - new technologies for mapping plant genomes. Trends Plant Sci 2002; 7:521-523

Weber D., Helentjaris T. Mapping RFLP loci in maize using B-A translocations. Genetics 1989; 121: 583-590

Weng Y., Lazar M.D. Comparison of homoeologous group-6 short arm physical maps of wheat and barley reveals a similar distribution of recombinogenic and gene-rich regions Theor Appl Genet 2002; 104: 1078-1085

Weng Y., Tuleen N.A., Hart G.E. Extended physhical map of the homoeologous group-6 chromosomes of wheat (*Triticum aestivum* L.). Theor Appl Genet 2000; 100: 519-527

Werner J.E., Endo T.R., Gill B.S. Toward a cytogenetically based physical map of the wheat genome. Proc Natl Acad Sci USA 1992; 89: 11307-11311

Wricke G., Dill P., Senft P. Linkage between a major gene for powdery mildew resistance and an RFLP marker on chromosome 1R of rye. Plant Breed 1996; 115: 71-73

Wu J., Kurata N., Tanoue H., Shimokawa T., Umehara Y., Yano M., Sasaki T. Physical mapping of duplicated genomic regions of two chromosome ends in rice. Genetics 1998 150: 1595-1603

Wu J.Z., Maehara T., Shimokawa T., Yamamoto S., Harada C., Takazaki Y., Ono N., Mukai Y., Koike K., Yazaki J., *et al.* A comprehensive rice transcript map containing 6591 expressed sequence tag sites. Plant Cell 2002; 14: 525-535

Wu K.S., Tanksley S.D. Genetic and physical mapping of telomers and microsatellites of rice. Plant Mol Biol 1993; 22: 861-872

Xie D.X., Devos K.M., Moore G., Gale M.D: RFLP-based genertic maps of the homoeologous group 5 chromosomes of bread wheat (*Triticum aestivum* L.). Theor Appl Genet 1993; 87: 70-74.

Xu G.W., Magill C.W., Schert K.F., Hart G.E. A RFLP linkage map of *Sorghum bicolor* (L.) Moench. Theor Appl Genet 1994; 89: 139-145

Yim Y., Davis G., Duru N., Musket T., Linton E., Messing J., McMullen M., Soderlund C., Polacco M., Gardiner J., Coe E.Jr. Characterization of three maize bacterial artificial chromosome libraries toward anchoring of the physical map to the genetic map using high-density bacterial artificial chromosome filter hybridization. Plant Physiol 2002; 130:1686-1696

Yoshimura A, Ideta O, Iwata N. Linkage map of phenotype and RFLP markers in rice.Plant Mol Biol 1997; 35: 49-60

Young N.D. Constructing a plant genetic linkage map with DNA markers. In: Phillips R.L. and Vasil I.K. (eds), DNA-Based Markers in Plants, Kluwer Academic Publishers, The Netherlands 2001, pp.31-48

Yuan Q.P., Liang F., Hsiao J., Zismann V., Benito M.I., Quackenbush J., Wing R., Buell R. Anchoring of rice BAC clones to the rice genetic map in silico. Nucl Acids Res 2000; 28: 3636-3641

Zhang H.N., Nasuda S., Endo T.R. Identification of AFLP markers on the satellite region of chromosome 1BS in wheat. Genome 2000; 43: 729-735

Zhang X.Q., Ross K., Gusatfson J.P. Physical location of homoeologous groups 5 and 6 molecular markers mapped in *Triticum aestivum* L. Jour Heredity 2000; 91:441-445

Zhao Q., Zhang Y., Cheng Z., Chen M., Wang S., Feng Q., Huang Y.C., Li Y., Tang Y.S., Zhou B., *et al.* A fine physical map of the rice chromosome 4. Genome Res 2002; 12: 817-823

Zhu Y.L., Song Q.J., Hyten D.L., Van Tassell C.P., Matukumalli L.K., Grimm D.R., Hyatt S.M., Fickus E.W., Young N.D., Cregan P.B. Single-nucleotide polymorphisms in soybean. Genetics 2003; 163: 1123-1134

Zwick M.S., Islam-Faridi M.N., Czeschin D.G., Wing R.A., Hart G.E., Stelly D.M., Price H.J. Physical mapping of the liguleless linkage group in *Sorghum bicolor* using rice RFLP-selected sorghum BACs. Genetics 1998; 148: 1983-1992

Chapter 4

ORGANIZATION OF RETROTRANSPOSONS AND MICROSATELLITES IN CEREAL GENOMES

Alan H. Schulman[1,*], Pushpendra K. Gupta[2] and Rajeev K. Varshney[3]

[1]*MTT Agrifood Research, Jokioinen, and University of Helsinki, Helsinki, Finland*; [2]*Ch. Charan Singh University, Meerut, U.P., India*; and [3]*Institute of Plant Genetics and Crop Plant Research (IPK), Correnssstrasse 3, D-06466 Gatersleben, Germany*
[*]Author for correspondence: alan.schulman@helsinki.fi

1. INTRODUCTION

Our understanding of genome organization has its roots in postwar interest in the effects of radiation. Observations on the relationship between doses of ionizing radiation and the frequency of mutations (Abrahamson *et al*., 1973; Trujillo and Dugan, 1975) indicated that the size of the genetic target receiving the radiation dose varied considerably between organisms. By the beginning of the 1970s, this phenomenon had come to be known as the "*C*-value paradox" (Thomas, 1971). The paradox was that the total genome size, or *C*-value, varied widely within a given clade of organisms and bore no relationship to organismal complexity. For example, two legumes within the same genus, *Vicia faba* and *Vicia sativa*, have haploid genomes of 13.1 x 10^9 and 2.2 x10^9 respectively, but differ very little morphologically. This observation has been fully confirmed by the large-scale determination of genome sizes of many plants (http://www.rbgkew.org.uk/cvalues), within which genome size varies from about 10^7 bp in *Cardamine* and Arabidopsis among the Crucifereae to nearly 10^{11} bp in *Fritillaria* among the Lilliaceae. Within the cereals, rice (*Oryza sativa*) has a compact genome of 4.8 x 10^8 bp, ranging upwards through sorghum (7.35 x 10^8), maize (26.7 x 10^8), and barley (54.4 x 10^8 bp).

P.K. Gupta and R.K. Varshney (eds.), Cereal Genomics, 83-118.

Resolution of the *C*-value paradox began with experiments carried out in the 1970s and early 1980s (*e.g.*, Flavell *et al.*, 1974; Hake and Walbot, 1980). Sheared DNA was melted and then allowed to reanneal over protracted periods of time to various $C_o t$ (combinations of concentration and time) levels. The experiments showed that genomes of higher organisms are comprised of single-copy DNA regions interspersed between repetitive DNA of various degrees of redundancy. These and later studies have led to the view that, in general, larger eukaryotic genomes have a high proportion of repetitive DNA. The maize nuclear genome for instance, contains about 60 to 80% repetitive DNA (Flavell *et al.*, 1974; Bennetzen *et al.*, 1994; Springer *et al.*, 1994; Heslop-Harrison, 2000; Meyers *et al.*, 2001) and the wheat nuclear genome contains 83 % repetitive DNA (Wicker *et al.*, 2001). These studies were, however, essentially static views of the genome lacking details of structure and function, and left unexplained why, within narrow clades of plants, genome sizes should show such variation.

A final resolution of the question of why eukaryotic genomes might vary in size derives from several fields of research. First, large-scale sequencing of ESTs from many plants and of the entire genomes of mouse, human, Arabidopsis, and rice indicates that the number of cellular genes is relatively constant, some 30,000 to 50,000 (TAGI, 2000; Goff *et al.*, 2002; Kurata *et al.*, 2002; Yu *et al.*, 2002). Given 30,000 genes with 4800 bp, 60 % of the rice genome but only 2.7 % of the barley genome encodes cellular functions.

Second, sequencing of long contiguous segments ("contigs") of the genome, such as in assembled BAC and YAC clones (respectively, Bacterial Artificial Chromosome and Yeast Artificial Chromosome), confirmed the hybridizaton results that much of the intervening DNA between cellular genes consists of repetitive DNA. Repetitive sequences can be found in the genome either in tandem arrays or in a dispersed fashion and can be classified into 3 categories: (a) transposable elements, which are mobile genetic elements, largely consisting of retrotransposons, in many plant species; (b) microsatellite sequences, which are tandemly repeated DNA sequences (also called simple sequence repeats); and (c) a special class of repetitive sequences including telomeric and centromeric sequences as well as tandem arrays of rDNA units.

Microsatellites and particularly retrotransposons are dynamic components of the genome whose abundance can change over time. As detailed below, the dynamics of retrotransposon replication and insertion between cellular genes can explain much of the size differences between otherwise similar plant genomes. Furthermore, cereal genomes are largely syntenic and colinear, having largely the same genes on homeologous chromosomes in mostly the

same order (Smilde *et al.*, 2001). The major part of cereal genomes is comprised of retrotransposons and microsatellites, and the major factor in the differences between their genome sizes is the varying proportions of these elements. Therefore, in the present chapter, we discuss the distribution and organization of retrotransposons and microsatellites and their impact on genome size in cereals.

2. RETROTRANSPOSONS AND THEIR ORGANIZATION WITHIN THE GENOME

Although the study of retroelements can be dated to about 1910 with the discoveries of avian leucosis virus (alv) and Rous sarcoma virus (Rous, 1911), their true nature was mysterious before the provirus hypothesis of Temin (Temin, 1964) was confirmed with the discovery of reverse transcriptase (Baltimore, 1970; Temin and Mizutani, 1970). The mode of retroviral replication was worked out during the 1970s, culminating in the first sequence of a retrovirus (Shinnick *et al.*, 1981). The presence of retroelements in vertebrate genomes, termed endogenous retroviruses, was subsequently established (Löwer *et al.*, 1996). By 1985, yeast was demonstrated to have endogenous elements, which were named *Ty*, transposing, as do the retroviruses, through an RNA intermediate (Boeke *et al.*, 1985; Kingsman and Kingsman, 1988). The presence and activity of similar elements in insects (Flavell *et al.*, 1980; Levis *et al.*, 1980; Shiba and Saigo, 1983) and in the plants (Shepherd *et al.*, 1984; Grandbastien *et al.*, 1989) was shown in the same period. The organization of retrotransposons in plant genomes has been investigated on four levels of increasing precision: Southern- and dot-blot hybridization studies giving a general view of copy number; *in situ* hybridization of chromosomes, yielding a picture of the distribution by chromosome region; mapping and sequencing of large contiguous segments of the genome, such as are present in BACs and YACs, detailing local organization; large-scale sequencing extending over whole genomes, making possible both a detailed local and global view.

2.1. Retrotransposon Structure

Retrotransposons are transposable elements that replicate via an RNA intermediate in a process involving transcription and reverse transcription, and are labeled Class I. Class II transposable elements are DNA transposons, which replicate only as part of the chromosome and move by a cut-and-paste mechanism. The Class I elements fall into three distinct groups. These are:

LINEs, or long-interspersed elements; SINEs, or short interspersed elements; LTR retrotransposons, which are bounded by long terminal repeats (LTRs). The LTR retrotransposons can be further divided into two categories, the *copia*-like and the *gypsy*-like elements. Members of each class display a greater similarity to members of the same class from different species and Kingdoms than to elements from the same species, supporting their ancient origin (Xiong and Eickbush, 1990). The three retrotransposon groups found in plants each occur in the animals and fungi as well, and therefore represent ancient genomic components that predate the divergence of these three major groups of eukaryotes.

The structure and life cycle of the plant LTR retrotransposons (reviewed in Kumar and Bennetzen, 1999) most resembles that of the mammalian retroviruses. The bounding LTRs contain the signals needed for RNA expression and processing. The 5' LTR serves as the promoter, whereas the 3' LTR functions as the terminator and polyadenylation signal. The LTRs can vary in length from several 100 bp (*Tos17*; Hirochika *et al.*, 1992) to over 5 kb (*Sukkula*; Shirasu *et al.*, 2000) and are themselves terminated by small inverted repeats giving the LTRs a universal 5' TG….CA 3' structure. Adjacent, respectively, to the 5' and 3' LTRs are the priming sites for the (-)-strand and (+)-strands of the cDNA that are recognized by the reverse transcriptase (RT) encoded by the retrotransposon. In addition to RT, LTR retrotransposons encode the same basic proteins, generally excepting the envelope protein, as do retroviruses: capsid protein (GAG), which forms the virus-like particle (VLP), aspartic proteinase (AP), which cleaves the expressed polyprotein into functional components, integrase (IN), which carries out the insertion of the cDNA into the genome, and RNase H (RH), which is required in replication (Turner and Summers, 1999). In the *copia*-like retrotransposons, the domains are organized 5' LTR-GAG-AP-IN-RT-RH-LTR 3', whereas *gypsy*-like elements are arranged as 5' LTR-GAG-AP-RT-RH-IN-LTR 3'. The GAG is often, though not necessarily, in a different reading frame than the other domains, which are expressed as single polyprotein referred to as POL. A newly recognized subset of the plant *gypsy*-like retrotransposons contains a coding domain (*env*), internal to the 3' LTR, specifying a putative envelope protein (Vicient, *et al*., 2001a). Similar domains have been found in *copia*-like elements as well (Laten *et al.*, 1998; Wright and Voytas, 2002). The *env*-bearing plant elements of the *gypsy*-like class are organized in a manner identical to the mammalian retroviruses, raising the still-unanswered question of their function as viruses. A new class of non-autonomous elements derived from retrotransposons but lacking protein-coding domains has been reported (Witte *et al.*, 2001).

The LTR retrotransposons are the most abundant and best-studied class of plant retroelements, but the LINEs and SINEs are important genomic components as well (Noma *et* al., 1999; Schmidt, 1999; Vershinin *et al*., 2002). The LINE elements are viewed as progenitors of the LTR retrotransposons, although the two groups replicate by different mechanisms (Eickbush, 1992). The LINE elements lack both LTRs and IN, and at least in mammals replicate via transcription from an internal promoter, followed by reverse transcription primed by genomic DNA at a cleaved target site (Luan *et al*., 1993; Weiner, 2002). Integration likely takes place via a break repair mechanism (Moore and Haber, 1996). In humans, the L1 family of LINE elements is a major mutagenic agent and responsive for genome remodeling via extensive deletions, at least in transformed cells (Kazazian and Goodier, 2002). Although few plant LINEs appear functional (Schmidt, 1999), the recently reported Karma element in rice can be activated by induced hypomethylation in tissue culture, leading to copy number increases in subsequent generations (Komatsu *et al.*, 2003).

The SINEs, in contrast to LINEs, encode no enzymatic functions, and appear to rely on those of LINEs for their propagation (Boeke, 1997; Lenoir *et al*., 2001). The SINEs are reverse-transcribed from RNA polymerase III products and, with the exception of some mammalian elements, are derived from tRNA; as such, they contain internal promoters that can render them transcriptionally active elements (Deragon and Capy, 2000). The SINE elements are mobile within the plants, and have been well studied in the genus Brassica (Gilbert *et* al., 1997; Arnaud *et al.* 2000; Tikhonov *et al.*, 2001). They have been used for phylogenetic markers in both Brassica and *Oryza*, in humans, and in many animals (Cheng *et al.*, 2002).

Since the discovery of the first active plant retrotransposon (Grandbastien *et al.*, 1989), considerable progress has been made in demonstrating that the various steps of the life cycle of LTR retrotransposons previously established for yeast and flies are carried out in the plants (Kumar and Bennetzen, 1999; Vicient *et al.*, 1999a). Among the cereals, transcription of retrotransposons appears to be a common phenomenon, corresponding to about 0.1% of total transcripts in EST databases (Vicient *et al.,* 2001b; Echenique, *et al.*, 2002). In the cereals, transcription has been directly demonstrated in rice (Hirochika, 1993; Hirochika *et al.,* 1996), maize (Turcich *et al.*, 1996), and barley (Suoniemi *et al.,* 1996a; Vicient *et al.,* 2001a). Translation, polyprotein processing, and formation of virus-like particles have been explicitly demonstrated in barley and other cereals (Jääskeläinen *et al.,* 1999; Vicient *et al.* 2001b) for the abundant *BARE*-1 retrotransposon family (Manninen and Schulman, 1993; Vicient *et al.*, 1999a) and transcriptional splicing has been for the retroviral-like *Bagy2* element (Vicient *et al.*, 2001a). Stress and tissue

culture activation of retrotransposons appears to be a general phenomenon (Wessler, 1996; Grandbastien, 1998), and has been well analyzed in tobacco and rice (Hirochika *et al.*, 1992; Hirochika *et al.*, 1996; Takeda *et al.*, 1998; Beguiristain *et al.*, 2001). In barley, although stress activation has not been directly demonstrated, both retrotransposon *BARE*-1 copy number and genome size is correlated with environmental factors associated with drought and temperature stress (Kalendar *et al.*, 2000).

2.2. Retrotransposon Copy Number Variation and its Relationship to Genome Size

Given actively replicating families of retrotransposons in plant genomes and gene numbers varying over a relatively small range, much of the considerably greater variation observed in genome size among the flowering plants can be attributed to variation in the bulk contribution of retrotransposons to the total genome. This appears true for eukaryotes generally; the compact genome of the yeast *Saccharomyces cerevisiae*, comprising 13 x 10^6 bp, contains 331 retroelements, of which only 51 are full-length retrotransposons (Kim *et al*., 1998). The great majority of the elements in yeast are, instead, solo LTRs, the remnant of recombinational loss of a retrotransposon. Occurrence of solo LTRs shall be revisited below with respect to the cereals. Taken together, the full-length and solo LTRs comprise 3.1 % of the yeast genome.

The genome of *Arabidopsis thaliana*, at 10^8 bp, is the smallest described in the plants (Goodman *et al.*, 1995). A careful survey of 1.7 x 10^7 bp (Le *et al.*, 2000) revealed the presence of 50 groups of LTR retrotransposons, divided about equally between gypsy-like and copia-like elements, each with about two members on average. In addition, three groups, or more correctly two groups (Lenoir *et al.,* 2001), of SINEs and 28 groups of LINEs were found, the latter with about one member per group. Extrapolating from the 134 Class I elements identified to the whole genome yields an estimated total of 779 elements in *A. thaliana*, comprising about 2 % of the DNA.

The cereals tend to have large genomes, and commensurately more retrotransposons. Cultivated rice has one of the smallest genomes among the cereals, containing 4.3 x 10^8 bp (Kurata *et al.*, 1997). Dispersed within the rice genome is about 10^3 retrotransposons when estimated by hybridization with a conserved probe (Hirochika *et al.*, 1992). The rice genome project has recently completed three chromosomes representing 100.18 Mb to date, about 23% of the total (Sasaki *et al.,* 2002; Feng *et al.,* 2002; The Rice

Chromosome 10 Sequencing Consortium, 2003). Annotation of the amassed sequence (supplementary material at http://sciencemag.org/cgi/data/300/5625/1566/DC1/1) reveals 11,423 retrotransposons covering 8.27 Mb. Extrapolating to the whole genome, this indicates a total of 4.9 x 10^4 retrotransposons covering 35 Mb, or 8.3% of the entire genome. The discrepancy with earlier estimates may be due in part to the annotation being based on scoring of LTRs rather than coding domains, and on the limitations of hybridization for detecting highly degenerate retroelements. An analysis of the genome size and retrotransposon content of barley, wild barley, and other species within the genus *Hordeum* confirmed the broad correlation of genome size with retrotransposon copy number (Vicient *et al.*, 1999b).

2.3. Chromosomal Distribution of Retrotransposons

The high copy number of retrotransposons in cereal and other genomes raises the question of how their bulk is distributed within the genome. The method of fluorescent *in situ* hybridization (FISH) offers an efficient strategy for acquiring a broad overview of genome organization, and has been applied to many retrotransposon systems. The 1 – 2 x 10^4 copies of the *BARE*-1 retrotransposon, for example, are dispersed along all seven haploid chromosomes of barley and related species, except in the centromeres, telomeres, and nucleolar organizing regions (Suoniemi *et al.*, 1996b; Vicient *et al.*, 1999b). The distribution of the wheat retrotransposon WIS 2-1A, which is sufficiently similar to *BARE*-1 to consider it a member of the *BARE*-1 family (Manninen and Schulman, 1993), was examined in the chromosomes of wheat and rye that are present in the allopolyploid triticale (x*Triticosecale*) by FISH (Muñiz *et al*., 2001). The elements were dispersed on all chromosomes of both constituent genomes in a manner similar to *BARE*-1 in barley. Likewise, a retroelement fragment from the oat genus *Avena*, 67 % and 70 % similar to *BARE*-1 and WIS-2-1A LTRs respectively, is dispersed along all chromosomes of diploid, tetraploid, and hexaploid species except in the centromeric and nucleolar organizer regions (Katsiotis *et al.*, 1996; Linares *et al.*, 1999). Other *copia*-like (Moore *et al.*, 1991; Svitashev *et al.*, 1994) and *gypsy*-like (Vershinin *et al*., 2002) elements show a distribution in barley broadly similar to that of *BARE-1*.

Few studies have been carried out on the distribution of the non-LTR retrotransposons, the LINEs and SINEs, in cereal genomes. A comparison of the distribution of *copia*-like, *gypsy*-like, and LINE elements in the wild barley *Hordeum spontaneum*, and in *Aegilops* speltoides, another member of the tribe Triticeae was made (Belyayev *et al.*, 2001). A clustered distribution

of *copia*-like elements was observed in *Ae. speltoides* and two main clusters on different chromosomes found in *H. spontaneum*. The *gypsy*-like elements were present as clusters in both species, whereas LINE elements showed a broader, but nevertheless clustered, distribution. Vershinin and coworkers (2002) reported on LINE distribution in several *Hordeum* species. They observed dispersion along most chromosomes, limited by the low copy numbers in cultivated barley, *H. vulgare*.

Although most families of retrotransposons in the cereals are excluded from the centromere and the nucleolar organizing regions, or are at least rare enough there to give weak FISH signals, some families show converse distributions. One of these, a gypsy-like element named *CEREBA*, is highly localized, but not exclusive, to cereal centromeric regions (Presting *et al.*, 1998; Hudakova *et al.*, 2001). The *gypsy*-like *RIRE7* element of rice also has reported centromeric association (Kumekawa *et al.*, 1999, 2001a; Nomomura *et al.*, 2001). In maize, the CentA retroelement is confined to centromeric regions, whereas the Huck and Prem2 retrotransposons with which it is associated were found not only in the centromeres but also elsewhere on the chromosomes (Ananiev *et al.*, 1998a). In addition, there are a number of other partially characterized retroelements that have been found within cereal centromeres (Dong *et al.* 1998; Miller *et al.* 1998; Nomomura and Kurata, 1999; Langdon *et al.*, 2000; Fukui *et al.*, 2001). A parallel situation occurs in Arabidopsis, where sequenced contigs for the centromeres of chromosomes 4 and 5 reveal a core of 180 bp repeats interrupted by the insertion of *gypsy*-like *Athila* retrotransposons (Kumekawa *et al.*, 2000,2001b). It is tempting to speculate that a single ancient family of *gypsy*-like elements is found in the centromeres of diverse plants, given that such elements can interact directly with conserved kinetochore proteins (Zhong *et al.*, 2002). As the sequences of more centromeres are determined through improved methods and our understanding of centromere function increases (Lamb and Birchler, 2003), a clearer picture will no doubt emerge.

Centromeres, telomeres, and nucleolar organizers have specific functions associated with particular classes of repetitive DNA (Pardue *et al.*, 1996). The absence of all but a few specific families of retrotransposons from these regions may be evidence for strong selective pressure to maintain structural integrity. The telomeres, in particular, appear retrotransposon-poor. However, there are the startling cases of HeT-A and TART in *Drosophila*. These non-LTR retrotransposons are exclusively found at telomeres as long head-to-tail arrays, where they deliver the essential replicative functions needed to maintain the chromosomal ends (Danilevskaya *et al.*, 1998; Casacuberta and Pardue, 2003). Integration at the correct site is orchestrated by the targeting of Gag to the telomeric regions (Rashkova *et al.*, 2003).

Whether similar roles are played by the retrotransposons that are present in structurally critical parts of plant chromosomes, or whether instead those families of retrotransposons are merely not sufficiently deleterious to be excluded, remains to be established.

2.4. Local Organization of Retrotransposons and Genes

Although *in situ* hybridization gives a general view of which compartments in the genome are favored insertion sites for particular classes of retrotransposons, its resolution is too low to reveal detail on how retrotransposons are situated with respect to genes. The advent of large-scale sequencing has now provided an opportunity to carry out in-depth organizational analyses of genes and retrotransposons on contiguous chromosomal segments extending over tens or hundreds of kilobases. The general view has emerged, consistent with earlier data derived from analytical centrifugation, that the genomes of cereals, and grasses more generally, are segmented into compact islands of genes, "gene space," surrounded by large expanses of repetitive DNA "seas" largely comprised of retrotransposons (Barakat *et al.*, 1997; Sandhu and Gill, 2002; see Chapter 12 of this volume also). These gene blocks appear to vary widely in size. The analysis of a 280 kb segment of the maize genome containing the *Adh1F* allele revealed long stretches of retrotransposons flanking two individual genes and amounting to 60% of the region (SanMiguel *et al.*, 1996). The retrotransposons, furthermore, were present as nests several elements deep, rather than as concatenated blocks. A later analysis of a 225 kb segment in the *Adh* region revealed retrotransposon stretches from 14 to 70 kb surrounding five single genes and a 39 kb block containing four putative genes and no retroelements (Tikhonov *et al.*, 1999). The *bz* locus of maize, which has a very high rate of recombination, is present on a unusually long retrotransposon-free expanse, 32 kb, containing ten genes (Fu *et al.*, 2001).

Detailed analyses of chromosome segments in cereals other than maize have extended and confirmed the model of cereal genomes as consisting of gene islands surrounded by retrotransposon seas. In a 60 kb stretch of the barley genome from chromosome 4HL that contains the *Mlo* locus, the Schulze-Lefert group found three genes clustered in a block of 18 kb (Panstruga *et al.*, 1998). An almost identical arrangement was found on a 66 kb contig from chromosome 2HL, which contains three genes in an interval of 18 kb flanked by a region that has undergone at least 15 retrotransposon integrations (Shirasu *et al*., 2000). A contig of 261 kb was sequenced from the barley *Mla* locus specifying resistance to powdery mildew (Wei *et al*., 2002). This segment contains 32 predicted genes organized into three gene islands

separated by two complexes of nested retrotransposons. The *BARE*-1 retrotransposon alone accounts for 17.5 % of the total contig. Analysis of four BACs, covering 417.5 kb, confirmed the gene island model of the barley genome (Rostoks *et al.*, 2002). The finishing and full annotation of the rice genome will give the first complete view of the organization of a cereal genome (Goff *et al.*, 2002). Initial glimpses, reconstructions, and annotation of sequenced regions (Nagano *et al.*, 1999; Yu *et al.*, 2002; The Rice Chromosome 10 Sequencing Consortium, 2003) indicate that the organization described for the few contigs of other cereals largely holds for the rice genome as a whole.

Most long segments of DNA that have been analyzed in detail are derived from genic regions. The cloning and sequence assembly of centromeres and other non-genic structures has been generally difficult, and these are under-represented both in general databases and in genome projects. As discussed above, the centromeres of maize contain tandem arrays of CentC elements, interrupted by variable numbers of centromere-specific CentA retrotransposons (Ananiev *et al.*, 1998a). Maize chromosomes possess as well, depending on the variety, unusual heterochromatic elements descriptively called knobs. An analysis of knob sequences revealed that retrotransposons interrupt tandem arrays of diagnostic 180 bp repeats (Ananiev *et al.*, 1998b). Huck elements were not found in any knob segments examined, although other elements such as Prem2 and Zeon were present and Grande seemed particularly abundant.

2.5. Retrotransposons and the Evolution of Genome Organization

Variations in the copy number of retrotransposon families within a species and between related species demonstrate that these elements dynamically contribute to genome evolution over time. To understand these changes, several aspects must be considered: integration site preference; the balance between gain and loss of integrated elements; selective forces acting on both copy number and integration pattern. One approach to reconstructing the evolutionary role of retrotransposons is to compare the sequences of equivalent regions of the genome from different cereals. This is aided by the twin phenomena of colinearity and synteny within the cereals, meaning that the order and content of genes is generally conserved over distances extending to the length of full chromosomes (Moore *et al.*, 1993; Bennetzen and Freeling, 1997). Sorghum and maize are ideal for comparison because of the fine-scale colinearity of their genes despite the maize genome being 3.5-

fold larger (Chen *et al.*, 1997). Sequencing and comparison of the *adh1* region in sorghum and maize showed that the retrotransposon nests comprising over 70% of the contig in maize are all missing from the orthologous region in sorghum (SanMiguel and Bennetzen, 1998; Gaut *et al.*, 2000). Similarly, the *a1* and *sh2* loci are separated by 140 kb in maize, but by only about 20 kb in sorghum and rice; most of the difference appears due to retrotransposon nests in maize (Chen *et al.*, 1998). Comparison of the LTR sequences indicated that the retrotransposons that expanded the *adh1* region in maize integrated between two and six million years ago, following the divergence of maize and sorghum 15 to 20 million years ago (SanMiguel *et al.*, 1998). Using similar estimation methods, Wei *et al.* (2002) estimated that the *Mla* locus of barley was invaded by five *BARE*-1 elements over the last 2 million years.

A major issue raised by evidence of massive genome expansion through integration of retrotransposons is the unidirectionality of the process: do genomes expand without end (Bennetzen and Kellogg, 1997a,b) Processes such as unequal crossing-over and deletion may remove retrotransposons, but do not appear sufficient alone to reverse the process of expansion. A third mechanism, LTR – LTR recombination, may play a major role. This removes one LTR and the internal domain of a retrotransposon from the genome and leaves behind a solo LTR. The genomes of *Hordeum* species, but apparently not that of maize, has 7 to 42 –fold more LTRs than full-length retrotransposons, corresponding in one barley cultivar to a total of 6×10^4 solo LTRs (Vicient *et al.*, 1999b). A striking contrast between the nests of retrotransposons observed in maize (SanMiguel *et al.*, 1996) and barley (Shirasu *et al.*, 2000) is that those in barley are comprised of solo LTRs but the nests in maize consist of full-length elements. In the compact yeast genome, likewise, 85% of the retroelements are present as solo LTRs (Kim *et al.*, 1998). The capacity of LTRs to recombine not only within a single retrotransposon, but also between elements along a chromosome arm, may well be a selective force for the gene island, repeat sea organization of cereal genomes. When LTRs from different elements recombine, the intervening DNA segment is lost. Selection would be unlikely to favor loss of genes by such a process, leading to segregation of retrotransposons and genes into different genomic compartments.

The dynamics of retrotransposon gain and loss in various regions of the genome has produced, much as for natural archipelagos, an array of both small and large gene islands separated by expanses of retrotransposons of varying sizes. These variations are reflected in differences in gene density observed in various contigs. For barley, Panstruga *et al.* (1998) and Shirasu and colleagues (2000) found a density of one gene per 20 kb overall, with

gene islands reaching one gene per 6 kb, Wei and coworkers (2002) observed average and island densities respectively of one gene per 8.1 kb and 4.6 kb, whereas Rostocks *et al.* (2002) reported an average of one gene per 21 kb, with variation from one gene per 12 kb to one gene per 103 kb. In maize, average densities of one gene per 50 kb, derived from islands of one per 9.8 kb surrounded by retrotransposon expanses of 14 to 70 kb, have been reported (Tikhonov *et al.*, 1999), though the *bz* locus and zein cluster reach densities as high as one gene per 3.2 kb and 5.2 kb respectively (Llaca and Messing, 1998; Fu *et al.*, 2001). Small gene islands of densities higher than one gene per 6 kb have been reported for barley, wheat, and rice as well (Feuillet and Keller, 1999), approaching the density of one gene per 4.5 kb found in the retrotransposon-poor *Arabidopsis thaliana.* The variability between different maize cultivars in the number of CentA elements in centromeres and retrotransposons in chromosomal knobs indicates that these regions, too, are subject to retrotransposon invasion and dynamic change over time (Ananiev *et al*., 1998a,b). Currently, we do not know if these widely varying densities represent a snapshot of a process of ongoing division of gene islands by newly forming retrotransposon nests, or if accumulation of retrotransposons is selected against at some loci in cereal genomes.

Retrotransposons depend, for their evolutionary survival, on being active enough transcriptionally that sufficiently many functional copies will be inherited to insure against loss and mutational inactivation. Insertion in silent heterochromatin does not favor this goal, but reducing host fitness through integrating into genes seems counterproductive as well. Integrating into nests near functional genes, however, may offer retrotransposons a solution. The specific means, if they exist, in cereal or other plant genomes to target many retrotransposons into nests of other retrotransposons remain to be defined. However, in the compact yeast genome, mechanisms have evolved to target integration into narrowly defined regions (Zhou *et al.*, 1996; Boeke and Divine, 1998). The *Ty3* retrotransposon, for example, is directed to its insertion site near tRNA genes by the affinity of its integrase to transcription factors of RNA polymerase III, which transcribes tRNA (Kirchner *et al.*, 1995). The general model of nests of retrotransposons built up over evolutionary time either through insertional preference or post-integration selection appears, however, to concern mainly the families of prevalent elements, present in thousands of copies. The *Tnt1* element is present in several hundred copies in the tobacco genome, and was originally isolated because of its ability to be active in tissue culture and mutate genes (Grandbastien *et al.*, 1989). The *Tos17* element is present in only one to four copies in the rice genome, but it is sufficiently activated by tissue culture where it is highly mutagenic to make it an excellent tool for gene tagging

(Hirochika *et al.*, 1996; Yamazaki *et al.*, 2001). Recent data for 20,000 integration sites show that *Tos17* inserts three time more frequently into genic regions than into intergenic regions, preferring as its insertion site the palindrome ANGTT – AACNT surrounding the target duplication, and that 76% of these motifs are found in genic regions in the rice genome (Miyao *et al.*, 2003). The *Tos17* element is also transiently activated by the crossing of wild rice (*Zamia latifolia*) with cultivated rice, leading to a stable increase in copy number in the succeeding introgression lines (Liu and Wendel, 2000).

Little or no information exists on the target site specificities of the non-LTR retrotransposons, the SINEs and LINES, in cereals. Intriguingly, SINE elements in Brassica show a preference for integrating into matrix attachment regions (MARs) although not strict sequence specificity (Tikhonov *et al.*, 2001). The authors propose two possible explanations: MARs are cleavage targets for the LINE endonucleases likely involved in SINE integration; insertion into MARs may lead to changes in chromatin organization, affecting gene expression and generating useful variation.

A general view has developed that abundant elements, which may be, as is *BARE*-1, active in many tissues of the plant (Suoniemi *e al.*, 1996a; Jääskeläinen *et al.*, 1999; Vicient *et al.*, 2000b), have been able to achieve high copy number because they generally do not disrupt genes, whereas other elements such as *Tos17* are under tight regulation and consequently rare because of their mutagenic potential. In this regard, the MITE elements, which are not retrotransposons but rather are derivatives of nonautonomous Type II transposons (Jiang *et al.*, 2003), offer a contrasting model. The MITE elements are also present in high copy numbers in the genome, but show a strong propensity for inserting in or near genes (Bureau et al., 1996; Casa *et al.*, 2000). One of the obvious differences between MITEs and retrotransposons is their size, 100 to 300 bp for MITEs and up to 10 kb for retrotransposons. It is tempting to speculate that MITEs are tolerated within genes because of their low disruptive potential (Walbot and Petrov, 2001).

Both MITEs and prevalent retrotransposons have been shown, however, to affect gene activity (Wessler *et al.*, 1995). The simplest cases are loss of function mutations caused by integration into individual genes (*e.g.*, Weil and Wessler, 1990; Varagona *et al.*, 1992). Larger-scale changes in gene expression may be possible through SINE insertions at MARs (Tikhonov *et al.*, 2001). Transposable elements may also insert into or near promoters, contributing novel regulatory functions (Bureau *et al.*, 1994; White *et al.*, 1994; Kloeckener-Gruissem and Freeling, 1995). A spectacular example of global effects on gene expression is the transcriptional activation of the *Wis*-2 retrotransposons in wheat following a wide cross, and subsequent silencing

of hundreds of nearby genes by chimeric transcripts originating from the LTRs (Kashkush *et al.*, 2003). If phenotypic variability provides a basis for selection, then retrotransposons that integrate into genes and affect their expression, rather than those that insertions solely into silent seas of repetitive DNA, will at times have a selective advantage.

2.6. Retrotransposon-Based Molecular Markers: A Practical Use of Genome Organization

Retrotransposons, as major agents of genome change and also as major, dispersed components of the genome, are appealing candidates for the development of molecular marker systems designed to track such changes. The transposition of retrotransposons is not linked, as it is with DNA transposons, with removal of the mother element from its locus; although LTR-LTR recombination may delete an element, the remaining LTR leaves many such losses transparent for marker systems based on LTRs. The greater unidirectionality of retrotransposon integration compared to that of point mutations, or to microsatellite expansion and contraction, confers great advantages in reconstructing pedigrees and phylogenies. Furthermore, the ancestral state of a retrotransposon insertion is obvious – it is the empty site, whereas for most genetic polymorphisms on which markers are built it cannot be inferred. In this way, SINE elements have been used to trace human roots to Africa (Batzer *et al.*, 1994), to establish the relationship of whales to even-toed ungulates (Shimamura *et al.*, 1997), and to infer the evolutionary relationships between wild rice species (Cheng *et al.*, 2002).

Most retrotransposon-based methods employ PCR primed on conserved motifs in the element and on some other widespread and conserved motif in the surrounding DNA. Waugh and colleagues (1997) exploited the dispersion and prevalence of *BARE*-1 in barley through modification of the AFLP (amplified fragment length polymorphism) technique. In their approach, coined S-SAP (sequence-specific amplified polymorphism), a primer anchored in the LTR replaces one of the adapter primers of AFLP. As discussed above, although retrotransposons are dispersed, they are also clustered in the genome. It is this phenomenon that makes possible the IRAP (inter-retrotransposon amplified polymorphism) method, in which amplifications are carried out between primers for two retroelements in the genome (Kalendar *et. al.*, 1999). If *BARE*-1 elements were fully dispersed in the barley genome, amplification templates would be a minimum of 50 kb long. Likewise, the proximity of microsatellites to retrotransposons in cereal genomes (see below) makes the REMAP method (Kalendar *et al.*, 1999)

function. In addition to S-SAP, IRAP, and REMAP, a fourth based method based on the polymorphic integration pattern of retrotransposons, RBIP (retrotransposon-based insertional polymorphism), has been developed (Flavell *et al.*, 1998). The RBIP method uses primers flanking retrotransposon insertions and scores the presence or absence of insertions at individual sites. Over the last five years, retrotransposon-based marker systems have proven their utility in phylogenetic, genetic diversity, breeding, and mapping projects not only in cereals including barley, wheat, rye, and oat, but also in other crop plants and tree species (*e.g.*, Ellis *et al.*, 1998; Gribbon *et al.*, 1999; Manninen *et al.*, 2000; Yu and Wise, 2000; Boyko *et al.*, 2000; Porceddu *et al.*, 2002; Leigh *et al.*, 2003; also see the chapter no. 3 of this volume).

3. ORGANIZATION OF MICROSATELLITES

Microsatellites are tandem repeats of DNA sequences, each repeat only a few base pairs (1-6 bp) long (Tautz and Renz, 1984). These are more popularly described as simple sequence repeats (SSRs) in plant systems (Morgante and Oliveri, 1993) and as short tandem repeats (STRs) in animal systems (Edwards *et al.,* 1991). These motifs consist of a single base pair or a small number of bases (usually ranging from 1 to 6) which are repeated several times and known as di-, tri-, tetra-nucleotide repeats, etc., accordingly. Microsatellites are abundant and occur frequently and randomly in all eukaryotic DNAs examined so far (see Gupta *et al.*, 1996; Gupta and Varshney, 2000). In the past, on an average, microsatellites in plant genomes were shown to be 10 fold less frequent than in the human genome (Powell *et al.*, 1996), although more recently, the frequencies were shown to be comparable in plants and animal systems. For instance, according to an earlier estimate on the basis of database search at that time, on an average, in monocots one SSR was observed every 64 kb, while in dicots one SSR occurred every 21kb (Wang *et al.*, 1984). However, according to recent reports based on searching of large genomic and EST (expressed sequence tags) sequences in different species, the frequencies of SSRs in plant genomes were observed to be much higher (1 SSR every 6-7 kb) than those reported earlier (Cardle *et al.*, 2000; Varshney *et al.*, 2002; Morgante *et al.*, 2002), and were comparable to those described for mammals (Beckmann and Weber, 1992). Also, in the rice genome sequences released as rough draft by Syngenta, one SSR (only di-, tri- and tetra-nucleotide repeats) was found to occur every 8 kb (Goff *et al.*, 2002).

Variation was also observed in the frequencies of individual microsatellite motifs among different organisms (Lagercrantz *et al.*, 1993; Morgante and Oliveri, 1993; Wang *et al.*, 1994; Gupta *et al.*, 1996). For instance, $(CA)_n$ motif is one of the most frequently occurring microsatellite in humans and several other mammals, but is comparatively less frequent in plants (Lagercrantz *et al.*, 1993; Morgante *et al.*, 2002). In plants, however, $(AT)_n$ microsatellite is the most abundant and the $(GA)_n$ is relatively more abundant than $(CA)_n$ repeats (Wang *et al.* 1994; Gupta *et al.*, 1996; Morgante *et al.*, 2002).

3.1. Densities/ Frequencies and Organization of Microsatellites in the Whole Genome

During the last decade and the early years of the present decade, microsatellite markers were developed in a large number of plant systems including major cereal species such as barley (Ramsay *et al.*, 2000; Thiel *et al.*, 2003), maize (Chin, 1996; Yu *et al.*, 2001), oats (Li *et al.*, 2000), rice (Akagi *et al.*, 1996; Temnykh *et al.*, 2000, 2001; McCouch *et al.*, 2002; Gao *et al.*, 2003), rye (Saal and Wricke, 1999), sorghum (Bhattramakki *et al.*, 2000) and wheat (Roder *et al.*, 1998b; Varshney *et al.*, 2000; Gupta *et al.*, 2002). In the majority of these studies involving several plant species, the two most common SSRs, whose densities in the genome were determined while screening genomic libraries for SSRs, included GA/CT and GT/CA (Table 1). The densities of $(AT)_n$ or $(GC)_n$ motifs could not be worked out, because of the difficulty in hybridization, due to self annealing. The densities of GA/CT and GT/CA in different plant species, determined as above, ranged from one SSR every 212 kb to 704 kb. These results differ from those for humans, with an estimated average density of one SSR every 6 kb (Beckmann and Weber, 1992). Estimates of the total number of SSRs at the genome level have also been made in several crops. For instance, the frequencies per haploid genome were found to be 3.6×10^4 $(GA)_n$ and 2.3×10^4 $(GT)_n$ in bread wheat (Roder *et al.*, 1995), and were estimated to be 1.36×10^3 $(GA)_n$ and 1.23×10^3 $(GT)_n$ in rice (Panaud *et al.*, 1995).

The organization of microsatellites in plant genomes has been studied in the past using both hybridization-based and PCR-based approaches. In hybridization-based approaches, synthetic oligonucleotides are used as probes either for hybridizing the gels or filters containing genomic DNA digested with a restriction enzyme, or for *in situ* hybridization of chromosomes. In PCR approaches, on the other hand, the primers flanking the microsatellites are designed and used in mapping/tagging experiments.

Table 1. Density of microsatellites in different portions of cereal genomes

Crop and source of SSRs	Density (kb of DNA per SSR)	Reference
Barley		
Genomic DNA	7.4	Cardle *et al.* (2000)
ESTs	7.5	Varshney *et al.* (2002)
	3.4	Kantety *et al.* (2002)
Maize		
Genomic DNA	4.5/5.71	Morgante *et al.* (2002)*
ESTs	8.1	Cardle *et al.* (2000)
	1.63/2.12	Morgante *et al.* (2002)*
	1.5	Kantety *et al.* (2002)
	7.5	Varshney *et al.* (2002)
	28.32	Gao *et al.* (2003)
Rice		
Genomic DNA	225-240	Wu and Tanksley (1993)§
	330-365	Panaud *et al.* (1995)¶
	7.4	Cardle *et al.* (2000)
	16/1.9	Temnykh *et al.* (2001)
	2.64/3.52	Morgante *et al.* (2002)*
BAC end sequences	40/3.7	Temnykh *et al.* (2001)
ESTs	3.4	Cardle *et al.* (2000)
	19	Temnykh *et al.* (2001)
	0.86/1.06	Morgante *et al.* (2002)*
	3.9	Varshney *et al.* (2002)
	4.7	Kantety *et al.* (2002)
	11.81	Gao *et al.* (2003)
Rye		
ESTs	5.5	Varshney *et al.* (2002)
Sorghum		
ESTs	5.5	Varshney *et al.* (2002)
	3.6	Kantety *et al.* (2002)

Table 1. Continued

Wheat		
Genomic DNA	440-704	Roder *et al.* (1995)[§]
	212-292	Ma *et al.* (1996)[§]
	3.35/5.16	Morgante *et al.* (2002)[*]
ESTs	1.33/1.67	Morgante *et al.* (2002)[*]
	6.2	Varshney *et al.* (2002)
	3.2	Kantety *et al.* (2002)
	17.2	Gao *et al.* (2003)
	9.2	Gupta *et al.* (2003)

[*]analysed frequency of imperfect and perfect SSRs separately. In each case frequency of imperfect SSRs is given first and followed by that of perfect SSRs; [§]studied the frequency of two DNR SSRs (GA/CT, GT/CA); [¶]classified SSRs in two categories- Class I SSRs >20 bp and ClassII SSRs >12 bp <20 bp. In each case the frequency of class I SSR is followed by that of class II.

Based on studies using the above approaches, conclusions have been drawn about the organization of microsatellites within a genome, although one should recognize the limitation of the hybridization approaches, which generally detect only high-density fragments representing the repetitive DNA.

3.1.1. *In-Gel* Hybridization (Oligonucleotide Fingerprinting)

The synthetic oligonucleotide probes complementary to SSR motifs have been successfully utilized for in-gel hybridization (and sometimes for Southern hybridization) with genomic DNA that was digested with individual restriction enzymes and electrophoresed on agarose gels (Ali *et al.*, 1986). The fragments that hybridize with synthetic oligonucleotides through in-gel hybridization are generally many, and the size of the hybridizing fragments range from a few hundred base pairs to more than 20kb, thus making the technique suitable for DNA fingerprinting (Beyermann *et al.*, 1992; Schmidt and Heslop-Harrison, 1996; Arens *et al.*,1995; Weising *et al.*, 1995, 1998). In crops like wheat, however, multilocus fingerprints due to SSR probes that are characteristic of this technique were not obtained during in-gel hybridization. Instead, a prominent solitary high molecular weight fragment (>23 kb), sometimes associated with a few low molecular weight bands, was obtained with a number of SSR probes (Varshney *et al.*, 1998). Similar type of high high molecular weight fragments (>23 kb) were observed in 14 individual species of *Triticum-Aegilops* group after in-gel hybridization (Sharma *et al.*, 2002). The presence of high molecular weight fragments measuring up to 30 kb and containing

$(GATA)_n$ and $(GTG)_n$ stretches were also earlier reported in barley (Beyermann *et al.,* 1992). In sugar-beet, also, fragments of >21.2 kb were obtained after hybridization with $(CA)_8$ probes; suggesting that tandem repeats harboring $(CA)_n$ occur in genomic regions, and that the organization of $(CA)_n$ differs from all the other microsatellites used in this study (Schmidt and Heslop-Harrison, 1996). It is also known that stretches carrying $(CA)_n$ form the subrepeats of a centromeric satellite (Schmidt and Metzlaff, 1991). In our opinion, the presence of high molecular weight bands (>20 kb) hybridizing with SSRs indicates the occurrence of microsatellites either within or in close association with the long tandem repeat units, *e.g.* retrotransposons, etc. Since the technique of in-gel hybridization is suitable for detection of only those sequences, which are relatively long and repetitive, the fragments that hybridized with synthetic oligonucleotides during in-gel hybridization may actually represent only those DNA fragments associated with SSRs, which are generally repetitive rather than representing both the repetitive and the unique sequences (Varshney *et al.*, 1998).

3.1.2. Association of Microsatellites with Retrotransposons

SSRs have been shown to be associated with 'short interspersed elements' (SINEs) in rice (Motohashi *et al.*, 1997). In barley also, association of microsatellites was observed with retrotransposons and other dispersed repetitive elements like *BARE*-1*, WIS2-1A, R-173*, *Pgr-1/PREM-1* (Ramsay *et al.*, 1999). In another study in rice, about 45% of $(AT)_n$ SSRs (harbored in BAC-ends) showed significant homology to 'Micropan' sequences, a new family of 'miniature inverted-repeat transposable elements' (MITEs) (Temnykh *et al.*, 2001). On the basis of the relative position of SSR and retrotransposons in barley, Ramsay *et al.* (1999) has drawn inferences about the origin of this association. They postulated that while some SSRs (proto-SSR; A- rich sequence) might have acted as 'landing pads' for insertion of transposable elements, there may be other SSRs that must have evolved as components of active transposable elements that are spread throughout the genome. It is also possible that, in some cases, expansion of SSRs in retroelements might have led to multiple SSR loci, which may also account for the difficulty in locus specific amplification, while using locus specific STMS primers. This may be true of the genomes like that of wheat, which contains a higher proportion of repetitive DNA (Varshney, 2001). The association between SSRs and mobile elements, as above, has also facilitated the development of the novel marker system REMAP (Kalender *et al.*, 1999; see section 2.6 earlier).

3.1.3. *In Situ* Hybridization with SSR Probes

In situ hybridization (ISH) was initially used to assess the chromosomal localization of various SSRs in humans and animals (Pardue *et al.*, 1987; Lohe *et al.*, 1993). This gave some useful information regarding physical organization of microsatellites in plant chromosomes, including those from cereals. For instance, a $(GAA)_7$ probe, when hybridized to barley chromosomes, gave a pattern conforming to the distribution of heterochromatin and the C-banding pattern (Pederson and Linde-Laursen, 1994). This distribution of the so-called (GAA) satellite in heterochromatic regions was later confirmed in some other cereal species also (Pederson *et al.*, 1996). In bread wheat, rye, and hexaploid triticale, ten different SSRs gave dispersed hybridization signals of varying strength on all chromosomes when used for ISH (Cuadrado and Schwaracher 1998). However, in wheat, microsatellite motifs $(AG)_{12}$, $(CAT)_5$, $(AAG)_5$, $(GCC)_5$, and in particular, $(GACA)_4$ hybridized strongly to pericentromeric and multiple intercalary sites on the B genome chromosomes and on chromosome 4A, resembling the N-banding pattern. In contrast to this, in rye, $(GACA)_4$ gave strong hybridization signals at many intercalary sites in all the chromosomes, which largely differed from the known banding pattern. Earlier in sugar beet also, when CA, GA, TA, CAC, GATA, GACA and GGAT repeats were used for ISH, it was noticed that each microsatellite had a characteristic genomic distribution and motif-dependent dispersion with site-specific enrichment or depletion of some motifs at centromeric or intercalary positions (Schmidt and Heslop-Harrison, 1996). From several ISH studies, as above, it has been inferred that the tandemly repeated sequences are located around all centromeres in blocks, more than 80 kb long, and that their homology in the microsatellite domain perhaps is responsible for some of the hybridization signals. Thus, ISH studies support the hypothesis of association of at least some SSRs with highly repetitive DNA that may sometimes involve retrotransposons.

3.1.4. Genome Mapping and Physical Mapping using STMS Primer Pairs

The sequences flanking specific microsatellite loci in the genome are believed to be conserved and, therefore, have been used for designing primers to amplify individual microsatellite loci: the technique was described as sequence tagged microsatellite site (STMS) analysis (Beckmann and Soller, 1990). STMS primers have come to be used extensively for genome mapping in a number of plant species including wheat, barley, rice, maize, etc. (for references see Gupta and Varshney, 2000). In almost all plant

species, the genome mapping suggested that microsatellites are dispersed throughout the genome and not clustered (Gupta and Varshney, 2000) except some reports in barley (Ramsay *et al.*, 2000; Li *et al.,* 2003), where centromeric clustering of genomic SSRs was observed. No clustering in barley genome was, however, observed, when EST-derived microsatellite loci were mapped later (Thiel *et al.*, 2003, Varshney *et al*., unpublished). In rice, microsatellite markers with different SSR motifs, regardless of whether they belonged to genomic DNA or cDNA, were also found to be rather uniformly distributed along all the rice chromosomes (Panaud *et al.*, 1996; Temynkh *et al.*, 2000; McCouch *et al.*, 2002). Similarly, in bread wheat, physical mapping of microsatellite markers on chromosomes of homoeologous group 2 showed an absence of microsatellite clustering (Roder *et al.*, 1998a).

3.1.5. Frequencies of Microsatellites in Whole Rice Genome

The distribution of SSRs in the rice genome has also been studied on the basis of the two whole genome draft sequences released, respectively, by Syngenta and by the Beijing Genome Institute (BGI). In the draft sequence released by Syngenta (Goff *et al.*, 2002), for instance, 48,351 SSRs (including di-, tri- and tetra-nucleotide repeats) were available, giving a density of 8 kb per SSR in the whole genome; SSRs represented by di-, tri-, and tetra-nucleotide repeats accounted respectively for 24%, 59% and 17% of the total SSRs. The most frequent dinucleotide repeats (DNRs) were AG/CT repeats, which accounted for 58% of all DNRs, and the most frequent trinucleotide repeats (TNRs) were CGG/CCG repeats, which accounted for 44% of all TNRs. In the predicted genes, among the 7000 SSRs that were available, SSRs were mainly TNRs (92%). This abundance of TNRs may be attributed to a lack of selection against length variation in these SSRs, since it will not cause any frameshift mutations. SSRs were also studied in the rough draft of rice genome sequences released by BGI (Yu *et al.*, 2002), and accounted for 1.7% of the genome, as against 3% of the human genome represented by SSRs. Interestingly, majority of rice SSRs were mononucleotides, primarily $(A)_n$ or $(T)_n$.

3.2. Density and Distribution of Microsatellites based on *in silico* Mining

During the last five years, in the genomics era, large-scale genome/EST sequencing projects were initiated in several plant species including cereals.

The data generated from these projects was utilized for studying the frequency, distribution and organization of microsatellites in the expressed portion of the genome, and in some cases also in the whole genome (Table 1). For development of EST-SSRs, ESTs have been scanned in different plant species, including cereals such as rice (Temnykh *et al.*, 2000, 2001), barley (Kota *et al.,* 2001; Thiel *et al.*, 2003), wheat (Eujayl *et al.*, 2002; Gao *et al.*, 2003; Gupta *et al.*, 2003), and rye (Hackauf and Wehling, 2002). These efforts also allowed estimation of the density of SSRs in expressed regions of the genomes. For instance, on the basis of a number of contiguous genomic sequences, the density of SSRs was found to be 7.4 kb per SSR in barley and 7.4 kb per SSR in rice. Similarly, on the basis of EST sequences, the density of SSRs was 3.4 kb per SSR in rice and 8.1 kb per SSR in maize (Cardle *et al.*, 2000). Another study was conducted, which involved survey of EST sequences amounting to 75.2 Mb in barley, 54.7 Mb in maize, 43.9 Mb in rice, 3.7 Mb in rye, 41.6 Mb in sorghum and 37.5 Mb in wheat; the overall average density of SSR in these species was found to be 6.0 kb per SSR (Varshney *et al.*, 2002). However in another study, the frequency of SSR was one every 11.81 kb in rice, 17.42 kb in wheat and 28.32 kb in maize (Gao *et al.*, 2003). Difference in the frequency of SSRs in the ESTs of a particular species in different studies may be attributed to criteria of SSR search and data quantity used to identify SSRs in the database mining approaches.

In the comprehensive study of Varshney et al. (2002), almost in every cereal species, the TNRs were the most frequent (54% to 78%) followed by the DNRs (17.1% to 40.4%). As mentioned earlier, the abundance of trimeric SSRs was attributed to the absence of frameshift mutations due to length variation in these SSRs (Metzgar *et al.*, 2000). Among the trinucleotide repeats also, codon repeats corresponding to small hydrophilic amino acids are perhaps easily tolerated, but strong selection pressures probably eliminate codon repeats encoding hydrophobic and basic amino acids. Such an inference was drawn from an analysis of coding DNA sequences in the whole genomes of fruitfly, the nematode *C. elegans* and the budding yeast (Katti *et al.*, 2001). Furthermore, in various cereal genomes, among the DNRs, the motif AG is the most frequent (38% to 59%) followed by the motif AC (20% to 34%) in all the species except rye, where these frequencies are 50% for AC and 37.9% for AG (Varshney *et al.*, 2002). The most infrequent motif is CG in all species (1.7% to 9.0%) except in barley, where AT is the least frequent (8.4%). Among the TNRs, the motif CCG is the most frequent, ranging from 32% in wheat, to 49% in sorghum followed by AGC (13% to 30%) in barley, maize, rice and sorghum, and AAC in wheat (27%) and rye (16%). The third most frequent motif is AGG in barley, rice, rye, sorghum, AGC in wheat, and AAC in maize.

In another study of SSRs in rice genome, a total of 57.8 Mb DNA sequences were used, which included the following: (i) 12,532 ESTs (6.3 Mb), (ii) a large set (74,127) of short BAC-end sequences (500 bp on average), and (iii) 27 fully sequenced large-insert (BAC/PAC) clones (150 kb on average) (Temynkh *et al.* 2000, 2001). In this study, SSRs were classified in two categories: class I SSRs, each with ≥20 bp, and class II SSRs, each with ≥12 bp to <20 bp. The density of class I SSRs was one SSR every 40 kb in BAC-end sequences, one SSR every 16 kb in fully sequenced BAC/PAC clones and one SSR every 19 kb in ESTs. Class II SSRs, similarly, occurred every 3.7 kb in BAC-ends and every 1.9 kb in BAC/PAC clones (Table 1). The proportion of GC-rich TNRs among the total SSRs also differed in each of the three classes of DNA sequences used, from a low of 10.5% of all SSRs in BAC-ends to a high of 59% of all SSRs in ESTs with intermediate frequencies (~27%) in fully sequenced BAC and PAC clones. The densities of $(AT)_n$ DNRs were in the reverse order to that of GC-rich TNR e.g. 38.2% in BAC ends, 27% in BAC/ PAC clones, and only 2.9% in ESTs. The pattern of the frequencies of tetra-nucleotide SSRs and (CA)n DNRs were similar to that of (AT)n.

The above pattern of variation demonstrated that different SSR motifs are not randomly distributed in the rice genome. Since the frequencies of GC-rich TNRs and (AT)n DNRs were identical in BAC/PAC clones and varied greatly and inversely in BAC-ends and ESTs, we may conclude that the BAC-ends and the ESTs represented different genomic domains, neither of which would have the SSR composition representative of the whole genome. However, the frequencies of different SSRs in the fully sequenced BACs/PACs perhaps represent more faithfully the frequencies of microsatellites in the whole genome. As a result, it can be concluded that regions of relatively high gene-density (e.g. BAC/PAC clones) have prevalence of GC-rich TNRs (known to be associated with genes) and the regions of lower gene density (e.g. BAC ends) has prevalence of (AT)n, (AC)n and tetra-nucleotide SSRs (which are abundant in non-coding, intergenic regions).

3.3. Comparative Distribution of Microsatellites in Transcribed and Non-transcribed Portions of the Genome

Recently, the abundance and relative distribution of microsatellites among transcribed and non-transcribed regions have been more critically assessed in some plant species including rice, maize and wheat, after examining a large data set of genomic sequences and ESTs (Morgante *et al.*, 2002). In general,

the frequency of microsatellites was significantly higher in ESTs than in genomic DNA across all species (Table 1). In the maize genome as well, the frequency of microsatellites in the non-repetitive fraction was actually found to be significantly higher (more than double) than that in the repetitive fraction. Furthermore, the average density of microsatellites in rice BAC clones that were representative of gene rich regions was almost double that of the reference rice genomic DNA, which again supports the hypothesis of higher frequencies of microsatellites in the non-repetitive regions. A highly significant and positive linear relationship was also observed between frequencies of microsatellites (including both, the perfect and the imperfect SSRs) and the proportion of single copy DNA in a genome. Thus microsatellite frequency is a function of not only the overall genome size, but also of the relative proportion of single-copy DNA.

The above results are in sharp contrast to the earlier observations, which assumed an occurrence of SSRs mainly in the repetitive DNA, and therefore, were used to suggest preferential origin of microsatellites from repetitive DNA in both animals (Arcot *et al.*, 1995; Nadir *et al.*, 1996) and plants (Ramsay *et al.*, 1999). However, these earlier results can now be explained on the basis of the limitation of the techniques, which perhaps failed to score SSRs in the unique sequences. Frequencies of individual SSRs also varied both within and between genomic and transcribed sequences. For instance, in ESTs there was a higher frequency of AG/CT repeats and lower frequency of AT repeats. Similarly, TNRs were significantly more abundant in ESTs and the CCG/CGG repeat motifs alone accounted for half of the TNRs in rice ESTs. In contrast to this, they were rare in dicots (e.g. Arabidopsis and soybean) and moderately abundant in monocots other than rice (e.g. maize and wheat). This difference was attributed to higher G+C content and consequent usage bias in monocot ESTs. Overall, on an average, microsatellites were also shorter in the transcribed regions.

On the basis of the *in silico* analysis of a large data set in various species and the rough drafts of rice genome that were released, the SSR frequency in plant, or at least in cereal, genomes is in the range of one SSR every 2- 10 kb, which is comparable with those for humans and mammals (Beckmann and Weber, 1992). Further, it can be concluded clearly that TNRs comprise the highest proportion of the total SSRs followed by the DNRs. The most frequent motifs are CCG, AGC, AAC among TNRs, and AG and AC among the DNRs. However, there are discrepancies in the frequencies of SSRs in a particular class within a given species (Table 1). These discrepancies may be explained by the varying search criteria for SSRs or by the different bias to 5' or 3' regions in the ESTs or by varying dataset, used by different groups. Among different species, the overall frequency of microsatellites is inversely

related to genome size and to the proportion of repetitive DNA but remained constant in the transcribed portion of the genome. This indicates that the distribution of microsatellites is a function of the dynamics and history of genome evolution and of selective constraints, because these microsatellites resided in different regions of the genome of a species pre-dating the recent genome expansions in the plants (Morgante *et al.*, 2002).

4. SUMMARY AND OUTLOOK

The wide variations in genome size, independent of biological complexity, phylogenetic relationship, or ploidy level, that were described as the *C*-value paradox in the early 1970s are now know to be due to the variations in abundance of the retrotransposons, DNA transposons, simple sequence repeats, and other repeats that constitute the greater portion of most plant genomes. The genome, rather than being a static association of genes recombining with a frequency based on their physical distance and accumulating point mutations, is a complex and dynamic landscape in which these various elements create both local and large-scale structures that change over time. The presence in cereals of simple sequence repeats in and near retrotransposons and genes, and the insertion of retrotransposon into the blocks of repetitive elements present in cereal centromeres, represent just a part of the landscape that genome sequencing is now revealing.

The retrotransposons, far from being the "junk DNA" they were initially in ignorance referred to as, function as self-replicating genomic viruses. They are able to effect vast changes in genome size over evolutionary time as well as alter gene expression patterns and gene products in single generations. Ultimately, the balance between gain and loss of retrotransposons, within the context of their integrational preferences and selective forces, shapes the genome. LTR-LTR recombination, gene conversion mechanisms, and the effects of repeated small deletions can offset gain through integration. At individual loci, reconstruction by means of the direct repeats generated by integrase reveals cycles of gain by integration and loss through LTR-LTR recombination (Shirasu *et al.*, 2000). Hot, dry conditions appear to favor gain but decrease loss of retrotransposons (Kalendar *et al.*, 2000). Loss of retrotransposons may in itself provide a purifying selection for functionality both of coding domains (Suoniemi *et al.*, 1998; Navarro-Quezada and Schoen, 2002) and of the promoter.

For the retrotransposons, successive cycles of nested integration, LTR-LTR recombination, and accumulation of point mutations and small deletions ultimately obliterates the historical record of events in the genome, limiting

our historical reach. The bidirectionality and rapidity with which microsatellites can change in size likewise obscures their origins. Although a small genome with few retrotransposons is generally seen as the ancestral state, proving this is difficult. Nevertheless, over the time scale represented by agriculture, 10 to 20 thousand years, retrotransposon insertions and microsatellite allelic variations are highly useful as molecular markers in a wide variety of applications. Mapping, phylogeny reconstruction, genetic diversity studies, pedigree analysis, and targeted breeding is proving the practical value of the "junk" and has reunited it with genetics.

5. ACKNOWLEDGEMENTS

We thank Dr Robbie Waugh, Scottish Crop Research Instititute, Invergowrie, Scotland for reviewing the MS and helpful comments.

6. REFERENCES

Abrahamson S., Bender M.A., Conger A.D., Wolff S. Uniformity of radiation-induced mutation rates among different species. Nature 1973; 245: 460-462

Akagi H., Yokozeki Y., Inagaki A., Fujimura T. Microsatellite DNA markers for rice chromosomes. Theor Appl Genet 1996, 93:1071-1077

Ali S., Müller C.R., Epplen J.T. DNA fingerprinting by oligonucleotide probes specific for simple repeats. Hum Genet 1986; 74:239-243

Ananiev E.V., Phillips R.L., Rines H.W. Chromosome-specific molecular organization of maize (*Zea mays* L.) centromeric regions. Proc Natl Acad Sci USA 1998a; 95: 13073-13078

Ananiev E.V., Phillips R.L., Rines H.W. Complex structure of knob DNA on maize chromosome 9. Retrotransposon invasion into heterochromatin. Genetics 1998b; 149: 2025-2037

Arcot S.S., Wang Z.Y., Weber J.L., Deininger P.L., Batzer M.A. Alu repeats- a source for the genesis of primate microsatellites. Genomics 1995; 29: 136-144

Arens, P., Odinot P., van Heusden A.W., Lindhout P., Vosman B. GATA and GACA repeats are not evenly distributed throughout the tomato genome. Genome 1995; 38:84-90

Arnaud P., Goubely C., Pélissier T., Deragon J.-M. SINE retroposons can be used in vivo as nucleation centers for de novo methylation. Mol Cell Biol 2000; 20: 3434-3441

Baltimore D. RNA-dependent DNA polymerase in virons of RNA tumor viruses. Nature 1970; 226: 1209-1211

Barakat A., Carels N., Bernardi G. The distribution of genes in the genomes of Gramineae. Proc Natl Acad Sci USA 1997; 94: 6857-6861

Batzer M.A., Stoneking M., Alegria-Hartman M., Bazan H., Kass D.H., Shaikh T.H., Novick G.E., Ioannou P.A., Scheer W.D., Herrera R.J., *et al*. African origin of human-specific polymorphic Alu insertions. Proc Natl Acad Sci USA 1994; 91: 12288-12292

Beckmann, J.S., Soller M. Towards a unified approach to genetic mapping of eukaryotes based on sequence tagged microsatellite sites**.** Bio/Technology 1990; 8: 930-932

Beckmann J.S. Weber J.L. Survey of human and rat microsatellites. Genomics 1992; 12: 627-631

Bennetzen J.L., Freeling M. The unified grass genome: synergy in synteny. Genome Res 1997; 7: 301-306

Bennetzen J.L., Kellogg E.A. Do plants have a one-way ticket to genome obesity? Plant Cell 1997a; 9: 1509-1514

Bennetzen J.L., Kellogg E.A. Reply. Plant Cell 1997b; 9: 1901-1902

Bennetzen J.L., Schrick K., Springer P.S., Brown W.E., Sanmiguel P. Active maize genes are unmodified and flanked by diverse classes of modified, highly repetitive DNA. Genome 1994; 37: 565-576

Beguiristain T., Grandbastien M.A., Puigdomènech P., Casacuberta J.M. Three Tnt1 subfamilies show different stress-associated patterns of expression in tobacco. Consequences for retrotransposon control and evolution in plants. Plant Physiol 2001; 127: 212-221

Belyayev A., Raskina O., Nevo E. Chromosomal distribution of reverse transcriptase-containing retroelements in two Triticeae species. Chrom Res 2001; 9: 129-136

Beyermann B., Nurnberg P., Weihe A., Mexiner M., Epplen J.T., Borner T. Fingerprinting plant genomes with oligonucleotide probes specific for simple repetitive sequences. Theor Appl Genet 1992; 83: 691-694

Bhattramakki D., Dong J.M., Chhabra A.K., Hart G.E. An integrated SSR and RFLP linkage map of *Sorghum bicolor* (L.) Moench. Genome 2000; 43: 988-1002

Boeke J.D. LINEs and Alus -- the polyA connection. Nat Genet 1997; 16: 6-7

Boeke J.D., Garfinkel D.J., Styles C.A., Fink G.R. *Ty* elements transpose through an RNA Intermediate. Cell 1985; 40: 491-500

Boeke J.D., Devine S.E. Yeast retrotransposons: finding a nice quiet neighborhood. Cell 1998; 93: 1087-1089

Boyko E., Kalendar R., Korzun V., Gill B., Schulman A.H. Combined mapping of *Aegilops tauschii* by retrotransposon, microsatellite, and gene markers. Plant Mol Biol 2002; 48: 767-790

Bureau T.E., Wessler S.R. *Stowaway* : A new family of inverted repeat elements associated with the genes of both monocotyledonous and dicotyledonous plants. Plant Cell 1994; 6: 907-916

Bureau T.E., Ronald P.C., Wessler S.R. A computer-based systematic survey reveals the predominance of small inverted-repeat elements in wild-type rice genes. Proc Natl Acad Sci USA 1996; 93: 8524-8529

Cardle L., Ramsay L., Milbourne D., Macaulay M., Marshall D., Waugh R. Computational and experimental characterization of physically clustered simple sequence repeats in plants. Genetics 2000; 156: 847-854

Casa A., Brouwer C., Nagel A., Wang L., Zhang Q.-., Kresovich S., Wessler S.R. The MITE family *Heartbreaker* (*Hbr*): molecular markers in maize. Proc Natl Acad Sci USA 2000; 97: 10083-10089

Casacuberta E., Pardue M.L. Transposon telomeres are widely distributed in the Drosophila genus: TART elements in the virilus group. Proc Natl Acad Sci USA 2003; 100: 3363-3368

Chen M., SanMiguel P., de Oliviera A.C., Woo S.-S., Zhang H., Wing R.A., Bennetzen J.L. Microcolinearity in *sh2*-homologous regions of the maize, rice, and sorghum genomes. Proc Natl Acad Sci USA 1997; 94: 3431-3435

Chen M., SanMiguel P.J., Bennetzen J.L. Seguence organization and conservation in *sh2*/*a1* -homologous regions of sorghum and rice. Genetics 1998; 148: 435-443

Cheng C., Tsuchimoto S., Ohtsubo H., Ohtsubo E. Evolutionary relationships among rice species with AA genome based on SINE insertion analysis. Genes Genet Syst 2002; 77: 323-334

Chin E.C.L. Maize simple repetitive DNA sequences: abundance and allele variation. Genome 1996; 39:866-873

Cuadrado A., Schwarzacher T. The chromosomal organization of simple sequence repeats in wheat and rye genomes. Chromosoma 1998; 107:587-594

Deragon J.-M., Capy P. Impact of transposable elements on the human genome. Annals of Medicine 2000; 32: 264-273

Danilevskaya O.N., Lowenhaupt K., Pardue M.L. Conserved families of the *Drosophila* HeT-A telomere-specific retrotransposon. Genetics 1998; 148: 233-242

Dong F.G., Miller J.T., Jackson S.A., Song J., Jiang J. Rice (*Oryza sativa*) centromeric regions consist of complex DNA. Proc Natl Acad Sci USA 1998; 95: 8135-8140

Echenique V., Stamova B., Wolters P., Lazo G., Carollo L., Dubcovsky J. Frequencies of *Ty1-copia* and *Ty3-gypsy* retroelements within the Triticeae EST databases. Theor Appl Genet 2002; 104: 840-844

Edwards A., Civitello H., Hammond H.A. and Caskey C.T. DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. Am J Hum Genet 1991; 49: 746-756

Eickbush T.H. Transposing without ends: the non-LTR retrotransposable elements. New Biol 1992; 4: 430-440

Ellis T.H.N., Poyser S.J., Knox M.R., Vershinin A.V., Ambrose M.J. *Ty1-copia* class retrotransposon insertion site polymorphism for linkage and diversity analysis in pea. Mol Gen Genet 1998; 260: 9-19

Eujayl I., Sorrells M.E., Baum M., Wolters P., Powell W. Isolation of EST-derived microsatellite markers for genotyping the A and B genomes of wheat. Theor Appl Genet 2002; 104: 399-407

Feng Q., Zhang Y., Hao P., Wang S., Fu G., Huang Y., Li Y., Zhu J., Liu Y., Hu X., *et al.,* Sequence and analysis of rice chromosome 4. Nature 2002; 420: 316-320

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

Flavell A.J., Ruby S.W., Toole J.J., Roberts B.E., Rubin G.M. Translation and developmental regulation of RNA encoded by the eukaryotic transposable element *copia*. Proc Natl Acad Sci USA 1980; 77: 7107-7111

Flavell R.B., Bennett MD, Smith J.B., Smith D.B. Genome size and the proportion of repeated nucleotide sequence DNA in plants. Biochem Genet 1974; 12: 257-268

Flavell A.J., Knox M.R., Pearce S.R., Ellis T.H.N. Retrotransposon-based insertion polymorphisms (RBIP) for high throughput marker analysis. Plant Jour 1998; 16: 643-650

Fu H., Park W., Yan X., Zheng Z., Shen B., Dooner H.K. The highly recombinogenic *bz* locus lies in an unusually gene-rich region of the maize genome. Proc Natl Acad Sci USA 2001; 98: 8903-8908

Fukui K.-N., Suzuki A., Lagudah E.S., Rahman S., Appels R., Yamamoto M., Mukai Y. Physical arrangement of retrotransposon-related repeats in the centromeric regions of wheat. Plant Cell Physiol 2001; 42: 189-196

Gao L.F., Tang J.F., Li H.W., Jia J.Z. Analysis of microsatellites in major crops assessed by computational and experimental approaches. Mol Breed 2003; 12: 245-261

Gaut B.S., Le Thierry d'Ennequin M., Peek A.S., Sawkins M.C. Maize as a model for the evolution of plant nuclear genomes. Proc Natl Acad Sci USA 2000; 97: 7008-7015

Gilbert N., Arnaud P., Lenoir A., Warwick S.I., Picard G., Deragon J.-M. Plant S1 SINEs as a model to study retroposition. Genetica 1997; 100: 155-160

Goodman H.M., Ecker J.R., Dean C. The genome of *Arabidopsis thaliana*. Proc Natl Acad Sci USA 1995; 92: 10831-10835

Goff S.A., Ricke, D., Lan T.-H., Presting G., Wang R., Dunn M., Glazebrook J., Sessions A., Oeller P., Varma, H., Hadley, D. *et al.* A draft sequence of the rice genome (*Oryza sativa* L. sspl. *japonica*). Science 2002; 296:92-100

Grandbastien M.-A., Spielmann A., Caboche M. *Tnt1*, a mobile retroviral-like transposable element of tobacco isolated by plant cell genetics. Nature 1989; 337: 376-380

Grandbastien M.-A. Activation of retrotransposons under stress conditions. Trends Plant Sci 1998; 3: 181-187

Gribbon B.M., Pearce S.R., Kalendar R., Schulman A.H., Paulin L., Jack P., Kumar A., Flavell A.J. Phylogeny and transpositional activity of *Ty1-copia* group retrotransposons in cereal genomes. Mol Gen Genet 1999; 261: 883-891

Gupta P.K., Balyan H.S., Edwards K.J., Isaac P., Korzun V., Roder M., Gautier M.F., Joudrier P., Schlatter A.R., Dubcovsky J., *et al.* Genetic mapping of 66 new microsatellite (SSR) loci in bread wheat. Theor Appl Genet 2002; 105: 413-422

Gupta P.K., Balyan H.S., Sharma P.C, Ramesh B. Microsatellites in plants: a new class of molecular markers. Curr Sci 1996; 70:45-54.

Gupta P.K., Rustgi S., Sharma S., Singh R., Kumar N., Balyan H.S. Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. Mol Gen Genom 2003; 270: 315-323

Gupta P.K., Varshney R.K. The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica 2000; 113:163-165.

Hackauf B., Wehling P. Identification of microsatellite polymorphisms in an expressed portion of the rye genome. Plant Breed 2002; 121:17-25

Hake S., Walbot V. The genome of *Zea mays*, its organization and homology to related grasses. Chromosoma 1980; 79: 251-270

Heslop-Harrison J.S. Comparative genome organization in plants: From sequence and markers to chromatin and chromosomes. Plant Cell 2000; 12: 617-635

Hirochika H., Fukuchi A., Kikuchi F. Retrotransposon families in rice. Mol Gen. Genet. 1992; 233: 209-216

Hirochika H. Activation of tobacco retrotransposons during tissue culture. EMBO J 1993; 12: 2521-2528

Hirochika H., Sugimoto K., Otsuki Y., Tsugawa H., Kanda M. Retrotransposons of rice involved in mutations induced by tissue culture. Proc Natl Acad Sci USA 1996; 93: 7783-7788

Hudakova S., Michalek W., Presting G.G., ten Hoopen R., dos Santos K., Jasencakova Z., Schubert I. Sequence organization of barley centromeres. Nucl Acids Res 2001; 29: 5029-5035

Jacob H.J., Brown D.M, Bunker R.K., Daly M.J., Dzau V.J., Goodman A., Kolike G., Kren V., Kurtz T., Lernmark A. *et al.* A genetic linkage map of the laboratory rat, *Rattus norvegicus*. Nature Genet 1995; 9: 63-69

Jääskeläinen M., Mykkänen A.-H., Arna T., Vicient C., Suoniemi A., Kalendar R., Savilahti H., Schulman A.H. Retrotransposon *BARE*-1: Expression of encoded proteins and formation of virus-like particles in barley cells. Plant Jour 1999; 20: 413-422

Jiang N., Bao Z., Zhang X., Hirochika H., Eddy S.R., McCouch S.R., Wessler S.R. An active DNA transposon family in rice. Nature 2003; 421: 163-167

Kalender R., Grob T., Regina M., Suoniemi A., Schulman A. IRAP and REMAP: two new retrotransposon-based DNA fingerprinting techniques. Theor Appl Genet 1999; 98:704-711

Kalendar R., Tanskanen J., Immonen S., Nevo E., Schulman A.H. Genome evolution of wild barley (*Hordeum spontaneum*) by *BARE*-1 retrotransposon dynamics in response to sharp microclimatic divergence. Proc Natl Acad Sci USA 2000; 97: 6603-6607

Kantety R.V., La Rota M., Matthews D.E., Sorrells M.E. Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. Plant Mol Biol 48: 501-510

Kashkush K., Feldman M., Levy A.A. Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. Nat Genet 2003; 32: 102-106

Katsiotis A., Schmidt T., Heslop-Harrison J.S. Chromosomal and genomic organization of *Ty1-copia*-like retrotransposon sequences in the genus *Avena*. Genome 1996; 39: 410-417

Katti M.V., Ranjeker P.K., Gupta V. Differential distribution of simple sequence repeats in eukaryotic genome sequences. Mol Biol Evol 2001; 18: 1161-1167

Kazazian H.H., Jr., Goodier J.L. LINE drive: Retrotransposition and genome instability. Cell 2002; 110: 277-280

Kim J.M., Vanguri S., Boeke J.D., Gabriel A., Voytas D.F. Transposable elements and genome organization: A comprehensive survey of retrotransposons revealed by the complete *Saccharomyces cerevisiae* genome sequence. Genome Res 1998; 8: 464-478

Kingsman A.J., Kingsman S.M. *Ty*: a retroelement moving forward. Cell 1988; 53: 333-335

Kirchner J., Connolly C.M., Sandmeyer S.B. Requirement of RNA polymerase III transcription factors for in vitro position-specific integration of retroviruslike element. Science 1995; 267: 1488-1491

Kloeckner-Gruissem B., Freeling M. Transposon-induced promoter scrambling: a mechanism for the evolution of new alleles. Proc Natl Acad Sci USA 1995; 92: 1836-1840

Komatsu M., Shimamoto K., Kyozuka J. Two-step regulation and continuous retrotransposition of the rice LINE-type retrotransposon Karma. Plant Cell 2003; 15: 1934-1944

Kota R., Varshney R.K., Thiel T., Dehmer K.J., Graner A. Generation and comparison of EST-derived SSR and SNP markers in barley *(Hordeum vulgare* L.). Hereditas 2001; 135:145-151

Kumar A., Bennetzen J. Plant retrotransposons. Annu Rev Genet 1999; 33:479-532.

Kurata N., Nomomura K.-I., Harushima Y. Rice genome organization: the centromere and genome interactions. Ann Bot 2002; 90: 427-435

Kumekawa N., Ohtsubo E., Ohtsubo H. Identification and phylogenetic analysis of *gypsy*-type retrotransposons in the plant kingdom. Genes Genet Syst 1999; 74: 299-307

Kumekawa N., Hosouchi T., Tsuruoka H., Kotani H. The size and sequence organization of the centromeric region of *Arabidopsis thaliana* chromosome 5. DNA Res 2000; 7: 315-321

Kumekawa N., Ohmido N., Fukui K., Ohtsubo E., Ohtsubo H. A new *gypsy*-type retrotransposon, *RIRE7*: preferential insertion into the tandem repeat sequence TrsD in pericentromeric heterochromatin regions of rice chromosomes. Mol Genet Genomics 2001a; 265: 480-488

Kumekawa N., Hosouchi T., Tsuruoka H., Kotani H. The size and sequence organization of the centromeric region of *Arabidopsis thaliana* chromosome 4. DNA Res 2001b; 8: 285-290

Lagercrantz U., Ellegren H., Andersson L. The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. Nucleic Acids Res 1993; 21: 1111-1115.

Lamb J.C., Birchler J.A. The role of DNA sequence in centromere formation. Genome Biology 2003; 4: 214

Langdon T., Seago C., Mende M., Leggett M., Thomas H., Forster J.W., Jones R.L., Jenkins G. Retrotransposon evolution in diverse plant genomes. Genetics 2000; 156: 313-325

Laten H., Majumdar A., Gaucher E.A. SIRE-1, a copia/Ty1-like retroelement from soybean, encodes a retroviral envelope-like protein. Proc Natl Acad Sci USA 1998; 95: 6897-6902

Le Q.H., Wright S., Yu Z., Bureau T. Transposon diversity in *Arabidopsis thaliana*. Proc Natl Acad Sci USA 2000; 97: 7376-7381

Leigh F., Kalendar R., Lea V., Lee D., Donini P., Schulman A.H. Comparison of the utility of barley retrotransposon families for genetic analysis by molecular marker techniques. Mol Genet Genom 2003; 269: 464-474

Lenoir A., Lavie L., Prieto J.-L., Goubely C., Côté J.-C., Pélissier T., Deragon J.-M. The evolutionary origin and genomic organization of SINEs in *Arabidopsis thaliana*. Mol Biol Evol 2001; 18: 2315-2322

Levis R., Dunsmuir P., Rubin G.M. Terminal repeats of the Drosophila transposable element *copia*: nucleotide sequence and genomic organization. Cell 1980; 21: 581-588

Li C., Rossnagel B.G., Scoles, G. J. The development of oats microsatellite markers and their use in identifying relationships among *Avena* species and oats cultivars. Theor Appl Genet 2000; 101:1259-1268

Li J.Z., Sjakste T.G., Röder M.S., Ganal MW. Development and genetic mapping of 127 new microsatellite markers in barley. Theor Appl Genet 2003; 107: 1021-1027

Linares C., Serna A., Fominaya A. Chromosomal organization of a sequence related to LTR-like elements of Ty1-*copia* retrotransposons in *Avena* species. Genome 1999; 42: 706-713

Liu B., Wendel J.F. Retrotransposon activation followed by rapid repression in introgressed rice plants. Genome 2000; 43: 874-880

Llaca V., Messing J. Amplicons of maize zein genes are conserved within genic but expanded and constricted in intergenic regions. Plant Jour 1998; 15: 211-220

Lohe A.R., Hilliker A.J., Roberts P.A. Mapping simple repeated DNA-sequences in heterochromatin of *Drosophila melanogaster*. Genetics 1993; 134:1149-1174

Löwer R., Löwer J., Kurth R. The viruses in all of us: Characteristics and biological significance of human endogenous retrovirus sequences. Proc Natl Acad Sci USA 1996; 93: 5177-5184

Luan D.D., Korman M.H., Jakubczak J.L., Eickbush T.H. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell 1993; 72: 595-605

Lucas H., Moore G., Murphy G., Flavell R.B. Inverted repeats in the long- terminal repeats of the wheat retrotransposon WIS 2-1A. Mol Biol Evol 1992; 9: 716-728

Ma Z.Q., Roder M., Sorrells M.E. Frequency and sequence characterstics of di-, tri- and tetra-nucleotide microsatellites in wheat. Genome 1996; 39: 123-130

Manninen I., Schulman A. *BARE*-1, a copia- like retroelement in barley (*Hordeum vulgare* L.). Plant Mol Biol 1993; 22: 829-846.

Manninen O., Kalendar R., Robinson J., Schulman A.H. Application of *BARE*-1 retrotransposon markers to map a major resistance gene for net blotch in barley. Mol Gen Genet 2000; 264: 325-334

McCouch S.R.,Teytelman L., Xu Y., Lobos K.B., Clare K., Walton K., Fu B., Maghirang R., Li Z., Xing Y., *et al.* Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.). DNA Res 2002; 9: 199-207

Mellor J., Malim M.H., Gull K., Tuite M.F., McCready S., Dibbayawan T., Kingsman S.M., Kingsman A.J. Reverse transcriptase activity and *Ty* RNA are associated with virus-like particles in yeast. Nature 1985; 318: 583-586

Metzgar D., Bytof J., Wills C. Selection against frameshift mutations limits microsatellite expansion in coding DNA. Genome Res 2000; 10:72-80

Meyers B.C., Tingley S.V., Morgante M. Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome. Genome Res 2001; 11: 1660-1676

Miller J.T., Dong F.G., Jackson S.A., Song J.Q., Jiang J. Retrotransposon-related DNA sequences in the centromeres of grass genomes. Genetics 1998; 150: 1615-1623

Moore G., Cheung W., Schwarzacher T., Flavell R. BIS 1, a major component of the cereal genome and a tool for studying genomic organization. Genomics 1991; 19: 469-476

Moore G., Gale M.D., Kurata N., Flavell R.B. Molecular analysis of small grain cereal genomes: current status and prospects. Bio/Technology 1993; 11: 584-589

Moore J.K., Haber J.E. Capture of retrotransposon DNA at the sites of chromosomal double-strand DNA breaks. Nature 1996; 383: 644-645

Morgante M., Oliveri A.M. PCR-amplified microsatellites as markers in plant genetics. Plant Jour 1993; 3: 175-182

Morgante M., Hanafey M., Powell W. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. Nature Genet 2002; 30: 194-200

Motohashi R., Mochizuki K., Ohtsubo H., Ohtsubo E. Structures and distribution of p-SINE1 members in rice genomes. Theor Appl Genet 1997; 95: 359-368

Muñiz L.M., Cuadrado A., Jouve N., Gonzales J.M. The detection, cloning, and characterisation of WIS 2-1A retrotransposon-like sequences in *Triticum aestivum* and x*Triticosecale* Wittmack and an examination of their evolution in related Triticeae. Genome 2001; 44: 978-989

Nadir E., Margalit H., Gallily T., BenSasson S.A. Microsatellite spreading in the human genome: Evolutionary mechanisms and structural implications Proc Natl Acad Sci USA 1996; 93: 6470-6475

Nagano H., Wu L., Kawasaki S., Kishima Y., Sano Y. Genomic organization of the 260 kb surrounding the *waxy* locus in a Japonica rice. Genome 1999; 42: 1121-1126

Navarro-Quezada A., Shoen D.J. Sequence evolution and copy number of Ty1-*copia* retrotransposons in diverse plant genomes. Proc Natl Acad Sci USA 2002; 99: 268-273

Noma K., Ohtsubo E., Ohtsubo H. Non-LTR retrotransposons (LINEs) as ubiquitous components of plant genomes. Mol Gen Genet 1999; 261: 71-79

Nomomura K., Kurata N. Organization of the 1.9 kb repeat unit RCE1 in the centromeric region of rice chromosomes. Mol Gen Genet 1999; 261: 1-10

Nomomura K., Kurata N. The centromere composition of multiple repetitive sequences on rice chromosome 5. Chromosoma 2001; 110: 284-291

Panaud O., Chen X., McCouch S.R. Frequency of microsatellite sequences in rice (*Oryza sativa* L.). Genome 1995; 38: 1170-1176

Panaud O., Chen X., McCouch S.R. Development of microsatellite markers and characterization of simple sequence length polymorphism (SSLP) in rice (*Oryza sativa* L.). Mol Gen Genet 1996; 258: 466-478

Panstruga R., Büschges R., Piffanelli P., Schulze-Lefert P. A contiguous 60 kb genomic stretch from barley reveals molecular evidence for gene islands in a monocot genome. Nucl Acids Res 1998; 26: 1056-1062

Pardue M.L., Lowenhaupt K., Rich A., Nordheim A. (dC-dA)n.(dG-dT)n sequences have evolutionarily conserved chromosomal locations in *Drosophila* with implications for roles in chromosome structure and function EMBO Jour 1987; 6:1781-1789

Pardue M.L., Danilevskaya O.N., Lowenhaupt K., Slot F., Traverse K.L. Drosophila telomeres: new views on chromosome evolution. Trends Genet 1996; 12: 48-52

Pedersen C., Linde-Laursen I. Chromosomal location of four minor rDNA loci and a marker microsatellite sequence in barley. Chrom Res 1994; 2: 67-71

Pedersen C., Rasmussen S.K., Linde- Laursen I. Genome and chromosome identification in cultivated barley and related species of the Triticeae (Poaceae) by in situ hybridization with the GAA- satellite sequence. Genome 1996; 39: 93-104

Porceddu A., Albertini E., Baracaccia G., Marconi G., Bertoli F.B., Veronesi F. Development of S-SAP markers based on an LTR-like sequence from *Medicago sativa* L. Mol Genet Genomics 2002; 267: 107-114

Powell W., Machray G.C., Provan J. Polymorphism revealed by simple sequence repeats. Trends Plant Sci 1996; 1: 215-222

Presting G.G., Malysheva L., Fuchs J., Schubert I. A Ty3/gypsy retrotransposon-like sequence localizes to the centromeric regions of cereal chromosomes. Plant Jour 1998; 16: 721-728

Ramsay L., Macaulay M., Cardle L., Morgante M., Ivanissevich D.S., Maestri E., Powell W., Waugh R. Intimate association of microsatellite repeats with retrotransposons and other dispersed repetitive elements in barley. Plant Jour 1999; 17: 415-425

Ramsay L., Macaulay M., Ivanissevich D.S., MacLean K., Cardle L., Fuller J., Edwards K.J., Tuvesson S., Morgante M., Massari A., *et al.* A simple sequence repeat-based linkage map of barley. Genetics 2000; 156: 1997-2005

Rashkova S., Athanasiadis A., Pardue M.L. Intracellular targeting of the *Drosophila* telomeric retrotransposons. J Virol 2003; 77: 6376-6384

Roder M.S., Korzun V., Gill B.S., Ganal M.W. The physical mapping of microsatellite markers in wheat. Genome 1998a, 41: 278-283

Roder M.S., Korzun V., Wendehake K., Plaschke J., Tixer M.H., Leroy P., Ganal M.W. A microsatellite map of wheat. Genetics 1998b, 149: 2007-2023

Roder M.S., Plaschke J., Konig S.U., Borner A., Sorells M.E., Tanksley S.D., Ganal M.W. Abundance, variability and chromosomal location of microsatellites in wheat. Mol Gen Genet 1995; 246: 327-333

Rostocks N., Park Y.J., Ramakrishna W., Ma J., Druka A., Shiloff B.A., SanMiguel P.J., Jiang Z., Brueggeman R., Sandhu D. and others. Genomic sequencing reveals gene content, genomic organization, and recombination relationships in barley. Functional and Integative Genomics 2002; 2: 51-59

Rous P. A sarcoma of the fowl transmissible by an agent separable from the tumor cells. Journal of Experimental Medicine 1911; 13: 397-411

Saal B., Wricke G. Development of simple sequence repeat markers in rye (*Secale cereale* L.). Genome 1999; 42: 964-972.

Sandhu D., Gill K.S. Gene-containing regions of wheat and other grass genomes. Plant Physiol 2002; 128: 803-811

SanMiguel P., Tikhonov A., Jin Y.K., Motchoulskaia N., Zakharov D., Melake-Berhan A., Springer P.S., Edwards K.J., Lee M., Avramova Z. and others. Nested retrotransposons in the intergenic regions of the maize genome. Science 1996; 274: 765-768

SanMiguel P., Bennetzen J.L. Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. Ann Bot 1998; 82: 37-44

SanMiguel P., Gaut B.S., Tikhoniv A., Nakajima Y., Bennetzen J.L. The paleontology of intergene retrotransposons in maize. Nat Genet 1998; 20: 43-45

Sasaki T. Matsumoto T., Yamamoto K., Sakata K., Baba T., Katayose Y., Wu J., Niimura Y., Cheng Z., Nagamura Y., *et al.* The genome sequence and structure of rice chromosome 1. Nature 2002; 420: 312-316

Schmidt T. LINEs, SINEs and repetitive DNA: non-LTR retrotransposons in plant genomes. Plant Mol Biol 1999; 40: 903-910

Schmidt T., Heslop-Harrison J.S. The physical and genomic hybridization of microsatellites in sugar beet. Proc Natl Acad Sci USA 1996; 93: 8761-8765

Schmidt T., Metzlaff, M. Cloning and characterization of a *Beta vulgaris* satellite DNA family. Gene 1991; 101:247-250

Sharma S., Balyan H.S., Kulwal P.L., Kumar N., Varshney R.K., Prasad M., Gupta P.K. Study of interspecific SSR polymorphism among 14 species from *Triticum-Aegilops* group. Wheat Info Serv 2002; 95:23-28

Shepherd N.S., Schwarz-Sommer Z., Blumberg vel Spalve J., Gupta M., Weinand U., Saedler H. Similarity of the *Cin1* repetitive family of *Zea mays* to eukaryotic transposable elements. Nature 1984; 307: 185-187

Shiba T., Saigo K. Retrovirus-like particles containing RNA homologoues to the transposable element *copia* in *Drosophila melanogaster*. Nature 1983; 302: 119-124

Shimamura M., Yasue H., Ohshima K., Abe H., Kato H., Kishiro T., Goto M., Munechika I., Okada N. Molecular evidence from retroposons that whales form a clade within even-toed ungulates. Nature 1997; 388: 666-670

Shinnick T.M., Lerner R.A., Sutcliffe J.G. Nucleotide sequence of Moloney murine leukaemia virus. Nature 1981; 293: 543-548

Shirasu K., Schulman A.H., Lahaye T., Schulze-Lefert P. A contiguous 66 kb barley DNA sequence provides evidence for reversible genome expansion. Genome Res 2000; 10: 908-915

Smilde W.D, Haluskova J., Sasaki T., Graner A. New evidence for the synteny of rice chromosome 1 and barley chromosome 3H from rice expressed sequence tags. Genome 2001; 44: 361-367

Springer P.S., Edwards K.J., Bennetzen J.L. DNA class organization on maize Adh1 yeast artificial chromosomes. Proc Natl Acad Sci USA 1994; 91: 863-867

Suoniemi A., Narvanto A., Schulman A.H. The *BARE*-1 retrotransposon is transcribed in barley from an LTR promoter active in transient assays. Plant Mol Biol 1996a; 31: 295-306

Suoniemi A., Anamthawat-Jónsson K., Arna T., Schulman A.H. Retrotransposon *BARE*-1 is a major, dispersed component of the barley (*Hordeum vulgare* L.) genome. Plant Mol Biol 1996b; 30: 1321-1329

Suoniemi A., Tanskanen J., Pentikäinen O., Johnson M.S., Schulman A.H. The core domain of retrotransposon integrase in *Hordeum*: Predicted structure and evolution. Mol Biol Evol 1998; 15: 1135-1144

Svitashev S., Bryngelsson T., Vershinin A., Pedersen C., Säll T., von Bothmer R. Phylogenetic analysis of the genus *Hordeum* using repetitive DNA sequences. Theor Appl Genet 1994; 89: 801-810

Takeda S., Sugimoto K., Otsuki H., Hirochika H. Transcriptional activation of the tobacco retrotransposon *Tto1* by wounding and methyl jasmonate. Plant Mol Biol 1998; 36: 365-376

Tautz D., Renz M. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nucl Acids Res 1994; 12: 4127-4138

Temin H.M. Nature of the provirus of Rous sarcoma. Natl Cancer Inst Monogr1964; 17: 557-570

Temin H.M., Mizutani S. RNA-directed DNA polymerase in virons of Rous sarcoma virus. Nature 1970; 226: 1211-1213

Temnykh S., Park W.D., Ayres N., Cartinhour S., Hauck N., Lipovich L., Cho Y.G., Ishii T., McCouch S.R. Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa* L.). Theor Appl Genet 2000; 100: 697-712

Temnykh S., DeClerck, G., Lukashova, A., Lipovich, L., Cartinhour, S. and McCouch, S. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. Genome Res 2001; 11: 1441-1452

TAGI, The Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 2000; 408: 796-815

The Rice Chromosome 10 Sequencing Consortium. In-depth view of structure, activity, and evolution of rice chromosome 10. Science 2003; 300: 1566-1569

Thiel T., Michalek W., Varshney R.K., Graner A. Exploiting EST databases for the development of cDNA derived microsatellite markers in barley (*Hordeum vulgare* L.). Theor Appl Genet 2003; 106: 411-422

Thomas C.A. The genetic organization of chromosomes. Annu Rev Genet 1971;5: 237-256. .

Tikhonov A.P., SanMiguel P.J., Nakajima Y., Gorenstein N.M., Bennetzen J.L., Avramova Z. Colinearity and its exceptions in orthologous *adh* regions of maize and sorghum. Proc Natl Acad Sci USA 1999; 96: 7409-7414

Tikhonov A.P., Lavie L., Tatout C., Bennetzen J.L., Avramova Z., Deragon J.-M. Target sites for SINE integration in Brassica genomes display nuclear matrix binding activity. Chrom Res 2001; 9: 325-337

Trujillo R., Dugan V.L. Radiosensitivity and radiation-induced mutability: an empirical relationship. Radiat Environ Biophys 1975; 12: 253-256

Turcich M.P., Mascarenhas J.P. PREM-1, a putative maize retroelement has LTR (long terminal repeat) sequences that are preferentially transcribed in pollen. Sex Plant Reprod 1994; 7: 2-11

Turner B.G., Summers M.F. Structural biology of HIV. J Mol Biol 1999; 285: 1-32

Varagona M.J., Purugganan M., Wessler S.R. Alternative splicing induced by insertion of retrotransposons into the maize *waxy* gene. Plant Cell 1992; 4: 811-820

Varshney R.K. *A study of Some Microsatellites in Some Hexaploid Wheats*. Ph.D. thesis, Ch. Charan Singh University, Meerut, India 2001; pp. 134

Varshney R.K. Kumar A., Balyan H.S., Roy J. K., Prasad M., Gupta P.K. Characterization of microsatellites and development of chromosome specific STMS markers in bread wheat. Plant Mol Biol Rep 2000; 18: 1-12

Varshney R.K., Sharma P.C., Gupta P.K., Balyan H.S., Ramesh B., Roy J.K., Kumar A., Sen A. Low level of polymorphism detected by SSR probes in bread wheat. Plant Breed 1998; 117: 182-184

Varshney R.K., Thiel T., Stein N., Langridge P., Graner A. *In silico* analysis on frequency and distribution of microsatellites in ESTs of some cereal species. Cell Mol Biol Lett 2002; 7: 537-546

Vershinin A.V., Druka A., Alkhimova A.G., Kleinhofs A., Heslop-Harrison J.S. LINE and *gypsy*-like retrotransposons in *Hordeum* species. Plant Mol Biol 2002; 49: 1-14

Vicient C.M., Kalendar R., Anamthawat-Jonsson K., Schulman A.H. Structure, functionality, and evolution of the *BARE*-1 retrotransposon of barley. Genetica 1999a; 107: 53-63

Vicient C.M., Suoniemi A., Anamthawat-Jónsson K., Tanskanen J., Beharav A., Nevo E., Schulman A.H. Retrotransposon *BARE*-1 and its role in genome evolution in the genus *Hordeum*. Plant Cell 1999b; 11: 1769-1784

Vicient C.M., Kalendar R., Schulman A.H. Envelope-containing retrovirus-like elements are widespread, transcribed and spliced, and insertionally polymorphic in plants. Genome Res 2001a; 11: 2041-2049

Vicient C.M., Jääskeläinen M., Kalendar R., Schulman A.H. Active retrotransposons are a common feature of grass genomes. Plant Physiol 2001b; 125: 1283-1292

Wang Z., Weber J.L., Zhong G., Tanksley S.D. Survey of plant short tandem repeats. Theor Appl Genet 1994; 88: 1-6

Walbot V., Petrov D.A. Gene galaxies in the maize genome. Proc Natl Acad Sci USA 2001; 98: 8163-8164

Waugh R., McLean K., Flavell A.J., Pearce S.R., Kumar A., Thomas B.B.T., Powell W. Genetic distribution of *BARE*-1-like retrotransposable elements in the barley genome revealed by sequence-specific amplification polymorphism (S-SAP). Mol Gen Genet 1997; 253: 687-694

Wei F., Wing R.A., Wise R.P. Genome dynamics and evolution of the *Mla* (powdery mildew) resistance locus in barley. Plant Cell 2002; 14: 1903-1917

Weiner A.M. SINEs and LINEs: the art of biting the hand that feeds you. Curr Opin Cell Biol 2003; 14: 343-350

Weising K., Nybom H., Wolff K., Meyer W. *DNA Fingerprinting in Plants and Fungi*. CRC Press, Boca Raton, Florida 1995; pp 322

Weising K., Winter P. Huttel B., Kahl G. Microsatellite markers for molecular breeding. Jour Crop Prod 1998; 1: 113-143

Weil C.F., Wessler S.R. The effects of plant transposable element insertion on transcription initiation and RNA processing. Annu Rev Plant Physiol Plant Mol Biol 1990; 41: 527-552

Wessler S.R. Turned on by stress. Plant retrotransposons. Curr Biol 1996; 6: 959-961

Wessler S.R., Bureau T.E., White S.E. LTR-retrotransposons and MITEs: Important players in the evolution of plant genomes. Curr Opin Genet Dev 1995; 5: 814-821

White S.E., Habera L.F., Wessler S.R. Retrotransposons in the flanking regions of normal plant genes: A role for *copia* -like elements in the evolution of gene structure and expression. Proc Natl Acad Sci USA 1994; 91: 11792-11796

Wicker T., Stein N., Albar L., Feuillet C., Schlagenhauf E., Keller B. Analysis of a contiguous 211 kb sequence in diploid wheat (*Triticum monococcum* L.) reveals multiple mechanisms of genome evolution. Plant Jour 2001; 26: 307-316

Witte C.P., Le Q.H., Bureau T., Kumar A. Terminal-repeat retrotransposons in miniature (TRIM) are involved in restructuring plant genomes. Proc Natl Acad Sci USA 2001; 98: 13778-13783

Wright D.A., Voytas D.F. *Athila4* of Arabidopsis and *Calypso* of soybean define a lineage of endogenous plant retroviruses. Genome Res 2002; 12: 122-131

Wu K.S., Tanksley S.D. Abundance, polymorphism and genetic mapping of microsatellites in rice. Mol Gen Genet 1993; 241: 225-235

Xiong Y., Eickbush T.H. Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO Jour 1990, 9: 3353-3362

Yamazaki M., Tsugawa H., Miyao A., Yano M., Wu J., Yamamoto S., Matsumoto T., Sasaki T., Hirochika H. The rice retrotransposon *Tois17* prefers low-copy-number sequences as integration targets. Mol Gen Genom 2001; 265: 336-344

Yu G.-X., Wise R.P. An anchored AFLP- and retrotransposon-based map of diploid *Avena*. Genome 2000; 43: 736-749

Yu J., Lu H., Bernardo R. Inconsistency between SSR groupings and genetic backgrounds of white corn inbreds. Maydica 2001; 46:133-139

Yu J., Hu S., Wang J., Wong G. K.-S., Li S., Liu B., Deng Y., Dai L., Zhou Y., Zhang X., *et al.* A draft sequence of the rice genome (*Oryza sativa* L. sspl. *indica*). Science 2002; 296:79-92

Zhong C.X., Marshall J.B., Topp C., Mroczek R., Kato A., Nagaki K., Birchler J.A., Jiang J., Dawe R.K. Centromeric retroelements and satellites interact with maize kinetochore protein CENH3. Plant Cell 2002; 14: 2825-2836

Zhou S., Ke N., Kim J.M., Voytas D.F. The *Saccharomyces* retrotransposon *Ty5* integrates preferentially into regions of silent chromatin at the telomeres and mating loci. Genes Dev 1996; 10: 634-645

Chapter 5

COMPARATIVE GENOMICS IN CEREALS

Andrew H. Paterson

University of Georgia, 111 Riverbend Road, Rm 228, GA 30602 Athens, USA
E-mail: paterson@uga.edu

1. INTRODUCTION

The cereal crops represent a well-studied group of plants, which have a relatively recent history. While the angiosperm (flowering plant) lineage is thought to be about 200 million years (MY) old, cereals such as maize (*Zea*), rice (*Oryza*), sorghum (*Sorghum*), and wheat (*Triticum*) diverged from a common ancestor only about 50-70 MYA (Kellogg, 2000). The cereals provide about half of calories consumed by humans and (together with closely-related forage grasses) an even larger share of calories consumed by livestock. Parallel and independent domestications of many cereals that are reproductively isolated from one another have provided humanity with a diverse set of crops cultivated for production of carbohydrate-rich seeds (grains) and/or biomass. These cereals are collectively also adapted to a very wide range of environments. Research on these diverse crops largely proceeded independently until the realization that DNA-level similarities offered a means for comparative studies.

Many structural and functional parallels appear to have persisted since divergence of the cereals from a common ancestor – indeed, domestication of diverse cereals may have involved mutations in genes for same traits (e.g. Hu *et al.*, 2003; Paterson *et al.*, 1995), thus providing one important motivation for comparative genomics studies. Equally important, however, are the many differences in gene sequences, arrangement, and expression pattern, which collectively contribute to much of the diversity that permitted cereals to adapt (both by natural and human selection) to such a wide range of environments. The relative importance of these factors in genetic determination of phenotypic variation remains to be elucidated, although early data suggest

P.K. Gupta and R.K. Varshney (eds.), Cereal Genomics, 119-133.

strongly the importance of transcriptional regulation of expression (Doebley and Lukens, 1998).

The 35-fold variation in genome size among major cereals, from about 0.5 pg (~490 Mb) per 1C for *Oryza sativa* (rice) to 17.33 pg (~16979 Mb) per 1C for *Triticum aestivum* (bread wheat) (Bennett and Leitch, 2003), is a second important motivation for comparative genomics research in the cereals. This variation appears to be largely the result of a dynamic and lineage-specific balance between origin and elimination of mobile dispersed repetitive DNA elements (Bennetzen, 2002). The complete sequence of the rice chromosomes (TRSC, 2003; Feng *et al.*, 2002; Sasaki *et al.*, 2002), generated by the integration of 'genomic shotgun' data (Goff *et al.*, 2002; Yu *et al.*, 2002) with extensive genetic and physical mapping efforts (Chen *et al.*, 2002; Wu *et al.*, 2002; Zhao *et al.*, 2002), provides a foundation for organizing information about diverse cereals. During the past five years, DNA sequence data for the cereals expanded much more rapidly than for the other taxa. For instance, while in January of 1998, the total of 4 million bases (Mb) of cereal sequence represented about 1% of the contents of GenBank, by the end of May of 2003, the total of 2,038-Mb of cereal sequences comprised about 6% of GenBank (Paterson *et al.*, 2003). Despite this rapid ramp-up, the total set of DNA sequence data for all cereals still remains smaller than the 1C-values of many individual cereal genomes, so that far more sequences will be worked out in future than those made available so far.

In view of the large size and complexity of cereal genomes, genetic maps remain the central tool used for comparative genomics. Detailed molecular genetic maps are now available for all major cereals, e.g. (Devos *et al.*, 2000; Lee *et al.*, 2002; McCouch *et al.*, 2002; Ming *et al.*, 1998; Qi *et al.*, 1996; Röder *et al.*, 1998; Sharopova *et al.*, 2002; Sorrells *et al.*, 2003; Wight *et al.*, 2003; Wu *et al.*, 2002). These maps are suitable both for studies on comparative biology, and for crop improvement. While these so called STS-maps have been successfully used for many applications using RFLP or SSR-based methods, mapped STS markers can also be readily used to discover SNPs or InDels (Nasu *et al.*, 2002), which can be used for genotyping through many new high-throughput and low-cost technologies. The ability to acquire such information on polymorphism for corresponding loci in many crops increases the value of STS-based genetic maps, reduces the costs associated with their wider utilization, and leverages the benefits of a wealth of prior results already attached to the loci comprising these maps.

Large-insert DNA libraries, made using cloning vectors such as bacterial artificial chromosomes (BACs), are an important complement to genetic maps in cereal comparative genomics. BAC clone libraries have now been

made for most major cereals. STS-based genetic maps provide an excellent means by which physical maps based upon large-insert clones can be integrated with genetic maps (Coe *et al.*, 2002; Draye *et al.*, 2001). 'Gene mapping', by hybridization of cloned or synthetic DNA probes (Cai *et al.*, 1998) to large-insert libraries, offers many of the advantages of somatic cell genetics, in particular obviating the need of the detectable DNA polymorphism which may impose a bias on the subsets of DNA probes that can be 'mapped.' Similarly, mapping based either upon radiation hybrids (Kynast *et al.*, 2001) or on genetic stocks containing partial deletions of individual chromosomes (Sorrells *et al.*, 2003; also Chapter 19 in this book) has accelerated progress in genomics research involving cereal taxa, for which detectable DNA polymorphism is relatively poor. The amenability of BAC clones to complete sequencing has also permitted insights into molecular evolution at a much finer scale than was possible with genetic maps (eg. Ilic *et al.*, 2003), albeit for a very small and biased sample of a genome.

2. MACROCOLLINEARITY, MICROCOLLINEARITY, AND THEIR EXCEPTIONS

In diverse cereals, gene repertoires available with individual chromosomes and the arrangement of these genes along the chromosomes has evolved much more slowly than overall genome size and organization (Feuillet and Keller, 2002). In fact, cereal lineages generally seem to show somewhat slower rates of chromosome structural divergence than what is shown by many other taxa (Paterson *et al.*, 1996). It has been possible to align mapped molecular markers and genes among and within most cereal genomes using early examples (Ahn *et al.*, 1993; Hulbert *et al.*, 1990), and based largely upon mapping of common sets of DNA probes as RFLPs in different taxa.

A positive correlation, though not perfect, generally exists between the antiquity of divergence of taxa from common ancestors, and the degree of major structural rearrangements that are found among their chromosomes. For example, *Sorghum* and *Saccharum* (sugarcane), which are believed to have diverged from a common ancestor perhaps as recently as 5 million years ago, show only one major structural rearrangement that clearly differentiates the two genera (although there appears to be additional structural polymorphism within highly-polyploid *Saccharum*) (Ming *et al.*, 1998). By contrast, sorghum and maize, which are believed to have diverged from a common ancestor more than 20 million years ago, differ by at least a dozen chromosome structural rearrangements that are large enough to be discerned

by alignment of genetic maps (Bowers *et al.*, 2003a). Curiously, sorghum and rice differ by a similar number of rearrangements as sorghum and maize despite more than twice as long a period of divergence from common ancestors. Recent genome-wide duplication in the maize lineage may be one factor that accelerated rearrangement in some lineages.

The availability of molecular genetic maps for a large and growing set of cereal taxa has identified several important exceptions to the generalizations of 'slow structural change correlated to taxonomic divergence'. For example, rapid chromosomal restructuring appears to have occurred in *Secale cereale* (rye) since its divergence from a common ancestor shared with wheat (Devos *et al.*, 1993), and in *Pennisetum glaucum* (pearl millet) since its divergence from closely-related foxtail millet (Devos *et al.*, 2000). The molecular mechanism causing these rapid rearrangements may be studied through fine-mapping and sequencing of rearrangement breakpoints (presumably contained within BACs), and their comparison to homologous regions in close relatives. The details of cereal genetic maps also reveal a large number of loci that deviate from synteny and/or collinearity, providing still additional evidence that genomes are far more fluid than would have been imagined even a few years ago. The comparative distributions of loci across most pairs of cereal taxa show parallels that can not be explained by chance – however the 'signal' from these parallels is accompanied by a substantial level of 'noise' that implicates many other factors in cereal genome evolution. For example, in detailed genetic maps of maize (>3400 loci; Lee *et al.*, 2002 http://www.maizegdb.org/map.php) and sorghum (>2500 loci; Bowers *et al.*, 2003a; http://www.plantgenome.uga.edu/) that include 952 comparative loci, only about 55% of loci show corresponding arrangement (Bowers *et al.*, 2003a).

Micro-synteny studies comparing the sequences of orthologous BACs offer new opportunities for the study of deviations from collinearity. In some cases, the arrangements of genes are similar or identical in diverse cereals such as rice, maize, and sorghum. Nonetheless, cases of gene gain/loss, inversion, and tandem duplication are by no means rare (e.g. Bennetzen 2000; Ramakrishna et al 2002; Feuillet and Keller 2002; Song *et al.*, 2002; Brunner *et al.*, 2003), and may result from a variety of mechanisms including transposition and illegitimate recombination (eg. Ilic et al, 2003; Bennetzen, 2002; Song *et al.*, 2002; Wicker *et al.*, 2003).

The overall distance between genes appears to be generally correlated with differences in genome size in different taxa. However, there are noteworthy exceptions in the form of 'gene-rich' regions that largely lack repetitive DNA, even in taxa with large amounts of repetitive DNA (Feuillet and Keller

1999). The spacing between genes has been profoundly altered by acquisition of both low-copy and repetitive DNA, most of which are rapidly-evolving and do not hybridize (at least to a specific locus) across species (Chen et al 1997; Brunner *et al.*, 2003). Recent-classical studies (San Miguel et al 1996) have shown that in maize, this DNA acquisition was largely a result of successive bursts of transposable element amplification and activity, often 'jumping' into transposable elements resulting from previous cycles of amplification. This model for fluidity and associated rapid divergence of intergenic regions is gaining support from additional taxa (Wicker *et al.*, 2003). Early notions that such activity may result in rapid growth in genome size (Bennetzen and Kellogg 1997) have recently been modified into models that invoke a dynamic and lineage-specific balance between origin and elimination of mobile dispersed repetitive DNA elements (Bennetzen, 2002; Wicker *et al.*, 2003).

3. GENOME DUPLICATION AND ITS CONSEQUENCES

The cereals closely resemble other angiosperms in the relative abundance of polyploid taxa. While the potential role of polyploidy and gene duplication in evolution has long been recognized (e.g. Ohno, 1970), the question of "Why are so many plants polyploid?" remains inadequately understood. The rarity of polyploidy in dioecious organisms (such as most animals and a few plants) is thought to be related generally to sex determination, and perhaps specifically to a need for balanced gene dosage between autosomes and sex chromosomes (Orr 1990). However, a growing body of data suggests that the rarity of such 'roadblocks' (heteromorphic sex chromosomes) is not sufficient to explain the prevalence of polyploids among angiosperms.

Many cases of polyploidy in the cereals are well known. Tetraploid and hexaploid wheats are central to world agriculture, and the latter are of very recent origin. Triticale, a synthetic polyploid made by joining the wheat and rye genomes, is one of only a few examples of major crop species that were intentionally produced by humans (i.e. that do not occur naturally). The *Saccharum* (sugarcane) species complex is one of the more striking polyploid series known. In the genus *Saccharum* the chromosome numbers are known to range from $2n = 36$ to 170, but despite their ploidy differences, most of these can be freely intercrossed to produce viable and fertile progeny. By contrast, chromosome numbers in the genus *Sorghum*, which is believed to have shared a common ancestor with *Saccharum* as recently as 5 million years ago (Ming *et al.*, 1998), include only $2n = 10$, 20, or 40 -- and little if

any genetic exchange is possible between the taxa with $2n = 10$ and those with $2n = 20$.

Molecular data also revealed that many cereals thought to be diploid were, in fact, ancient polyploids. As early as 1986, non-random patterns of duplication of isozyme loci were predicted to be a result of ancient chromosomal duplication of maize (Wendel *et al.*, 1986) – a prediction that is now generally accepted. Early suggestions of ancient polyploidy in rice (Kishimoto *et al.*, 1994; Nagamura *et al.*, 1995; Wang *et al.*, 2000) were also strongly supported by analysis of genomic shotgun sequence (Goff *et al.*, 2002). This has received further support from analysis of preliminary genome-wide sequence assemblies, although there remains some controversy about whether such duplication was genome-wide (Paterson *et al.*, 2003) or limited to a subset of chromosomes (Vandepoele *et al.*, 2003). Genomic duplication in sorghum appears similar to the pattern observed in rice (Bowers *et al.*, 2003a; Chittenden *et al.*, 1994).

Recent polyploidy and ancient duplication both have important consequences for comparative genomics. In early cereal genomics studies, it quickly became obvious that only a subset of duplicated ('homoeologous') genes could be genetically mapped due to availability of limited polymorphism. The likelihood of being able to map each of two homoeologues was approximately the square of the likelihood of being able to map one – for taxa such as wheat with very little polymorphism, this was prohibitive. Even in rice, where intersubspecific crosses are common, this factor obscured the clear delineation of most duplicated chromatin until the genomic sequence became available. In maize and sorghum with high level of polymorphism, it is often possible to map duplicate loci. However, duplicate loci are far more abundant in maize than in sorghum, consistent with the occurrence of a genome-wide duplication in maize following divergence of these taxa.

Ancient duplication also appears to be followed by 'diploidization', or loss of many single members of homoeologous pairs, obscuring and complicating analysis of collinearity on both a macro- and a micro-scale. DNA sequence elimination appears to begin very soon after polyploid formation in some cereals (Eckhardt, 2001), and in polyploids as ancient as maize, many genes have been 'diploidized'. This process of loss of duplicate genes appears to be continuous. A detailed analysis of episodic duplication in the Arabidopsis lineage reveals that progressively more ancient chromatin duplications contain progressively lower percentages of genes for which both duplicated copies still exist (Bowers *et al.*, 2003b).

A sound understanding of the timing of duplication events relative to divergence of taxa from a common ancestor, is also necessary for understanding the effects of duplications on genomes having a common ancestor (Bowers *et al.*, 2003b; Kellogg, 2003). If taxon divergence postdates duplication only then the traditional 'one-to-one' comparisons commonly performed among many cereals could be truly appropriate. If duplication in one or both lineages postdates taxon divergence, more complex approaches are needed to deconvolute the consequences of 'diploidization' from the effects of chromosome structural rearrangement. Recent progress in revealing the history of ancient duplication events in dicots (Bowers *et al.*, 2003b) suggested that parallel studies are needed (and are in progress – Paterson *et al.*, 2003) in the cereals and related monocots.

4. EXTRAPOLATING INFORMATION FROM BOTANICAL MODELS ACROSS THE CEREALS

The utilization of molecular tools across diverse cereals is at least 15 years old (Hulbert *et al.*, 1990) and hardly needs to be revisited. The primary challenge associated with this approach is that cross-utilization of genomic tools to study genetic diversity requires one to identify the genomic sequences that are conserved (largely or wholly) across diverse taxa. The taxa that are compared should also each exhibit detectable polymorphisms that can serve as DNA markers. RFLPs have been used in most such studies to date, visualizing signal using a cDNA or other conserved sequence, but usually based on polymorphisms that occur in less-conserved flanking sequence. The availability of the rice genomic sequence, together with the finding that intron positions are often conserved across considerable taxonomic distances (Fedorov *et al.*, 2002), suggests the possibility of using PCR-based approaches, designing primers from conserved exons and detecting polymorphisms in intervening introns (Iwata *et al.*, 2001; Schneider *et al.*, 1999). Such approaches are readily compatible with single-stranded conformational polymorphism (SSCP)-based detection (McCallum *et al.*, 2001; Schneider *et al.*, 2002) or high-throughput methods that are sufficiently sensitive to detect single-nucleotide polymorphisms (SNPs).

The similarity of genes and inferred protein sequences across diverse taxa reflects functional conservation that has been highlighted by discoveries of independent mutations in corresponding genes in taxa as diverse as the cereals and Arabidopsis (Spielmeyer *et al.*, 2002; Yano *et al.*, 2000). A large number of genes are now known, which control agriculturally-important phenotypes in rice and analogous phenotypes in Arabidopsis. Through

association approaches, many of the genes were found to be the determinants of key traits not only in some major crops like maize (Thornsberry *et al.*, 2001), but also in less-well studied taxa such as many 'orphan crops' (Goodman *et al.*, 2002) for which genomic information is presently lacking. These 'orphan crops' are essential to sustain low-income human populations in harsh climates where inputs such as fertilizer and water are limited.

A more complex situation is available in cases, where information about positional similarities is limited. For example, in diverse cereals, several relatively complex traits such as mass per seed (Paterson *et al.*, 1995) and rhizomatous habit (Hu *et al.*, 2003) are controlled by sets of QTLs that occur at corresponding locations more often than can be explained by chance. This finding is important, since such convergence would only be likely if there exists only a few genes in which mutation frequently modifies the target phenotype. This is though in sharp contrast to the classical quantitative inheritance models that invoke infinitely large numbers of genes, but allows positional cloning of genes/QTLs for such traits (Paterson *et al.*, 1995). More specifically, such correspondence of QTLs suggests that one can use whole genome sequences of models such as rice, to identify genes/QTLs controlling traits for which no candidate genes are available. Even in the absence of corresponding QTLs, clues to the identities of single QTLs in (for example) sugarcane, might be derived from analysis of the rice sequence.

5. METHYL-FILTRATION AND Cot ANALYSIS FOR DNA SEQUENCE DIVERSITY IN CEREALS WITH LARGE-GENOMES

While comparative genomics can help us identify features shared by different cereal genomes, enhanced knowledge of diversity is key to understanding the evolution of cereal genes and genomes. A major obstacle in obtaining information about cereal DNA sequence diversity is repetitive DNA, which accounts for a major part of most cereal genomes. This repetitive DNA also largely determines the cost of sequencing these genomes by shotgun approaches. EST sequencing is an economical first step in gene discovery, but only a fraction of the transcriptome is expressed in any single source tissue. Even by studying cDNA libraries from multiple tissues, diminishing returns typically accrue after about 10^5 sequences, since many genes that are expressed only rarely or at low levels are likely to be missed, and no information is obtained on regulatory sequences or other important low-copy elements.

Two approaches have recently been suggested by which one might efficiently continue to capture low-copy DNA sequence diversity well beyond the point of diminishing returns for EST sequencing. These two approaches include methyl-filtration and Cot analysis. Methyl-filtration (Rabinowicz *et al.*, 2003) involves the production of genomic libraries enriched in hypomethylated sequences, which are rich in unique gene sequences. Although such libraries have been used for more than a decade as a source of DNA probes for RFLP mapping of single-copy loci, but this hypomethylated fraction has rarely been used for sequencing. While methyl-filtration clearly enriches the genomic DNA for genic sequences, a complication arises, since the pattern and significance of DNA methylation differs markedly between species, developmental stages, genes within an organism, and regions of a gene (Peterson *et al.*, 2002b). Consequently, exclusion of hypermethylated DNA (silenced genes) may result in the loss of important/interesting genes. Early comparisons of genomic sequences from BAC clones with sequences from libraries enriched in hypomethylated sequences suggest that as few as 50% of genes are recovered by methylation-based gene enrichment techniques (see abstract of M. Vaudin *et al.*, 44th Maize Genetics Conference, 2002, Kissimmee, FL; www.agron.missouri. edu).

The second approach of Cot analysis is based strictly on separation of DNA elements based upon their relative iteration frequency in the genome. This old approach is a powerful biochemical technique, to 'fractionate' large, repetitive DNA fractions of the genomes. Developed by Roy Britten and colleagues nearly 35 years ago, Cot analysis is based on the observation that in a solution of heat-denatured sheared genomic DNA, a specific sequence reassociates at a rate proportional to the number of times it occurs in the genome. Cot-based cloning and sequencing (CBCS: Peterson *et al.*, 2002a; Peterson *et al.*, 2001) involves fractionation of a genome into 'components' based on reassociation rates. The components of Cot-based fractionation are cloned and the number of clones sequenced from each fraction is proportional to the kinetic complexity of the corresponding fraction. By this means, the cost of capturing the unique sequence information from a large repetitive genome (such as those of most cereals) can be reduced by 50-90% (Peterson *et al.*, 2002b).

CBCS also provides for sampling of representative sequences from most repetitive element families in a genome. For example, Peterson et al (2002a) identified highly-repetitive elements that collectively accounted for 15% of the sorghum genome, in only 253 sequence reads of clones randomly-chosen from a 'HRCot' (highly-repetitive Cot fraction) library. A single element, Retrosor-6, accounted for about 13% of the highly-repetitive clones, and 6% of the total genomic DNA of sorghum. While repetitive DNA is an enormous

impediment to cereal gene and genome research, knowledge of the sequences and distribution of repeats may circumvent many problems, and indeed create new research opportunities. For example, methods based on use of the 'Alu' element family permitted many advances in human genome research (Batzer and Deininger, 2002) much before the complete sequence became available. Similar methods for 'transposon display' (Lam *et al.*, 1996) have been recently applied to rice and maize (Casa *et al.*, 2002; Jiang *et al.*, 2003). Better understanding of the physical distributions of repetitive DNA families at a resolution compatible with cloning technologies (such as over different BAC clones) may provide the means to identify 'gene-rich' genomic domains that are priorities for early sequencing. Complete physical mapping of large genomes will benefit substantially from a comprehensive knowledge of the sequences and distributions of repetitive DNA families. Identification of repetitive DNA is also valuable for masking repeats out of EST databases, significantly improving the quality of unigene sets.

6. SUMMARY AND OUTLOOK

Detailed structural, functional, and comparative analyses of cereal genes and genomes, anchored to the rice genome sequence promise to shed much light on the early events that shaped the cereal lineage. This information will be invaluable as a framework for organizing comparative information for both major and 'orphan' (Goodman *et al.*, 2002) cereals. The value of cross-utilizing genomic tools in other cereals is well known, and the finished rice sequence will offer further opportunities (such as efficient identification of large numbers of 'intron scanning' primer pairs).

However, the rice sequence is only a beginning. Comparative phenotypic, genomic and sequence information from many additional taxa will be needed to elucidate the specific events responsible for the morphological and physiological diversity that adapts different cereals to different climates, production regimes, and human needs. Such information promises to grow at an accelerating rate by virtue of efficient new methods, and will help to reveal the relative roles of different genes, and different types of genomic changes, in the evolution of phenotypic diversity among and within cereal lineages. As the identities of growing numbers of key cereal genes become known in individual taxa (e.g. Spielmeyer *et al.*, 2002), growing attention to comparative biology (e.g. Lukens and Doebley, 2001) promises to facilitate understanding of the relationship between DNA polymorphism and biological diversity.

The relationship between extensive sequence data for diverse cereals and phenotypic variation, will allow careful selection of germplasm, which has been precisely characterized for a suite of phenotypes in many diverse environments. Reliable phenotypic information, together with efficient re-sequencing approaches, applied using detailed knowledge of population structure and genetic relationships, is a promising means by which to implicate small subsets of sequences in the control of key traits (Thornsberry *et al.*, 2001). Such 'association approaches' benefit from a good understanding of the extent of linkage disequilibrium (LD) in plant populations; this information has only recently been assembled, and only for a few selected taxa to date. In outcrossing species such as maize, LD often decays to virtually undetectable levels even at opposite ends of a single gene, although rates of decay vary widely for different genes (Tenaillon *et al.*, 2002). Further, the LD is moderately stronger in low-copy regions of predominantly-selfing taxa (M. Hamblin, A. H. Paterson, S. Kresovich, unpubl. data; P. Morrell, J. M. Chandler, A.H. Paterson, unpubl. data) but still much less than might have been anticipated after the relatively short time since the domestication of many plants. Fine-scale taxon-specific information will be needed in order to comprehensively scan transcriptomes for specific genes under selection in crop gene pools. Rapid growth in information about the cereals, and continuing technological improvements, will make such goals possible sooner than we might anticipate.

7. REFERENCES

Ahn S., Anderson J., Sorrells M., Tanksley S. Homeologous relationships of rice, wheat, and maize chromosomes. Mol Gen Genet 1993; 241: 483-490

Batzer M. A., Deininger P.L. Alu repeats and human genomic diversity. Nature Rev Genet 2002; 3: 370-379

Bennett M.D., Leitch I.J. Angiosperm DNA C-values Database (release 4.0, Jan. 2003). http://www.rbgkew.org.uk/cval/homepage.html, 2003

Bennetzen J.L. Comparative sequence analysis of plant nuclear genomes: Microcolinearity and its many exceptions. Plant Cell 2000;12:1021-1029

Bennetzen J.L. Mechanisms and rates of genome expansion and contraction in flowering plants. Genetica 2002; 115: 29-36

Bennetzen J.L., Kellogg E.A. Do plants have a one-way ticket to genomic obesity? Plant Cell 1997; 9:1509-1514

Bowers J.E., Abbey C., Anderson S., Chang C., Draye X., Hoppe A H., Jessup R., Lemke C., Lennington J., Li Z., *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 2003a; 367-386

Bowers J.E., Chapman B.A., Rong J.K., Paterson A.H. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature 2003b; 422: 433-438

Brunner, S., Keller, B., and Feuillet, C. A large rearrangement involving genes and low-copy DNA interrupts the microcollinearity between rice and barley at the Rph7 locus. Genetics 2003; 164: 673-683

Cai W.W., Reneker J., Chow C.W., Vaishnav M., Bradley A. An anchored framework BAC map of mouse chromosome 11 assembled using multiplex oligonucleotide hybridization. Genomics 1998; 54: 387-397

Casa A.M., Mitchell S.E., Smith O.S., Register J.C., Wessler S.R., Kresovich S. Evaluation of Hbr (MITE) markers for assessment of genetic relationships among maize (*Zea mays* L.) inbred lines. Theor Appl Genet 2002; 104: 104-110

Chen M., Presting G., Barbazuk W.B., Goicoechea J.L., Blackmon B., Fang G., Kim H., Frisch D., Yu Y., Sun S., *et al.* An integrated physical and genetic map of the rice genome. Plant Cell 2002; 14: 537- 545

Chen M., SanMiguel P., deOliveira A.C., Woo S.S., Zhang H., Wing R.A., Bennetzen J.L. Microcolinearity in *sh2*-homologous regions of the maize, rice, and sorghum genomes. Proc Natl Acad Sci USA 1997; 94: 3431-3435

Chittenden L.M., Schertz K.F., Lin Y.R., Wing R.A., Paterson A.H. A Detailed RFLP map of *Sorghum bicolor* x *S. propinquum*, suitable for high-density mapping, suggests ancestral duplication of sorghum chromosomes or chromosomal segments. Theor Appl Genet 1994; 87: 925-933

Coe E., Cone K., McMullen M., Chen S.S., Davis G., Gardiner J., Liscum E., Polacco M., Paterson A., Sanchez-Villeda H., *et al.* Access to the maize genome: An integrated physical and genetic map. Plant Physiol 2002; 128: 9-12

Devos K.M., Atkinson M.D., Chinoy C.N., Francis H.A., Harcourt R.L., Koebner R.M.D., Liu C.J., Masojć P., Xie D.X., Gale M.D. Chromosomal rearrangement in the rye genome relative to that of wheat. Theor Appl Genet 1993; 85: 673-680

Devos K.M., Pittaway T.S., Reynolds A., Gale M.D. Comparative mapping reveals a complex relationship between the pearl millet genome and those of foxtail millet and rice. Theor Appl Genet 2000; 100: 190-198

Doebley J., Lukens L. Transcriptional regulators and the evolution of plant form. Plant Cell 1998; 10: 1075-1082

Draye X., Lin Y.-R., Q. X.-Y., Bowers J.E., Burow G.B., Morrell P.L., Peterson D.G., Presting G.G., Ren S.X., Wing R.A., *et al.* Toward integration of comparative genetic, physical, diversity, and cytomolecular maps for grasses and grains, using the *Sorghum* genome as a foundation. Plant Physiol 2001; 125: 1325-1341

Eckhardt N. A sense of self: The role of DNA sequence elimination in allopolyploidization. Plant Cell 2001.; 13: 1699-1704

Fedorov A., Merican A. F., Gilbert W. Large-scale comparison of intron positions among animal, plant, and fungal genes. Proc Natl Acad Sci USA 2002; 99: 16128-16133.

Feng Q., Zhang Y., Hao P., Wang S., Fu G., Huang Y., Li Y., Zhu J., Liu Y., Hu X., *et al.,* Sequence and analysis of rice chromosome 4. Nature 2002; 420: 316-320

Feuillet C., Keller B. High gene density is conserved at syntenic loci of small and large grass genomes. Proc Nat Acad Sci USA 1999; 96: 8265-8270

Feuillet C., Keller B. Comparative genomics in the grass family: molecular characterization of grass genome structure and evolution. Ann Bot 2002; 89:3-10

Goff S.A., Ricke D., Lan T.H., Presting G., Wang R.L., Dunn M., Glazebrook J., Sessions A., Oeller P., Varma H., *et al.* A draft sequence of the rice genome (*Oryza sativa* L. ssp *japonica*). Science 2002; 296: 92-100

Goodman R.M., Naylor R., Tefera H., Nelson R., Falcon W. The rice genome and the minor grains. Science 2002; 296: 1801-1801

Hu F.Y., Tao D.Y., Sacks E., Fu B.Y., Xu P., Li J., Yang Y., McNally K., Khush G.S., Paterson A.H., *et al.* Convergent evolution of perenniality in rice and sorghum. Proc Natl Acad Sci USA 2003; 100: 4050-4054

Hulbert S.H., Richter T.E., Axtell J.D., Bennetzen J. L. Genetic-mapping and characterization of sorghum and related crops by means of maize DNA Probes. Proc Natl Acad Sci USA 1990; 87: 4251-4255

Ilic K., SanMiguel P., Bennetzen J. A complex history of rearrangement in an orthologous region of the maize, sorghum, and rice genomes. Proc Natl Acad Sci USA 2003; 100: 12265-12270

Iwata H., Ujino-Ihara T., Yoshimura K., Nagasaka K., Mukai Y., Tsumura Y. Cleaved amplified polymorphic sequence markers in sugi, *Cryptomeria japonica* D. Don, and their locations on a linkage map. Theor Appl Genet 2001; 103: 881-895

Jiang N., Bao Z.R., Zhang X.Y., Hirochika H., Eddy S.R., McCouch S.R., Wessler S.R. An active DNA transposon family in rice. Nature 2003; 421: 163-167

Kellogg E.A. The grasses: A case study in macroevolution. Annu Rev Ecol System 2000; 31: 217-238

Kellogg E. A. It's all relative. Nature 2003; 422: 383-384

Kishimoto N., Higo H., Abe K., Arai S., Saito A., Higo K. Identification of the duplicated segments in rice chromosomes 1 and 5 by linkage analysis of cDNA markers of known functions. Theor Appl Genet 1994; 88: 722-726

Kynast R.G., Riera-Lizarazu O., Vales M.I., Okagaki, R.J., Maquieira S.B., Chen G., Ananiev E.V., Odland W.E., Russell C. D., Stec A. O., *et al.* A complete set of maize individual chromosome additions to the oat genome. Plant Physiol 2001; 125: 1216-1227

Lam W., Lee T., Gilbert W. Active transposition in zebrafish. Proc Natl Acad Sci USA 1996; 93: 10870-10875

Lee M., Sharopova N., Beavis W.D., Grant D., Katt M., Blair D., Hallauer A. Expanding the genetic map of maize with the intermated B73 x Mo17 (IBM) population. Plant Mol Biol 2002; 48: 453-461

Lukens L., Doebley J. Molecular evolution of the teosinte branched gene among maize and related grasses. Mol Biol Evol 2001; 18: 627-638

McCallum J., Leite D., Pither-Joyce M., Havey M.J. Expressed sequence markers for genetic analysis of bulb onion (*Allium cepa* L.). Theor Appl Genet 2001; 103: 979-991

McCouch S.R.,Teytelman L., Xu Y., Lobos K.B., Clare K., Walton K., Fu B., Maghirang R., Li Z., Xing Y., *et al.* Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.). DNA Res 2002; 9: 199-207

Ming R., Liu S.C., Lin Y.R., da Silva J., Wilson W., Braga D., van Deynze A., Wenslaff T.F., Wu K.K., Moore P.H., *et al.* Detailed alignment of Saccharum and Sorghum chromosomes: Comparative organization of closely related diploid and polyploid genomes. Genetics 1998; 150: 1663-1682

Nagamura Y., Inoue T., Antonio B., Shimano T., Kajiya H., Shomura A., Lin S., Kuboki Y., Harushima Y., Kurata N., *et al.* Conservation of Duplicated Segments between Rice Chromosomes 11 and 12. Breed Sci 1995; 45: 373-376

Nasu S., Suzuki J., Ohta R., Hasegawa K., Yui R., Kitazawa N., Monna L., Minobe Y. Search for and analysis of single nucleotide polymorphisms (SNPs) in rice (*Oryza sativa*, *Oryza rufipogon*) and establishment of SNP markers. DNA Res 2002; 9: 163-171

Ohno S. *Evolution by Gene Duplication*, Springer Verlag, Berlin, 1970

Orr H.A. “Why polyploidy is rarer in animals than in plants” revisited. American Naturalist 1990; 136:759-770

Paterson A., Bowers J., Peterson D., Estill J., Chapman B. Structure and evolution of cereal genomes. Curr Opin Genet Dev 2003; 13: 644-650

Paterson A.H., Lan T.H., Reischmann K.P., Chang C., Lin Y.R., Liu S.C., Burow M.D., Kowalski S.P., Katsar C.S., DelMonte T.A., *et al.* Toward a unified genetic map of higher plants, transcending the monocot-dicot divergence. Nature Genet 1996; 14: 380-382

Paterson A.H., Lin Y.R., Li Z.K., Schertz K.F., Doebley J.F., Pinson S.R.M., Liu S.C., Stansel J.W., Irvine J.E. Convergent domestication of cereal crops by independent mutations at corresponding genetic-loci. Science 1995; 269: 1714-1718

Peterson D.G., Schulze S.R., Sciara E.B., Lee S.A., Bowers J.E., Nagel A., Jiang N., Tibbitts D.C., Wessler S.R., Paterson A.H. Integration of Cot analysis, DNA cloning, and high-throughput sequencing facilitates genome characterization and gene discovery. Genome Research 2002a; 12: 795-807

Peterson D.G., Schulze S.R., Sciara E.B., Lee S.A., Bowers J.E., Nagel A.N.J., Tibbitts D.C., Wessler S.R., Paterson A.H. Integration of Cot analysis, DNA cloning, and high-throughput sequencing facilitates genome characterization and gene discovery. http://www.ncbi.nlm.nih.gov/entrez. 2001

Peterson D.G., Wessler S.R., Paterson A.H. Efficient capture of unique sequences from eukaryotic genomes. Trends Genet 2002b; 18: 547-550

Qi X.Q., Stam P., Lindhout P. Comparison and integration of four barley genetic maps. Genome 1996; 39: 379-394

Rabinowicz P.D., McCombie W.R., Martienssen R.A. Gene enrichment in plant genomic shotgun libraries. Curr Opin Plant Biol 2003; 6: 150-156

Ramakrishna W., Dubcovsky J., Park Y.J., Busso C., Emberton J., SanMiguel P., Bennetzen J. L. Different types and rates of genome evolution detected by comparative sequence analysis of orthologous segments from four cereal genomes. Genetics 2002; 162:1389-1400

Röder M.S., Korzun V., Wendehake K., Plaschke J., Ti×ier M., Leroy P., Ganal M.W. A microsatellite map of wheat. Genetics 1998; 149: 2007-2023

SanMiguel P., Tikhonov A., Jin Y.K., Motchoulskaia N., Zakharov D., Melake-Berhan A., Springer P.S., Edwards K.J., Lee M., Avramova Z., *et al.* Nested retrotransposons in the intergenic regions of the maize genome. Science 1996; 274: 765-768

Sasaki T. Matsumoto T., Yamamoto K., Sakata K., Baba T., Katayose Y., Wu J., Niimura Y., Cheng Z., Nagamura Y., *et al.* The genome sequence and structure of rice chromosome 1. Nature 2002; 420: 312-316

Schneider K., Borchardt D.C., Schafer-Pregl R., Nagl N., Glass C., Jeppsson A., Gebhardt C., Salamini F. PCR-based cloning and segregation analysis of functional gene homologues in Beta vulgaris. Mol Gen Genet 1999; 262: 515-524

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

Sharopova N., McMullen M.D., Schultz L., Schroeder S., Sanchez-Villeda H., Gardiner J., Bergstrom D., Houchins K., Melia-Hancock S., Musket T. *et al.* Development and mapping of SSR markers for maize. Plant Mol Biol 2002; 48: 463-481

Song R., Llaca V., Messing J. Mosaic organization of orthologous sequences in grass genomes. Genome Res 2002; 12: 1549-1555

Sorrells M.E., La Rota M., Bermudez-Kandianis C.E., Greene R.A., Kantety R., Munkvold J.D., Miftahudin, Mahmoud A., Ma X., Gustafson P.J., *et al.* Comparative DNA sequence analysis of wheat and rice genomes. Genome Res 2003; 13:1818-1827

Spielmeyer W., Ellis M.H., Chandler P.M. Semidwarf (sd-1), "green revolution" rice, contains a defective gibberellin 20-oxidase gene. Proc Natl Acad Sci USA 2002; 99: 9043-9048

Tenaillon M.I., Sawkins M.C., Anderson L.K., Stack S.M., Doebley J., Gaut B.S. Patterns of diversity and recombination along chromosome 1 of maize (*Zea mays* ssp *mays* L.). Genetics 2002; 162: 1401-1413

Thornsberry J.M., Goodman M.M., Doebley J., Kresovich S., Nielsen D., Buckler E.S. *Dwarf8* polymorphisms associate with variation in flowering time. Nature Genet 2001; 28: 286-289

TRCS, The Rice Chromosome 10 Sequencing Consortium: Yu Y., Rambo T., Currie J., Saski C., Kim H.R., Collura K., Thompson S., Simmons J., Yang T.-J., Nah G. *et al.* In-depth

view of structure, activity, and evolution of rice chromosome 10. Science 2003; 300: 1566-1569

Vandepoele K., Simillion C., Van de Peer Y. Evidence that rice and other cereals are ancient aneuploids. Plant Cell 2003; 15: 2192-2202

Wang S.P., Liu K.D., Zhang Q.F. Segmental duplications are common in rice genome. Acta Botanica Sinica 2000; 42: 1150-1155

Wendel, J. F., Stuber, C. W., Edwards, M. D., and Goodman, M. M. Duplicated Chromosome Segments in Maize (Zea-Mays-L) - Further Evidence from Hexokinase Isozymes. Theor Appl Genet 1986; 72: 178-185.

Wicker, T., Yahiaoui, N., Guyot, R., Schlagenhauf, E., Liu, Z. D., Dubcovsky, J., and Keller, B. Rapid genome divergence at orthologous low molecular weight glutenin loci of the A and A(m) genomes of wheat. Plant Cell 2003; 15: 1186-1197.

Wight, C. P., Tinker, N. A., Kianian, S. F., Sorrells, M. E., O'Donoughue, L. S., Hoffman, D. L., Groh, S., Scoles, G. J., Li, C. D., Webster, F. H., *et al.* A molecular marker map in 'Kanota' x 'Ogle' hexaploid oat (Avena spp.) enhanced by additional markers and a robust framework. Genome 2003; 46: 28-47.

Wu, J. Z., Maehara, T., Shimokawa, T., Yamamoto, S., Harada, C., Takazaki, Y., Ono, N., Mukai, Y., Koike, K., Yazaki, J., *et al.* A comprehensive rice transcript map containing 6591 expressed sequence tag sites. Plant Cell 2002; 14: 525-535.

Yano, M., Katayose, Y., Ashikari, M., Yamanouchi, U., Monna, L., Fuse, T., Baba, T., Yamamoto, K., Umehara, Y., Nagamura, Y., *et al.* Hd1, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the arabidopsis flowering time gene CONSTANS. Plant Cell 2000; 12: 2473-2483.

Yu, J., Hu, S. N., Wang, J., Wong, G. K. S., Li, S. G., Liu, B., Deng, Y. J., Dai, L., Zhou, Y., Zhang, X. Q., *et al.* A draft sequence of the rice genome (Oryza sativa L. ssp indica). Science 2002; 296: 79-92.

Zhao, Q., Zhang, Y., Cheng, Z. K., Chen, M. S., Wang, S. Y., Feng, Q., Huang, Y. C., Li, Y., Tang, Y. S., Zhou, B., *et al.* A fine physical map of the rice chromosome 4. Genome Research 2002; 12: 817-823.

Chapter 6

POPULATION GENETIC STRUCTURE OF WILD BARLEY AND WHEAT IN THE NEAR EAST FERTILE CRESCENT: REGIONAL AND LOCAL ADAPTIVE PATTERNS

Eviatar Nevo

Institute of Evolution, University of Haifa, Haifa 31905, Israel
E-mail: nevo@research.haifa.ac.il

1. INTRODUCTION

Wheat and barley are the two principal grain crops domesticated some 10,000 years ago in the Pre-pottery Neolithic Near East (Zohary and Hopf, 2000; Badr *et al*., 2000; Lev Yadun *et al*., 2000; Salamini *et al*., 2002;). Spectacular progress has been made during the last few decades in genomics enabling the tracking of major milestones in their domestication evolution (Peng *et al*., 2003; Chen *et al*., 2004) and, most importantly, in their spatiotemporally adaptive genomic diversity patterns and genetic structure of populations (Nevo, 1992; Nevo *et al*., 2002). The wild cereal research program at the Institute of Evolution, Haifa, Israel, was part of a general research program on the evolution of genic and genomic diversity in natural populations of diverse organisms from bacteria through plants, animals and humans (Nevo, 2001a) including *global, regional,* and *local* scales (Nevo, 1978, 1988, 1998a, 2001a; Nevo and Beiles, 1988; Nevo *et al*., 1984, 2002) (see detailed list of publications in http://evolution.haifa.ac.il). Major problems addressed were to estimate genetic diversity in nature (Nevo, 2001b) and the relative importance of the driving forces of biological evolution in an attempt to rank their importance and assess their universality. Generally, these forces can be categorized as largely stochastic (i.e., mutation, recombination, migration, and genetic drift; see Maruyama 1970,

P.K. Gupta and R.K. Varshney (eds.), Cereal Genomics, 135-163.

1977) as opposed to deterministic (i.e., natural selection). We explored the following questions. (i) How much of the genomic and phenomic diversity in nature is adaptive and processed by natural selection? (ii) What is the origin and evolution of adaptation and speciation processes under spatiotemporal variables and stressful macrogeographic and microgeographic environments?

We advanced ecological genetics into ecological genomics and globally analyzed ecological, demographic, and life history variables in 1,200 diverse species across life (Nevo *et al*., 1984), thousands of populations, and tens of thousands of individuals tested mostly for allozyme and partly for DNA diversity (Nevo, 2001b; Nevo and Beiles, 1988). Likewise, we tested thermal, chemical, climatic, and biotic stresses in several model organisms. Our results indicate abundant genotypic and phenotypic diversity in natural populations. The organization and evolution of molecular and organismal diversity in nature at global, regional, and local scales are generally nonrandom and structured; display regularities across life; and are positively correlated with, and partly predictable by, abiotic and biotic environmental heterogeneity and stress. Biodiversity evolution, even in small isolated populations, is primarily driven by natural selection including diversifying, balancing, cyclical, and purifying selective regimes, interacting with, but ultimately overriding, the effects of mutation, migration, and stochasticity. More specifically, an attempt has been made to study the population genetic structure (Nei, 1973, 1977; Crow, 2003) generally and specifically of wild cereals, as biologically and economically important model organisms that underline human civilization, and to find out if the population subdivision is an indicator of the driving forces involved in evolution (Nevo, 2001a).

Our research program at the Institute of Evolution on genetic diversity, devoted to wild cereals [mainly wild barley (Nevo, 1992) and wild emmer wheat (Nevo *et al.,* 2002)], derived from the following reasons: (i) The Near East Fertile Crescent (including Israel) is the center of origin and diversity of Old World plants (Zohary and Hopf, 2000) including the progenitors of some of the most important plants of human nutrition (e.g., wheat) and feed (e.g., barley). (ii) Israel, in particular, is a rich natural laboratory of the progenitors of cultivated plants including cereals (Nevo, 1986). (iii) Israel's high biodiversity is derived from its location at the junction of three continents, each contributing to its extraordinary biodiversity. (iv) The geological history, climatic diversity, paleoecological, and topographical diversities of Israel all contribute to its remarkable biodiversity. (v) The sharp climatic division between the northern mesic Mediterranean and southern xeric Negev Desert generates a gradient of increasing aridity, primarily southwards and secondarily eastwards toward the surrounding deserts, thereby singling out

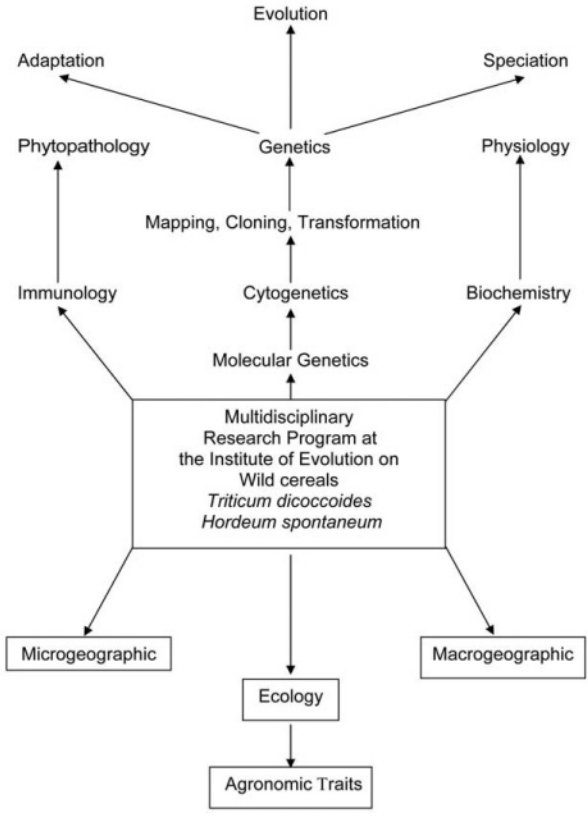


Figure 1. Multidisciplinary long-term research program of wild cereals at the Institute of Evolution, University of Haifa, Israel.

aridity stress as a major evolutionary driving force affecting genetic diversity and population genetic structure.

The above research project involved both regional and local microsites (macroevolution and microevolution within the Near East Fertile Crescent and Israel, respectively). These microsites represent sharply contrasting patterns of temperatures, aridity stress, lithology, soil types, topography, etc., i.e., climatic and edaphic stresses (Nevo, 2001a; http://evolution.haifa.ac.il) (Figs. 1 and 2a,b). Particularly, the studies of populations collected from the microsites, exhibiting contrasting ecological conditions, provide excellent critical tests for evaluating the dynamics of genome and phenome evolution. Microsite studies highlight and sharpen the study of population subdivisions while contrasting and accentuating the relative importance of evolutionary

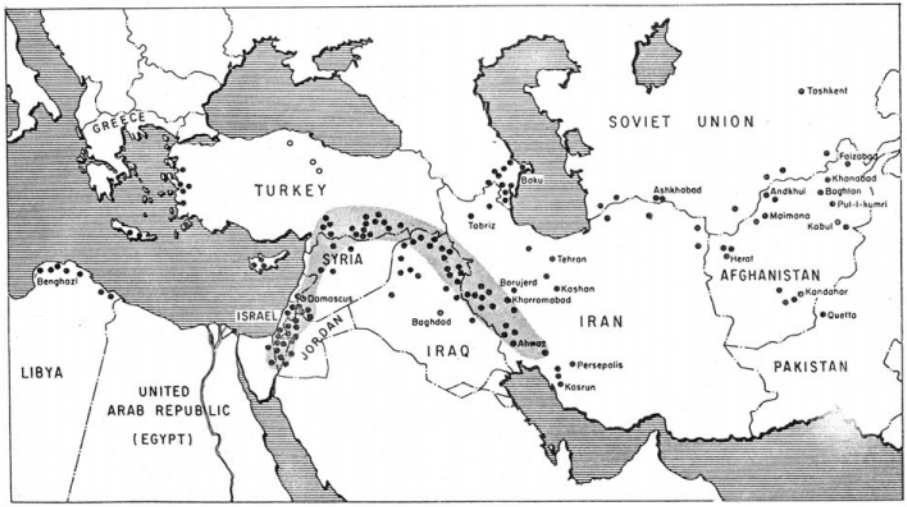


Figure 2a. Distribution of known and reasonably certain sites of wild barley. Massive stands in fairly primary habitats may occur within the shaded area. Elsewhere, wild barley may be abundant, but confined to highly disturbed habitats. The shaded area also includes the distribution of wild emmer wheat. (From Harlan and Zohary, 1966).

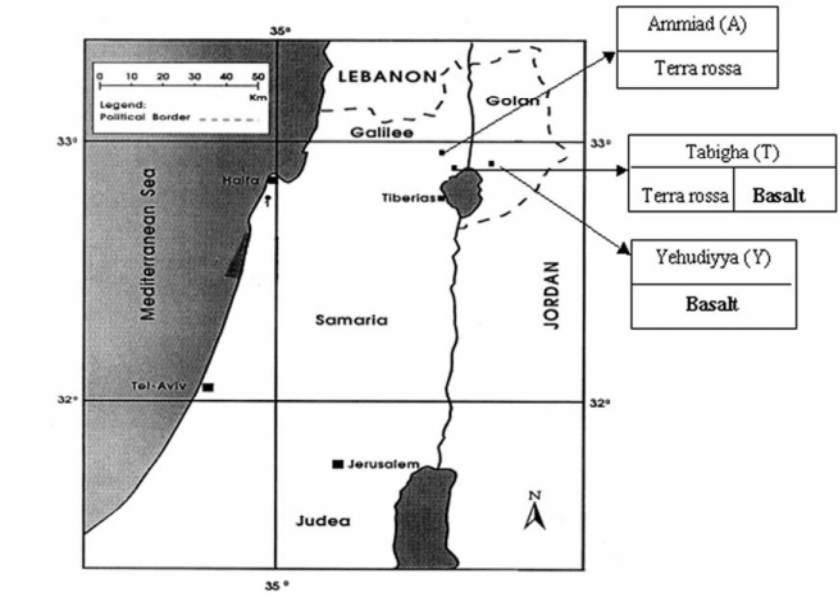


Figure 2b. The three microsites of wild emmer wheat located north of the Sea of Galilee. These sites were studied for protein and DNA diversities across divergent microclimates, soils, and topographies (Li *et al*., 2000a-d, 2001, 2002a,b, 2003).

Table 1. Gene Bank collections of wild cereals at the Institute of Evolution, University of Haifa, Israel, used for genetic and agronomic research programmes

Species	Countries	Populations	Genotypes
Hordeum spontaneum (wild barley)	5	133***	3390
Triticum dicoccoides (wild emmer wheat)	6*	32**	1659
Aegilops species	3	220	6600
Avena (wild hexaploid oats)	1	38	985
Total	6****	423	12634

*including Iran, Syria, and Iraq, which are represented by a few genotypes each; **excluding Iran, Syria, and Iraq; ***Including 3 populations of Tabigha, 4 populations of Newe Ya'ar and 7 populations of "Evolution Canyon" (see Nevo, 2001a, c); ****countries involved in the overall count

forces that cause differentiation leading to adaptation and speciation (Nevo, 2001a).

The aforementioned *regional* and *local* studies demonstrated differential viability of allozyme and DNA genotypes, where diversity and divergence were selected at either macro- and microscales contrasting conditions and ecologies or under critical empirically controlled lab conditions. In particular, dramatic population subdivisions were found in wild barley and emmer wheat populations in several microsites climaxing in the "Evolution Canyon" model (Nevo, 1995, 1997, 2001a) (Fig. 3). The *global, regional*, and *local* studies suggested that environmental stresses (climatic, thermal, chemical, and biotic) played a major role in driving evolution. In this chapter, an attempt has been made to review the results of population genetic structure of a long-term multidisciplinary research program undertaken on wild barley, *H. spontaneum* (Nevo, 1992), and wild emmer wheat, *T. dicoccoides* (Nevo *et al*., 2002), which are the progenitors of barley and wheat, respectively. The review describes genetic diversity coupled with population subdivisions *within* and *between* populations across the Near East Fertile Crescent of these wild cereals at their center of origin and diversity and periphery representing the whole ecogeographical range of the progenitors. The review will also highlight the adaptive evolutionary processes operating in these populations and the utility of these populations as potential genetic resources for crop improvement. The framework of the wild cereal research program appears in Fig. 1, and the Gene Bank collections on which the studies were based appear in Table 1.

2. SAMPLING LOCALITIES

The molecular diversity and divergence of wild barley and wild emmer wheat were examined, *regionally* in the Near East Fertile Crescent (including the countries Israel, Turkey, Iran, Syria, and Jordan) (Fig. 2a), and *locally* in five microsites in northern Israel including Ammiad, Tabigha, Yehudiyya, Newe Ya'ar, and "Evolution Canyon" (Nevo, 2001a). The five microsites in Israel (three of them shown in Figure 2b and a fourth in Fig. 3), used for the collection of wild barley and wild emmer wheat material, have the following features:(1) Ammiad, north of the Sea of Galilee, Israel (Nevo *et al*., 1991 and tables/figures therein) is a wild emmer microsite study area represented by the following habitats and subhabitats: (i) four habitats including the North-facing slope, valley, Ridge, and Karst, which display topographical and ecological heterogeneity, subdivided into (ii) 11 subhabitats that were found primarily in space over very short distances (four transects totaling 800 m and ranging in altitude from 240 to 340 m above sea level), and secondarily over time. (2) Yehudiyya, also north of the lake of Galilee, involved two sharply divergent climatic microniches in the open Tabor oak forest (i) *sunny* between trees and (ii) *shady* under trees' canopies in an area of less than 1000 m^2 involving 12 paired plots (12 trees) distributed in a mosaic pattern. (3) Tabigha, also north of the Sea of Galilee, is a microsite involving both wild emmer and wild barley, which is divided into mesic 'basalt' and xeric 'terra rossa' soil types. The terra rossa soil niche is drier and more stressful and narrow-niched than the basalt throughout the growing season (November to May); (4) Newe Ya'ar, a microsite in the lower Galilee involving only wild barley, is represented by the following four microniches: (i) *sun-soil*, (ii) *sun-rock,* (iii) *shade-soil,* (iv) *shade-rock,* and the following two contact zones: (i) soil periphery of the *sun-rock* microniche, and (ii) soil periphery of the *shade-rock* microniche; (5) "Evolution Canyon" represented by the xeric "African" and mesic "European" opposite slopes separated by 50-100 m at the bottom; the former is described as the south-facing slope or SFS, and the latter as the north-facing slope or NFS (Fig. 3). In this microsite, 2500 species have been identified, including wild barley on the opposite slopes.

3. GENETIC DIVERSITY AND ADAPTIVE SUBDIVISION OF WILD BARLEY, *HORDEUM SPONTANEUM*, IN THE NEAR EAST AND ISRAEL

Proteins (allozymes) as well as DNA markers (RAPDs, SSRs, AFLPs, rDNA) were used to study the genetic diversity among wild barley

Figure 3. "Evolution Canyon" I, Lower Nahal Oren, Mount Carmel, Israel.
Note the plant formation on the opposite slopes. The green, lush, "European", temperate, cool-mesic north-facing slope (NFS) sharply contrasts with the open park forest of warm-xeric, tropical, "African-Asian" savanna on the south-facing slope (SFS) (from Nevo, 2001a).

populations collected not only from the different regions in the Near East Fertile Crescent, but also from different microsites in Israel, each representing a contrast of edaphic, topographical, and climatic variables. The material and the markers used for the study of genetic diversity included the following (see references below). (a) Regional analyses: (i) allozymic variation in proteins encoded by 27-shared loci examined in 2125 individuals representing 52 populations of wild barley from three countries (Israel, Turkey, and Iran); (ii) genetic variability in RAPDs examined in 21 populations, seven from each of the three countries (Israel, Turkey, and Iran); (iii) DNA variation at 204 polymorphic loci out of the 268 AFLP loci examined in 39 genotypes from the Fertile Crescent; (iv) DNA variation at 18 SSR loci examined in 94 individual plants of 10 wild barley populations from Israel and in 306 individuals of 16 populations from Jordan. (b) Local analyses: Wild barley from several microsites from Israel was also examined for genetic differentiation due to edaphic and microclimatic contrasts. The barley material collected from different microsites and the markers used included the following: (i) from Newe Ya'ar representing six microniches, 356 individuals from a dense population in an area of 3182m^2 was examined for 25 allozyme loci; (ii) from Tabigha, 278 individual plants were examined for allozymes, 88 plants were examined for RAPDs and 63 plants were examined for ribosomal DNA; (iii) from "Evolution Canyon", wild barley accessions from each of the two slopes were examined for 28 allozyme loci, 51 RAPD loci, 19 SSRs, and 16 rDNA space length variants (slvs).

3.1. Allozymes: Regional and Local Studies

3.1.1. Regional Studies

The results of allozyme analysis in populations of wild barley collected from three countries (Israel, Turkey, and Iran) in the Fertile Crescent (Nevo et al., 1986a and Fig. 2a) indicated that:(a) *H. spontaneum* in the Near East Fertile Crescent is very variable genetically; (b) genetic differentiation of populations includes some clinal, but primarily regional and local patterns, often displaying sharp geographic differentiation over short distances at both single- and multilocigenome organization structures (following Brown et al., 1980, which explains the concept and analysis of genome organization); (c) the average relative genetic differentiation (GST) was 54% within populations, 39% among populations (range 29-48%) within countries, and 8% among the three countries (Table 7 in Nevo et al., 1986a); (d) allele distribution is characterized by a high proportion of unique alleles (51%) and a high proportion of common alleles that are distributed either *locally* or *sporadically*, as well as displaying an "archipelago" genetic structure; (e) discriminant analysis by allele frequencies successfully clustered wild barley of each of the three countries (96% correct classification); (f) a substantial portion of the patterns of allozyme variation in the wild gene pool is significantly correlated with the environment and is predictable ecologically, chiefly by a combination of humidity and temperature variables; (g) natural populations of wild barley are, on the average, more variable than two composite crosses and landraces of cultivated barley (see detailed analysis in Nevo, 2004). The spatial patterns and environmental correlates and predictors of genetic variation of *H. spontaneum* in the Fertile Crescent indicate that genetic variation in wild barley populations is not only rich, but, at least partly, adaptive and predictable by ecology and allozyme markers (Nevo, 1987; Nevo and Payne, 1987; Nevo *et al.*, 1993). Consequently, *conservation* and *utilization* programs should optimize sampling strategies by following the ecological-genetic factors and molecular markers as effectively predictive guidelines (see Tables 1-10 and Figs. 1, 2 in Nevo *et al*., 1986a; Likewise, Chalmers *et al*., 1992; Volis *et al*., 2001, 2002).

3.1.2. Local Studies

As mentioed above, allozymic variation in proteins of wild barley was also examined at three different microsites in Israel, Newe Ya'ar with six microniches (Nevo *et al*., 1986c), Tabigha with two soil types (Nevo *et al*., 1981), and “Evolution Canyon” with contrasting microclimates on the opposite slopes (Nevo *et al.,* 1997). *Aegilops peregrinus* also show intersoil

allozymic differences in Tabigha (Nevo *et al*., 1994). Discriminant analysis indicated significant multilocus allozymic differentiation *between* the climatic, edaphic, and topographic contrasts of the microniches. The results also suggested that allozyme polymorphisms in wild barley are at least partly adaptive and differentiate predominantly by microniche ecological selection (including soil and microclimatic types), rather than by stochastic processes and/or neutrality of allozymic variants.

3.2. DNA-Based Molecular Markers: Regional and Local studies

A variety of DNA markers were used for the study of genetic diversity in different wild barley populations listed above.

3.2.1. RAPDs: Regional and Local studies

3.2.1.1. Regional studies

Genetic variability in RAPDs (Randomly Amplified Polymorphic DNA) was studied in wild barley studied from Israel, Turkey, and Iran (Baum *et al*., 1997; Nevo *et al*., 1998). In general, high RAPD genetic diversity indices were associated with stressful environments, either with hot or cold steppes and deserts. Interpopulational genetic distances showed no association with the geographic distance between the populations' provenance. In contrast, significant Spearman rank correlations were observed between RAPD band frequencies and ecogeographical parameters of the provenance used in the study. The genetic distances *within* and *between* the three countries did not support the isolation by distance model because populations of the three countries are *intermixed* (see Fig. 2 in Nevo *et al*., 1998). Thus, *ecological* rather than *geographical factors* played a major role in genetic patterning of RAPD diversity. The correlation data indirectly suggested that natural selection appears to be the major determinant of both RAPD and allozyme diversities both of which are correlated with environmental stress (For regional adaptive RAPD variation in *H. spontaneum,* see also Volis *et al*., 2001, 2002)

.

3.2.1.2. Local studies

(i) Newe Ya'ar. A parallel study in wild barley from the aforementioned microsite Newe Ya'ar (Fig. 2b) examined variation of 75 RAPD loci (Owuor *et al*., 2003 and tables/figures therein). Significantly higher polymorphism and gene diversity ($P_{0.05}$ = 0.920; H_e = 0.411) were observed on the more

stressful *sun-rock* microniche, compared to the least stressful *shade-soil* microniche ($P_{0.05}$ = 0.653; H_e = 0.188). Fifty-six loci (74.7%) out of the 75 loci varied significantly in allele frequencies between the microniches. *Gst* analysis revealed that on an average, 75.7% of the total genetic diversity exists within the four microniches, while 24.3% exists *between* the microniches. The highest genetic distance was between *shade-soil* and *sun-rock* (D = 0.220), which confirmed their sharp microecological contrasts. Genetic classification of individual genotypes into their respective microniches, based on the presence/absence of bands of the 2, 3, 4, 5, and 7 best differentiation loci, obtained by stepwise discriminant analysis, were 79, 84, 89, 95, and 100%, respectively. We concluded that ecological contrasts at the Newe Ya'ar microsite caused molecular diversifying selection (presumably at both *coding* and *noncoding* genomic regions of wild barley) overriding gene flow effects. *(ii) Tabigha.* RAPD-PCR was also used to examine populations from the two soil types at Tabigha near the Sea of Galilee, Israel (Owuor *et al*., 1999). Analysis of 118 putative loci revealed significant ($p < 0.05$) genetic differentiation in polymorphism ($P_{0.05}$) between the two soils, basalt and terra rossa, across transects with P being higher in the more heterogeneous basalt (mean $P_{0.05}$ = 0.902) than in terra rossa (mean $P_{0.05}$ = 0.820). Gene diversity (H_e) was higher in basalt (mean H_e = 0.371) than in terra rossa (mean H_e = 0.259). Furthermore, unique alleles were confined to one soil transect II only. Gametic phase disequilibria showed a larger multilocus association of alleles in basalt than terra rossa and in transect I than in transect II. Spearman rank correlation (r_s) revealed a strong association between *specific* loci, soil types, and transects. Also, analysis of multilocus organization revealed *soil-specific* multilocus-genotypes. Therefore, our results suggest an edaphically differentiated genetic structure, which corroborate the niche width-variation hypothesis, and can be explained, in part, by natural selection on RAPDs covering both *coding* and *noncoding* regions of the genomes or linked blocks of genes. This pattern of RAPD diversity is in agreement with allozyme (Nevo *et al*., 1981) and hordein protein (Nevo *et al*., 1983) diversities in the same subpopulations studied previously.

3.2.2. Microsatellites or Simple Sequence Repeats (SSR): Regional and Local Studies

Microsatellite (SSR) diversity was examined regionally in wild barley populations from Israel across a southward transect of increasing aridity (Turpeinen *et al*., 2001 and tables/figures therein), from Jordan across a southward transect of increasing aridity (Baek *et al*., 2003 and tables/figures therein), and from the Fertile Crescent (Ivandic *et al*., 2002), and locally from

Newe Ya'ar microsite in Israel (Huang *et al*., 2002 and tables/figures therein). Allelic distribution in populations was *nonrandom*. Estimates of mean gene diversity were highest in stressful, arid-hot environments. Based on these observations, it was suggested that SSRs are at least partly adaptive in nature and are subjected to natural selection either *within* or *between* genes (Li *et al*., 2002a, 2004; Ivandic *et al*., 2002) and at least partly involved in regulatory processes.

3.2.3. AFLP: Regional Studies

In wild barley (*H. spontaneum*) populations from Israel, adaptive patterns similar to those for SSRs, were also obtained by AFLP (Amplified Fragment Length Polymorphisms) analysis (Turpeinen *et al*., 2003). AFLP diversity in *H. spontaneum* in the Fertile Crescent was based on 39 genotypes (Pakniyat *et al*., 1997). There were 268 AFLP loci of which 204 proved to be polymorphic. Genotypes were grouped according to the area of origin. Shoot Na^+ content and carbon isotope (δ^{13}) reflecting drought resistance were associated with site of origin of the ecogeographic data. In Israel (Turpeneinin *et al*., 2003) 204 AFLP loci were tested out of which 189 (93%) were polymorphic. Genetic diversity was 31% *between* populations and 69% *within* populations. Associations between ecogeographical variables and the mean gene diversity were found at one primer pair.

3.2.4. Ribosomal DNA (rDNA): Regional and Local Studies

Microecological divergence in *H. spontaneum* was also estimated through ribosomal DNA in (i) the study of sixty-three genotypes from: "Evolution Canyon" (Fig. 3) representing 36 genotypes from the contrasting microclimates on the opposite slopes and 27 genotypes from Tabigha (Gupta *et al.*, 2002), (ii) 42 genotypes from Newe Ya'ar in Israel (Gupta *et al*., 2004), and in (iii) 285 genotypes of *H. spontaneum* across Jordan (Sharma *et al*., 2004). Spacer length variants (slvs) and the slv phenotypes largely correlated with environmental conditions. These studies suggested that ecogeographical factors, rather than geographical factors per se, seem to affect the distribution of rDNA alleles. Overall, it can be concluded that rDNA repeat unit length polymorphism (rDNA diversity) in wild barley is distributed non-randomly and adaptively in different microniches. A novel feature i.e. homogenization of intergenic speacer (IGS) length at two rDNA loci located at two nonhomologous chromosomes was also demonstrated in one study (Gupta *et al.*, 2004).

4. GENETIC DIVERSITY AND ADAPTIVE SUBDIVISIONS OF WILD EMMER WHEAT IN ISRAEL AND TURKEY

In wild emmer wheat, the following populations from Israel and Turkey were examined utilizing allozymes and DNA markers. (i) 1815 plants representing 37 populations (33 from Israel and 4 from Turkey) for allozymes; (ii) 135 genotypes representing 15 populations from Israel and Turkey for DNA markers involving a wide range of ecological conditions of soil, temperature and water availability.

Microscale analysis was also conducted in wild emmer wheats from several microsites using the following markers and material: (i) from Ammiad, 812 individual plants were examined over 6 years for allozyme variation at 43 loci; this study was extended to examine spatiotemporal variation at 35 allozyme loci in 1207 individuals representing 4 populations (see references below); (ii) in another study, 75-175 individuals from Ammiad were examined for variation at 35 allozymes, 60 RAPDs and 25 SSRs, in four habitats that differed in aridity stress; (iii) from Yehudiyya, 137 individuals were examined for 48 allozyme loci and 105 individuals were examined for 28 SSR loci; (iv) from Tabigha, representing two soil types (mesic basalt and xeric terra rosa), 356 individuals were examined for 47 allozyme loci and 28 SSRs.

4.1. Allozyme Variation: Regional and Local Studies

4.1.1. Variation in Populations from Israel and Turkey

Allozyme variation in the tetraploid wild emmer wheat, Triticum dicoccoides, the progenitor of all cultivated wheats, was studied for proteins encoded by 42 gene loci in 1815 plants representing 37 populations (33 from Israel and 4 from Turkey) sampled from 33 localities during 1979 to 1987 (Nevo and Beiles, 1989 and tables/figures therein). The results showed that: (a) 6 loci (14%) were monomorphic in all populations, 15 loci (36%) were locally polymorphic, and 21 loci (50%) were regionally polymorphic. These results are similar to those obtained earlier for 12 Israeli populations (Nevo et al., 1982). All polymorphic loci (except 4) displayed high levels of polymorphism (10%) locally. (b) The proportion of polymorphic loci per population averaged 0.220 (range, 0.050 - 0.415), and the mean number of alleles per locus was 1.252 (range: 1.050-1.634); genic diversity, H_e,

averaged 0.059 (range: 0.002 - 0.119). (c) Altogether there were 119 alleles at the 42 putative loci tested, 114 (96%) of these in Israel. (d) Genetic differentiation was primarily *regional* and and *local*, not clinal; 70% of the variant alleles were common (≥10%) and not widespread, but rather *localized* or *sporadic* (see classification in Marshall and Brown, 1975) displaying an *"archipelago"* population genetics and ecology structure. "Archipelago" genetic structure alleles are built up locally in high frequency, but are often missing in neighboring localities. This phenomenon may even occur in the central continuous populations in which alternative fixation of up to eight alleles was described over ten to hundreds of meters in the Golan Heights between Qazrin and Yehudiyya (Nevo *et al*., 2002, pp. 75-77). The coefficients of genetic distance between populations were high and averaged $D = 0.134$; range, 0.018 – 0.297, an indication of sharp genetic differentiation over short distances. (e) Discriminant analyses differentiated Israel, between central and 3 marginal regions as well as those having different soil-type populations (Fig. 2a-f in Nevo and Beiles, 1989). (f) In G_{ST} analysis, *allozymic* variation comprised 40% *within* and 60% *between* populations (Table 8 in Nevo and Beiles, 1989) indicating dramatic interpopulation divergence. (g) Gametic phase disequilibria were abundant, their number being positively correlated ($r_s = 0.60$, $P<0.01$) with the humidity. (h) Multilocus organization (Brown *et al.,* 1980) was substantive, also positively correlated with humidity. (i) Allozyme diversity, overall and at single loci, was significantly correlated with, and partly predictable by, climatic and edaphic factors. (j) The distribution of the significant positive and negative values and the absence of autocorrelations in the correlogram revealed no similar geographic patterns across loci eliminating migration as a prime factor of population genetic differentiation.

The above results suggested the following. (i) During the evolutionary history of wild emmer, diversifying natural selection, through climatic and edaphic factors, was a major agent of genetic population structure and differentiation at both single and multilocus levels; and (ii) wild emmer wheat harbors large amounts of genetic diversity exploitable as genetic markers in sampling and as abundant adaptive genetic resources for wheat improvement (see also Nevo, 2001c; Nevo *et al*., 2002; and Peng *et al*., 2000, 2003).

4.1.2. Allozyme Loci Exhibiting Spatiotemporal Ecological Differentiation at Ammiad

In wild emmer wheat, *Triticum dicoccoides* from microsites at Ammiad, allozymic variation in proteins encoded by 43 loci was analyzed during four consecutive growing seasons during 1983-1984 to 1986-1987 (Nevo *et al*.,

1991 and tables/figures therein). In four vegetationally and topographically defined habitats, and 11 subhabitats (see above), significant genetic differentiation was found primarily in space over very short distances (four transects totaling 800 m, and ranging in altitude from 240 to 340 m above sea level), and secondarily over time. The highest gene diversity occurred in the Karst formation where soil moisture was most variable, i.e., in the habitat displaying the broadest niche. Genetic distance, D averaged 0.049 (range 0.005-0.134). Genetic differentiation among populations for 24 polymorphic allozyme loci across 4 years of study was G_{ST}=0.12 (range 0.002 to 0.34). In other words, 87.8% of the allozymic variation was *within* and 12.2% was *between* the 4 major habitats. This is an extraordinarily high G_{ST} estimate for a microsite whose largest distance between the habitats is only 0.5 km. The results suggest that allozyme polymorphisms in wild emmer wheat are partially adaptive. Genetic differentiation appears to be primarily affected by environmental factors related to topography and temporal microclimatic changes, probably through drought, i.e., aridity stress because the geological background (Middle Eocene Limestone) is similar in all habitats, except the Karst.

4.1.3. Allozyme Variation Related to Aridity Stress at Ammiad

Another spatiotemporal allozyme study of wild emmer, *T. dicoccoides* at Ammiad examined 35 allozyme loci over 6 years (Li *et al*., 2001). This analysis used new methods and two additional sample sets (1988 and 1993) and previous allozymic data (1984-1987, analyzed by Nevo *et al*., 1991). Significant temporal and spatial variations in allele frequencies and levels of genetic diversity were detected in 11 subpopulations from 11 subhabitats and in four populations with a mean G_{ST} of 0.152 *between habitats* and 0.173 *between years* (Figs. 4, 5). Significant associations were observed among allele frequencies and gene diversities at different loci, indicating that many allele frequencies change over time in the same or opposite directions. Multiple regression analysis showed that variation in soil-water content and rainfall distribution in the growing season significantly affected frequencies of ten alleles, numbers of alleles at eight loci, and gene diversity at four loci. Random genetic drift and hitchhiking models may not explain such locus-specific spatiotemporal divergence and strong allelic correlation or locus correlation as well as the functional importance of allozymes. Natural ecological selection, presumably through water stress, might be an important force adaptively directing spatiotemporal allozyme and DNA diversity in wild wheat both regionally and locally.

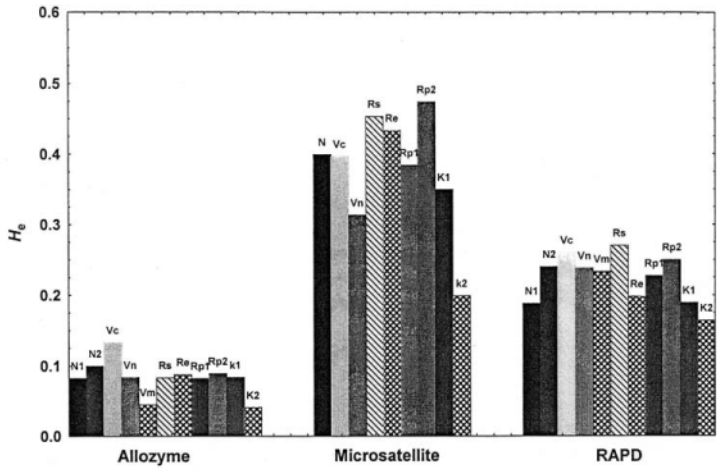


Figure 4. Histogram of average gene diversity (H_e) in the 11 subpopulations of *T. dicoccoides* from the 11 subhabitats in Ammiad revealed by 35 allozymes, 25 SSRs, and 60 RAPDs loci. *N1* Upper North, *N2* Middle North, *Vn* Narrow Valley, *Vc* Main Valley center, *Vm* Main Valley margins, *Rs* Ridge south, *Re* Ridge east, *Rp1* Ridge shoulder, *Rp2* Ridge top, *K1* Upper Karst, *K2* Lower Karst (from Nevo *et al*., 2002a, based on Li *et al*., 2000c).

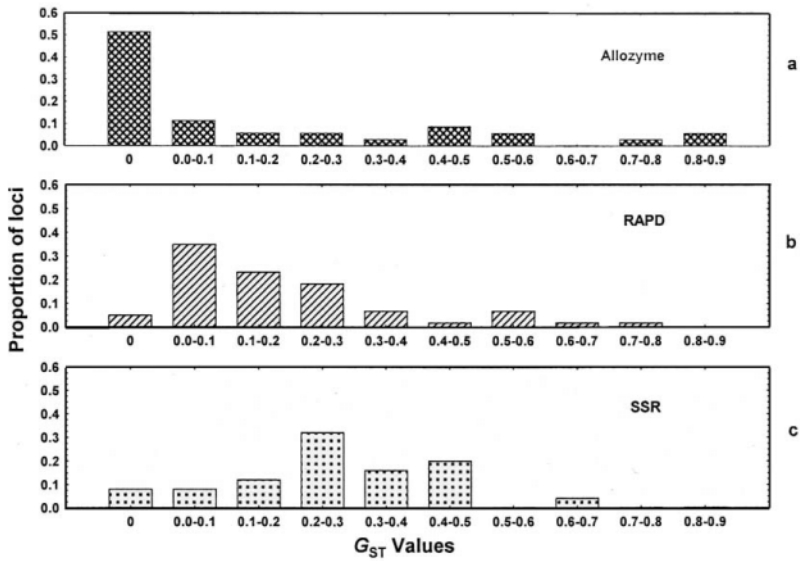


Figure 5. Histogram showing the proportional distribution of GST values among the four subpopulations of *T. dicoccoides* at 35 allozymes, 60 RAPDs, and 25 microsatellite loci in Ammiad (based on Li et al., 2000c) '0' – no GST value, i.e., monomorphic loci (from Nevo *et al*., 2002)**.**

4.2. DNA-Based Molecular Markers (SSRs)

In 15 populations (135 genotypes) of wild emmer wheat, *Triticum dicoccoides*, representing a wide range of ecological conditions of soil, temperature, and water availability in Israel and Turkey, diversity was examined at 20 microsatellite loci (Fahima *et al*., 2002). Extensive diversity at these SSR loci was observed despite the predominantly selfing nature of this plant species. The 20 Gatersleben wheat microsatellites (gwm), located on 13 chromosomes of the genomes A and B of wheat, revealed a total of 364 alleles with an average of 18 alleles per gwm marker (range: 5-26), which is comparable to an average frequency of 18.1 alleles per locus, registered in 998 accessions of hexaploid cultivated bread wheat (*Triticum aestivum*) originating in 68 countries (Huang, X. Q. *et al*., 2002). The proportion of polymorphic loci per population of the wild emmer averaged 0.90 (range: 0.45-1.00); genic diversity, H_e, averaged 0.50 (range 0.094-0.736); and Shannon's information index averaged 0.84 (range 0.166-1.307). The coefficients of genetic distance between populations were high and averaged $D = 1.862$ (range 0.876-3.320), an indication of sharp genetic divergence over short distances. Interpopulation genetic distances showed no association with geographic distances between the population sites of origin, which ruled out a simple isolation by distance model. Genetic dissimilarity values between genotypes were used to produce a dendrogram of the relationships among wild wheat populations by the unweighted pair-group method with arithmetic averages (UPGMA). Microsatellite analysis was found to be highly effective in distinguishing genotypes of *T. dicoccoides*, originating from diverse ecogeographical sites in Israel and Turkey, with 88% of the 135 genotypes correctly classified into sites of origin by discriminant analysis. The patterns of microsatellite distribution were *nonrandom* and were in agreement with the previously obtained allozyme (Nevo *et al*., 1982; Nevo and Beiles, 1989) and RAPD (Fahima *et al*., 1999) patterns although the genetic-diversity values obtained with microsatellites are much higher. Significant correlates of microsatellite markers with various *climatic* and *soil* factors suggest that, as in allozymes and RAPDs, natural selection causes adaptive microsatellite ecogeographical differentiation, not only in *coding*, but, most importantly, in *noncoding* genomic regions. Hence, the concept of "junk DNA" needs to be replaced, at least partly, by that of regulatory DNA. The results suggest that microsatellite markers are useful for the estimation of genetic diversity *within* and *between* natural populations of *T. dicoccoides* displaying the unique "archipelago" shown earlier by allozymes (Nevo and Beiles, 1989) (see below for "archipelago" genetic structure).

4.3. Parallel Protein and DNA Genetic Differentiation at Ammiad and Yehudiyya Microsites

A parallel analysis of proteins (isozymes) and DNA polymorphism in wild emmer wheat populations from the microsites of Ammiad and Yehudiyya in Israel, suggested that relatively higher genetic differentiation (G_{ST}) existed between populations in the microsites-similar to the regional pattern, and that the genetic diversity increased with increasing aridity.

4.3.1. Populations from Ammiad

At Ammiad, larger genetic distance was found at SSR loci, followed by RAPDs and allozyme loci (Fig. 4). The populations in drier habitats tend to have higher allozyme, RAPD, and SSR diversities (H_e). Allele distributions at microsatellite loci were *nonrandom* and associated with habitats. Significant genetic differentiation and variation in repeat numbers were found among subpopulations in the four major habitats and nine subhabitats. Habitat-*specific* and *–unique* alleles and linkage disequilibria were observed in the Karst population. The populations dwelling in drier habitats and subhabitats showed higher genetic diversities at microsatellite loci. These results suggest that natural selection, presumably through aridity stress, acts upon microsatelllite divergence in coding and noncoding sequences, thereby contributing to differences in fitness. These results have profound implications for genetic conservation both *in situ* and *ex situ* (Nevo, 1998b).

4.3.2. Populations from Yehudiyya

At Yehudiyya, significant genetic differentiation in allozymes (Nevo *et al.,* 1988a) and SSRs (Li *et al*., 2002b, 2003), at single-, two- and multilocus structures were found between neighboring climatic niches, *sunny* and *shady*, which were only separated by a few meters compared across a series of trees whose canopies shade the wheat plants that are exposed to sun between the trees. The results suggested that in wild emmer wheat, allozyme and SSR polymorphism, both in *coding* and *noncoding* regions of the genome, are partly adaptive and differentiate primarily at the multilocus level by climatic factors presumably related to aridity stress. The average allozyme genetic differentiation *between* the two climatic *sunny* and *shady* microniches was remarkably high for a microsite G_{ST} = 0.22 (see Table 4 in Nevo *et al*., 1988a). Significant microclimatic divergence characterized many SSR loci displaying asymmetric and *nonrandom* distribution of repeat numbers.

Niche-*specific* and niche-*unique* alleles and *linkage disequilibria* diversified the *sunny* and *shady* populations, separated by a few meters. At both single- and two-locus levels, the microclimatic environment clearly affects microsatellite diversity. Allele distribution at SSR loci is clustered and constrained with lower or higher boundary. This may imply that SSRs have functional significance and natural constraints (Li *et al*., 2003). Genetic factors involving genome, chromosome, motif, and locus significantly affected SSR diversity. Genome B appeared to have a larger average repeat number (ARN), but lower variance in repeat number (σ^2_{ARN}), and smaller number of alleles per locus than genome A. SSRs with compound motifs showed larger ARN than those with perfect motifs. The effects of replication slippage and recombinational effects (e.g., unequal crossing over) on SSR diversity varied with SSR motifs. Ecological stresses (*sun* vs. *shade*) may affect mutational mechanisms influencing the level of SSR diversity by both processes. The evidence reflects effects of ecological stresses and genetic constraints on diversifying natural selection on SSR diversity resulting presumably in adaptive structures.

4.4. Edaphic Adaptive Population Subdivision at the Tabigha Microsite

Allozymic variation in proteins encoded by 47 loci was analyzed during 1983-84 and 1984-85 in 356 individual plants of wild emmer wheat, *Triticum dicoccoides,* from Tabigha (Nevo *et al*., 1988b). Each year the test involved two 100-meter transects, each equally subdivided into mesic basalt and xeric terra rossa soil types. Comparisons were based on 16 common polymorphic allozyme loci. Significant genetic differentiation, genetic phase disequilibria, and genome organization according to soil type differentiated the populations from the two soil types. We also examined 28 SSRs, in emmer wheat populations from the same microsite (Li *et al.*, 2000b,d). The SSRs also significantly diverged in allele distribution repeat length, genetic diversity, and linkage disequilibria between the two soil types. Soil-*specific* and -*unique* SSR alleles and linkage disequilibria were observed in the terra rossa and basalt populations. A permutation test showed that the effects of random genetic drift were very low for the significant genetic diversity at microsatellite loci between the two populations, suggesting that an adaptive molecular pattern derived by edaphic selection may act upon variation of the microsatellites that represent both *coding* and *noncoding* genomic regions (Morgante *et al*., 2002; Li *et al.,* 2004). The results suggest that allozyme and SSR polymorphisms in wild emmer wheat are partly adaptive and that they differentiate at both single and multilocus structures primarily caused by

environmental stress of ecological factors, such as soil type, topography, and temporal changes probably through aridity stress. The functional significance of microsatellites in genes across plants and animals was reviewed by Li et al. (2004).

In a more comprehensive study of 27-dinucleotide SSR loci diversity, correlation was found between SSR diversity and soil variation in the three microsites (Yehudiyya, Tabigha, and Ammiad) in northern Israel (Li *et al*., 2000b). The results demonstrated that SSR diversity is correlated with the interaction of ecological and genetic factors. Genetic factors including genome (A vs. B), chromosome, motif, and locus affected average repeat number (ARN), variance in repeat number (σ^2), and number of alleles (NA) of SSRs, but the significance of some factors varied among populations. Genome effect on SSR variation may result from different motif types, particularly compound (or imperfect) versus perfect motifs, affected SSR variation. Soil-*unique* and soil-*specific* alleles were found in two edaphic groups dwelling on terra rossa and basalt soils across *macro*- and *micro*geographical scales. The largest contributions of genetic and ecological effects were found for diversity for ARN and NA, respectively. Multiple regressions indicated that replication slippage and unequal crossing over could be important mutational mechanisms, but their significance varied among motifs. Edaphic stresses may affect the *probability* of replication errors and recombination intermediates and thus control diversity level and divergence of SSRs. The results may indicate that SSR diversity is adaptive, channeled by natural selection, and influenced by both internal genetic and ecological external factors and their interactions.

5. "EVOLUTION CANYON"

The "Evolution Canyon" model is a research program examining canyon biodiversity evolution from bacteria to mammals (Nevo 1995, 1997, 2001a). To date, 2500 species have been identified in an area of 7000 m^2 (Fig. 3). Collections of wild barley, *H. spontaneum*, were compared on the "European" NFS and "African" SFS for 28 allozyme loci, (Nevo *et al*., 1997), 51 RAPD loci (Owuor *et al*., 1997), 19 nuclear SSR loci (Meyer *et al*., 2004; Nevo *et al*., 2004), and 2 rDNA loci (Gupta *et al*., 2002). Intra- and inter-slope *nonrandom* differences were observed in all molecular markers (allozymes, RAPDs, AFLP, SSRs, and rDNAs). The interslope differences include polymorphism, genic diversity, slope *specific* and *unique* alleles, linkage disequilibria, G_{ST}, and genetic distances (D), e.g., the interslope SSR D_A = 0.481 between midstations on the opposite slopes. This D_A found between populations separated by 200 m, is as large as that found between

Talpiyyot near Jerusalem and Sede Boqer in the northern Negev Desert separated by 100 km. Genetic distance due to allozyme loci even between up and midslope stations separated by 30 m on the "African" slope, sharply differed microclimatically (Pavlicek *et al*., 2003); D = 0.113, as large as between the Afiq Golan population and Wadi Qilt Judean Desert population, separated by 100 km, i.e., 3000-fold larger (Nevo *et al*., 1997). The largest genetic distance due to RAPD (D = 0.402), was between the "European" low and "African" mid-station (Owuor *et al*., 1997).

6. "ARCHIPELAGO" POPULATION: GENETIC STRUCTURE OF WILD EMMER WHEAT AND WILD BARLEY

Wild emmer grows in lush and extensive stands in the center of its origin and diversity in the catchment area of the Upper Jordan Valley (in northern Israel, in the eastern Upper Galilee Mountains, and the Golan Heights). However, elsewhere in the Fertile Crescent (Fig. 2.4 in Nevo *et al*., 2002), populations of wild emmer are semi-isolated or isolated, mostly displaying a patchy structure. At least in Israel, but possibly also elsewhere across the range of wild emmer in the Fertile Crescent, populations are subdivided into demes or clumps of varying sizes including large, medium, and small patches (see Fig. 2b in Nevo and Beiles, 1989). This "archipelago structure" (Table 4.4 in Nevo *et al*., 2002, and Fig. 5 here) was originally described for a smaller sample of wild emmer in Israel (Nevo *et al*., 1982) and Turkey (Nevo *et al*., 1986b), and for wild barley, *H. spontaneum*, in the Near East (Nevo 1992; Nevo *et al*., 1986a). The main characteristics of the "archipelago" genetic structure relate to allele distributions and are described below.

The highly subdivided, "archipelago"-type ecological population structure of wild emmer is even more distinct than that of wild barley and is matched by its genetic population structure. Gene differentiation *within* and *between* populations, which were sometimes geographically very close, was more pronounced than those *between* wild emmer wheat populations from Israel and Turkey (Table 3 in Nevo and Beiles, 1989). Of the total genetic diversity of *T. dicoccoides*, 40% was within populations and 60% was between populations whereas only 5% was found between Israel and Turkey as metapopulations (Table 8 in Nevo and Beiles, 1989 - presents G_{ST} ranges 0.05-0.91 for different loci and 0.44-0.83 for central, marginal, and isolated populations). This conclusion is reinforced by the microgeographic allozyme analysis based on edaphic, topographic, and temporal differentiation (Nevo *et al*., 1988b), and on microclimatic spatial differentiation (Nevo *et al*., 1982;

Golenberg and Nevo, 1987). Similar patterns to those of the allozyme analysis were obtained with DNA analyses at *macro-* (Fahima *et al*., 1999, 2002) and *micro*-levels (Li *et al*., 1999, 2000a-d, 2001, 2002a,b, 2003).

The high genetic differentiation within and between populations of *T. dicoccoides* is primarily reflected by the analysis of allele distribution. The latter revealed sharp local and regional differentiation in overall genetic indices (Table 4.1 in Nevo *et al*., 2002) as well as in individual allele frequencies (Appendix in Nevo and Beiles, 1989). Out of the 119 alleles at 42-shared allozyme loci across Israel and Turkey, five alleles (8.2%) were *unique* for Turkey. In Israel, allele *uniqueness* was also found for the ecogeographic subdivisions as follows. (1) Out of 93 alleles occurring in central populations, 17 alleles (18.3%) were *unique* in *marginal* populations: (2) The southeastern margins had 84 alleles of which six were *unique* (7.1%). (3) The western margins had 79 alleles, 6 were *unique* (7.6%) (4) The northeast (Mt. Hermon) had 53 alleles of which one was *unique* (1.9%). (5) Based on *soil types*: basalt had 89 alleles, of which 11 were *unique* (12.4%). And finally, (6) regarding *population size*: "large" populations had 93 alleles of which 17 were *unique* (18.3%); "medium" populations had 78 alleles, 11 were *unique* (14.1%); and "small" populations had 82 alleles, 10 were *unique* (12.2%) (See discriminant analysis in Fig. 4.2 in Nevo *et al*. 2002 and Table 3 and 8 in Nevo and Beiles, 1989).

It was also shown that 70% of all variant isozyme alleles were not widespread, but revealed *localized* and *sporadic* distribution (Following the method of Marshall and Brown, 1975). Likewise, the analysis of genetic distances between populations supports the conclusion based on genetic differentiation and allele distribution that sharp *local* differentiation over short geographic distances is the rule, and the frequency of some common alleles (>10%) is *localized* and high (a high frequency of an allele may be neighbored by a low or zero frequency). The population genetic structure of wild emmer is *mosaic*. This genetic mosaicism appears to reflect the underlying ecological heterogeneity, which derives from local and regional geological, edaphic, climatic, and biotic differentiations. The resulting structure, as in wild barley, is an ecological genetic "archipelago", where the genetic structures are in accordance with the ecological patterns, defying the geographical isolation by distance model.

7. THE ADAPTIVE NATURE OF GENIC DIVERSITY

The molecular diversity and divergence of wild emmer wheat and wild barley, *regionally* in the Near East Fertile Crescent, and *locally* in five

natural populations at Ammiad, Tabigha, Yehudiyya, Newe Ya'ar, and "Evolution Canyon" microsites in northern Israel display parallel ecological-genetic patterning (Nevo, 2001a). The *regional* and *local* results demonstrated significant spatial and temporal molecular divergence at the DNA and protein levels in *T. dicoccoides* and *H. spontaneum* subdivided populations. Specifically, these patterns revealed the following: (1) significant genetic diversity exists at single-, two-, and multilocus structures of allozymes, RAPDs, AFLPs, and SSRs diverging over very short distances of several to a few dozen meters. (2) Largely *noncoding* (RAPDs and SSRs, but see Morgante *et al*., 2002) genomic regions are correlated with, and predictable by, environmental stress (*climatic, edaphic,* and *biotic*) and heterogeneity (the niche-width variation hypothesis) displaying significant *niche-specific* and *–unique* alleles and genotypes. (3). The genomic organization of wild cereals is *nonrandom*, heavily *structured*, and at least partly, if not largely, *adaptive* regardless of the "archipelago" genetic population structure. It defies explanation by genetic drift, neutrality, or near neutrality models as the primary driving forces of wheat molecular evolution. The only viable model explaining the genomic organization of wild cereals is natural selection, primarily *diversifying, balancing,* and *cyclical selection* regimes over space and time according to the two- or multiple-niche ecological models. Spatial models are complemented by temporal models of genetic diversity and change (Kirzhner *et al*., 1995, 1996, 1998, 1999; Korol *et al*., 1996). Natural selection may interact with mutation, migration, and stochastic factors but overrides them in orienting wild barley and wild emmer wheat evolutionary processes.

Based on mathematical modeling, we established that stabilizing selection with a cyclically moving optimum appears to efficiently protect polymorphism for linked loci, affecting the selected trait (Korol *et al*., 1994, 1996; Kirzhner *et al*., 1995). In particular, unequal gene action and/or dominance effects may lead to local polymorphism stability with substantial polymorphism attracting domain. Moreover, under strong cyclical selection, complex dynamic patterns were revealed including "supercycles" (with periods comprising hundreds of environmental oscillation periods) and "deterministic chaos" (Kirzhner *et al*., 1996, 1998; Korol *et al*., 1998; Ryndin *et al*., 2001). These patterns could substantiate polymorphism and increase the range of temporal variation of allele frequencies. We believe that this previously uncharacterized evolutionary mechanism may increase genetic diversity over long-term periods and contribute to overcoming local and regional extinctions.

8. CENTER OF ORIGIN AND DIVERSITY

The center of origin and diversity of wild emmer and wild barley, the progenitor of most wheat and barley, and that of other progenitors of cultivated plants, e.g., wild oats (Kahler *et al.,* 1980; Somersalo *et al*., 1998; Zohary and Hopf, 2000) is the Near East Fertile Crescent. Particularly in Israel, with its extraordinary biotic and physical diversity, wild cereals developed, both within and between populations, a wide range of adaptive diversity for resistance against multiple diseases, pests, and ecological stresses over a long evolutionary history. Most importantly, this diversity is neither random nor neutral. By contrast, it displays adaptive genetic diversity at all levels for genetic biochemical, morphological, and immunological characteristics, which contribute to the species' ability to adapt to widely variable climatic and edaphic conditions with diverse and complex fitness syndromes. The long-lasting coevolution of wild cereals with parasites and with ecologically heterogeneous abiotic nature of Israel in particular, and the Fertile Crescent in general, led to the development of single *multi-allelic* genes and *multilocus* structures embracing both *coding* and *noncoding* genomic regions *locally* and *regionally* coadapted for both *short-* and *long-term* survival (Nevo, 1992, 2001a; Nevo *et al*., 2002).

9. GENETIC RESOURCES

Wild cereals harbor rich genetic resources and are the best hope for cereal improvement. The desirable traits that can be transferred from wild barley and wild emmer wheat to cultivated barley and wheat include the following: (i) resistance to a variety of *abiotic* (e.g., drought, cold, heat, and salt tolerances) and biotic (viral, bacterial, fungal, and herbicide resistances) stresses, (ii) high-quantity and –quality storage proteins (glutenins, gliadins, and hordeins), (iii) differential richness of amino acids, (iv) amylase and photosynthetic yield (Nevo, 1992, 2001c; Nevo *et al*., 2002). Genes for most of these traits available in wild barley and wild emmer wheat are still largely untapped and provide potential precious sources for cereal improvement.

Domestication-related quantitative trait loci (QTL) have been recently mapped in *T. dicoccoides* (Peng *et al*., 2000, 2003) and include brittle rachis, heading date, plant height, grain size, yield, and yield components. We mapped 549 molecular markers and approximately 70 domestication QTL effects, *nonrandomly* distributed among and along chromosomes. Seven domestication syndrome factors were proposed, each affecting 5-11 traits. The following important observations were made with respect to

domestication-related QTLs in wild emmer wheat. (*i*) Clustering and strong effects of some QTLs; (*ii*) Remarkable genomic association of strong domestication-related QTLs with gene-rich regions; and (iii) Unexpected predominance of QTL effects in the A genome. Similar genetic mapping was conducted in wild barley (Chen *et al*., 2004). The cryptic beneficial alleles at specific QTLs derived from wild cereals may contribute to cereal improvement (Fig. 1 in Peng *et al*., 2003). These beneficial alleles could be introduced into cultivated cereals (simultaneously eliminating agronomically undesirable alleles) by using the strategy of marker-assisted selection. Wild cereals harbor very valuable wild germplasm resource for future cereal improvement.

10. SUMMARY AND OUTLOOK

The studies conducted at the Institute of Evolution, University of Haifa, during 1975-2004 and briefly summarized in this chapter, demonstrated genetic differentiation *within* and *between* populations of wild barley and wild emmer wheat. This genetic differentiation is reflected both at the protein (isozyme) and DNA (molecular markers) levels and is largely adaptive in nature, driven by selective evolutionary forces. Conceptually, in-depth probing of comparative genome structure and function are the major challenges, in particular, the intimate relationship of the *coding* and *noncoding* (largely regulatory) genomes. Such studies will unravel the regulation of genome evolution and highlight the rich genetic potentials for cereal improvement residing in wild cereals. In particular, the following areas deserve major attention: molecular cloning of adaptation genes based on integrated genomic strategies and novel methodologies including genetic and physical mapping of molecular markers and expressed sequence tags (EST), sequencing of gene rich regions of cereal genomes, microarray expression analysis, retrotransposons (Kalendar *et al*., 2000), and genetic transformation involving defined target genes/alleles. The population subdivision reviewed here should be an important guideline in exploring genome structure and dynamics.

11. ACKNOWLEDGMENTS

I thank all my colleagues and collaborators for their long-term dedicated collaborations. I thank the Ancell-Teicher Foundation for Molecular Genetics and Evolution for financial support. I also wish to acknowledge the financial support of the Wolfson Foundation, the United States-Israel Agricultural

Research and Development Fund, BARD; the United States-Israel Binational Science Foundation, BSF; Israel Science Foundation; Israeli Ministries of Science and Agriculture and the Israeli Plant Gene Bank; The Humana Foundation: and the German-Israeli Cooperation Project (DIP project funded by the BMBF and supported by BMBF's International Bureau at the DLR.

12. REFERENCES

Badr A., Müller K., Schafer-Pregl R., El Rabey H., Effgen S., Ibrahim H.H., Pozzi C., Rohde W., Salamini F. On the origin and domestication history of barley. Mol Biol Evol 2000; 17: 499-510

Baek H.J., Beharav A., Nevo E. Ecological-genomic diversity of microsatellites in wild barley, *Hordeum spontaneum*, populations in Jordan. Theor Appl Genet2003; 106: 397-410

Baum B.R., Nevo E., Johnson D.A., Beiles A. Genetic diversity in wild barley (*Hordeum spontaneum* Koch) in the Near East: a molecular analysis using random amplified polymorphic DNA (RAPD). Genet Resou Crop Evol 1997; 44: 147-157

Brown A.H.D., Feldman M.W., Nevo E. Multilocus structure of natural populations of *Hordeum spontaneum*. Genetics 1980; 96: 523-536

Chalmers K.J., Waugh R., Waters J., Forster B.P., Nevo E., Powell W. Grain isozyme and ribosomal DNA variability in *Hordeum spontaneum* populations from Israel. Theor Appl Genet 1992; 84: 313-322

Chen G., Krugman T., Fahima T., Chen K., Hu Y., Röder M. S., Korol A.B., Nevo E. Chromosomal regions controlling seedling drought resistance of Israeli wild barley, *Hordeum spontaneum,* populations. Theor Appl Genet 2004 (submitted)

Crow J. Measuring population subdivision. In: S.P. Wasser (ed.) *Evolutionary Theory and Processes: New Horizons.* Kluwer Academic Press, Dordrecht, The Netherlands, 2003, pp. 35-42

Fahima T., Röder M.S., Wendehake K., Kirzhner V.M., Nevo E. Microsatellite polymorphism in natural populations of wild emmer wheat, *Triticum dicoccoides*, in Israel. Theor Appl Genet 2002; 104: 17-29

Fahima T., Sun G.L., Beharav A., Krugman T., Beiles A., Nevo E. RAPD polymorphism of wild emmer wheat population, *Triticum dicoccoides*, in Israel. Theor Appl Genet 1999; 98: 434-447

Golenberg E.M., Nevo E. Multilocus differentiation and population structure in a selfer, wild emmer wheat, *Triticum dicoccoides*. Heredity 1987; 58: 451-456

Gupta P.K., Sharma S., Kumar S., Balyan H.S., Beharav A., Nevo E. Adaptive ribosomal DNA polymorphism in wild barley at a mosaic microsite in Newe Ya'ar, Israel. Plant Science 2004, DOI:10.1016/j.plantsci2004.02.017

Harlan J. R., Zohary D. Distribution of wild emmer wheat and barley. Science 1966; 153: 1074-1080

Huang A., Beharav A., Li Y.C., Kirzhner V.M., Nevo E. Mosaic microecological differential stress causes adaptive microsatellite divergence in wild barley*, Hordeum spontaneum*, at Newe Ya'ar, Israel. Genome 2002; 45: 1-14

Huang X.Q., Börner A., Röder M.S., Ganal M.W. Assessing genetic diversity of wheat (*Triticum aestivum* L.) germplasm using microsatellite markers. Theor Appl Genet 2002; 105: 699-707

Ivandic V., Hackett C.A., Nevo E., Keith R., Thomas W.T.B., Forster B.P. Analysis of simple sequence repeats (SSRs) in wild barley from the Fertile Crescent: associations with ecology, geography and flowering time. Plant Mol Biol 2002; 48: 511-527

Kahler A.L., Allard R.W., Krzakowa M., Wehrhahn C.R., Nevo E. Associations between isozyme phenotypes and environment in the slender wild oat (*Avena barbata*) in Israel. Theor Appl Genet 1980; 56: 31-47

Kalendar R., Tanskanen J., Immonen S., Nevo E., Schulman A. Genome evolution of wild barley (*Hordeum spontaneum)* by BARE-1 retrotransposon dynamics in response to sharp microclimatic divergence. Proc Natl Acad Sci USA 2000; 97: 6603-6607

Kirzhner V.M., Korol A.B., Ronin Y.I., Nevo E. Genetic supercycles caused by cyclical selection. Proc Natl Acad Sci USA 1995; 92: 7130-7133

Kirzhner V.M., Korol A.B., Nevo E. Complex dynamics of multilocus systems subjected to cyclical selection. Proc Natl Acad Sci USA 1996; 93: 6532-6535

Kirzhner V.M., Korol A.B., Nevo E. Complex limiting behaviour of multilocus genetic systems in cyclical environments. Jour Theor Biol 1998; 190: 215-225

Kirzhner V.M., Korol A.B., Nevo E. Abundant multilocus polymorphisms caused by genetic interaction between species on trait-for-trait bases. Jour Theor Biol 1999; 198: 61-70

Kirzhner V.M., Korol A.B., Ronin Y.I., Nevo E. Genetic supercycles caused by cyclical selection. Proc Natl Acad Sci USA1995; 92: 7130-7133

Korol A.B., Kirzhner V.M., Nevo E. Dynamics of recombination modifiers caused by cyclical selection: Interaction of forced and auto-oscillations. Genet Res 1998; 72:135-148

Korol A.B., Kirzhner V.M., Ronin Y.I., Nevo E. Cyclical environmental changes as a factor of maintaining genetic polymorphism. II. Diploid selection for an additive trait. Evolution 1996; 50: 1432-1441

Korol A.B., Preygel I.A., Preygel S.A. *Recombination Variability and Evolution.* Chapman and Hall, London, England, 1994

Lev-Yadun S., Gopher A., Abbo S. The cradle of agriculture. Science 2000; 288:1602-1603

Li Y.C., Fahima T., Beiles A., Korol A.B., Nevo E. Microclimatic stress and adaptive DNA differentiation in wild emmer wheat (*Triticum dicoccoides*). Theor Appl Genet 1999; 98: 873-883

Li Y.C., Fahima T., Korol A.B., Peng J., Röder M.S., Kirzhner V.M., Beiles A., Nevo E. Microsatellite diversity correlated with ecological-edaphic and genetic factors in three microsites of wild emmer wheat in North Israel. Mol Biol Evol 2000b; 17: 851-862

Li Y.C., Fahima T., Krugman T., Beiles A., Röder M.S., Korol A.B., Nevo E. Parallel microgeographic patterns of genetic diversity and divergence revealed by allozyme, RAPD, and microsatellites in *Triticum dicoccoides* at Ammiad, Israel. Cons Genet 2000c; 1: 191-207

Li Y.C., Fahima T., Peng J.H., Röder M.S., Kirzhner V.M., Beiles A., Korol A.B., Nevo E. Edaphic microsatellite DNA divergence in wild emmer wheat, *Triticum dicoccoides* at a microsite, Tabigha, Israel. Theor Appl Genet 2000d; 101: 1029-1038

Li Y.C., Fahima T., Röder M.S., Kirzhner V.M., Beiles A., Korol A.B., Nevo E. Genetic effects on microsatellite diversity in wild emmer wheat (*Triticum dicoccoides*) at the Yehudiyya microsite, Israel. Heredity 2003; 90: 150-156

Li Y.C., Korol A.B., Fahima T., Beiles A., Nevo E. Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. Mol Ecol 2002a; 11: 2453-2465

Li Y.C., Korol A.B., Fahima T., Nevo E.. Microsatellites within genes: Structure, function, and evolution Mol Biol Evol 2004 (in press).

Li Y.C., Krugman T., Fahima T., Beiles A., Korol A.B., Nevo E. Spatiotemporal allozyme divergence caused by aridity stress in natural population of wild emmer wheat, *Triticum dicoccoides*, at the Ammiad microsite. Theor Appl Genet 2001; 102: 853-864

Li Y.C., Röder M.S., Fahima T., Kirzhner V.M., Beiles A., Korol A.B., Nevo E. Natural selection causing microsatellite divergence in wild emmer wheat at the ecologically variable microsite at Ammiad, Israel. Theor Appl Genet 2000a; 100: 985-999

Li Y.C., Röder M.S., Fahima T., Kirzhner V.M., Beiles A., Korol A.B., Nevo E. Climatic effects on microsatellite diversity in wild emmer wheat (*Triticum dicoccoides*) at the Yehudiyya microsite, Israel. Heredity 2002b; 89: 127-132

Marshall D.R., Brown A.H.D. Optimum sampling strategies in genetic conservation. In: O.H Frankel, J.G. Hawles (eds.) *Crop Genetic Resources for Today and Tomorrow*. Cambridge University Press, Cambridge, UK, 1975, pp. 53-70

Maruyama T. Effective number of alleles in a subdivided population. Theor Pop Biol 1970; 1: 273-306

Maruyama T. Stochastic Problems in Population Genetics. *Lecture Notes in Biomathematics*, Springer-Verlag, Heidelberg, Germany, 1977; 17: 1-245

Meyer R.C., Hackett C.A., Forster B.P., Russell J.R., Lawrence P., Young G.R., Booth A., Beharav A., Nevo E., Powell W. Fine-scale analysis of "Evolution Canyon" wild barley based on nuclear and choloroplast microsatellite variation. Mol Ecol 2004, submitted

Morgante M., Hanafey M., Powell W. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. Nature Genet 2002; 30: 194-200

Nei M. Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci USA, 1973; 70: 3321-3323

Nei M. F-statistics and analysis of gene diversity in subdivided populations. Ann Hum Genet 1977; 41: 225-233

Nei M. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 1978; 89: 583-590

Nevo E. Genetic variation in natural populations: Patterns and theory. Theor Pop Biol 1978; 13: 121-177

Nevo E. Genetic resources of wild cereals and crop improvement: Israel, a natural laboratory. Israel Jour Bot 1986; 35: 255-278

Nevo E. Plant genetic resources: Prediction by isozyme markers and ecology. In: M. Rattazi, J. Scandalios, G.S. Whitt (eds.) *Isozymes: Current Topics in Biological Research Agriculture, Physiology and Medicine,* Alan R Liss Inc, New York, USA, 1987; 16: 247-267

Nevo E. Genetic diversity in nature: Patterns and theory. Evol Biol 1988; 23: 217-247

Nevo E. "Origin, evolution, population genetics and resources for breeding of wild barley, *Hordeum spontaneum*, in the Fertile Crescent". In: P. Shewry (ed.) *Barley: Genetics, Molecular Biology and Biotechnology.* C.A.B. International, Wallingford, UK, 1992, pp. 19-43

Nevo E. Asian, African and European biota meet at "Evolution Canyon", Israel: Local tests of global biodiversity and genetic diversity patterns. Proc Royal Soc Lond 1995; 262: 149-155

Nevo E. Evolution in action across phylogeny caused by microclimatic stresses at "Evolution Canyon". Theor Pop Biol 1997; 52: 231-243

Nevo E. Molecular evolution and ecological stress at global, regional and local scales: The Israeli perspective. Jour Exp Zool 1998a; 282: 95-119

Nevo E. Genetic diversity in wild cereals: Regional and local studies and their bearing on conservation *ex situ* and *in situ.* Genet Resou Crop Evol 1998b; 45:355-370

Nevo E. Evolution of genome-phenome diversity under environmental stress. Proc Natl Acad Sci USA, 2001a; 98: 6233-6240

Nevo E. Genetic Diversity. In: *Encyclopedia of Biodiversity.* Academic Press, London, UK, 2001b; 3: 195-213

Nevo E. Genetic resources of wild emmer, *Triticum dicoccoides*, for wheat improvement in the third millennium. Israel Jour Plant Sci 2001c; 49: 77-91

Nevo E. Genomic diversity in nature and domestication. In: R. Henry (ed.) *Diversity and Evolution of Plants. Genotypic and Phenotypic Variation in Higher Plants,* CAB International, Wallingford, UK, 2004 (in press)

Nevo E., Baum B.R., Beiles A., Johnson D.A. Ecological correlates of RAPD DNA diversity of wild barley, *Hordeum spontaneum*, in the Fertile Crescent. Genet Resou Crop Evol 1998; 45: 151-159

Nevo E., Beharav A., Meyer R.C., Hackett C.A., Forster B.P., Russell J.R., Handley L., Powell W. Genomic mirosatellite adaptive divergence of wild barley by microclimatic stress in "Evolution Canyon", Israel. Biol Jour Linn Soci 2004, submitted

Nevo E., Beiles A. Genetic parallelism of protein polymorphism in nature: Ecological test of the neutral theory of molecular evolution. Biol Jour Linn Soc 1988; 35: 229-245

Nevo E., Beiles A. Genetic diversity of wild emmer wheat in Israel and Turkey: Structure, evolution, and application in breeding. Theor Appl Genet 1989; 77: 421-455

Nevo E., Beiles A., Ben-Shlomo R. The evolutionary significance of genetic diversity: Ecological, demographic and life history correlates. *Lecture Notes in Biomathematics*, Springer-Verlag, Heidelberg, Germany, 1984; 53: 13-213

Nevo E., Beiles A., Kaplan D. Genetic diversity and environmental associations of wild emmer wheat, in Turkey. Heredity 1986b; 61: 31-45

Nevo E., Beiles A., Kaplan D., Golenberg E.M., Olsvig-Whittaker L., Naveh Z. Natural selection of allozyme polymorphisms: A microsite test revealing ecological genetic differentiation in wild barley. Evolution 1986c; 40: 13-20

Nevo E., Beiles A., Krugman T. Natural selection of allozyme polymorphisms: a microgeographic climatic differentiation in wild emmer wheat, *Triticum dicoccoides*. Theor Appl Genet 1988a, 75: 529-538

Nevo E., Beiles A., Krugman T. Natural selection of allozyme polymorphisms: a microgeographic differentiation by edaphic, topographical and temporal factors in wild emmer wheat *Triticum dicoccoides.* Theor Appl Genet 1988b; 76: 737-752

Nevo E., Beiles A., Storch N., Doll H., Andersen B. Microgeographic edaphic differentiation in hordein polymorphisms of wild barley. Theor Appl Genet1983; 64: 123-132

Nevo E., Beiles A., Zohary D. Genetic resources of wild barley in the Near East: Structure, evolution and application in breeding. Biol Jour Linn Soc 1986a; 27: 355-380

Nevo E., Brown A.H.D., Zohary D., Storch N., Beiles A. Microgeographic edaphic differentiation in allozyme polymorphisms of wild barley (*Hordeum spontaneum*, Poaceae). Plant Syst Evol 1981; 138: 287-292

Nevo E., Golenberg E.M., Beiles A., Brown A.H.D., Zohary D. Genetic diversity and environmental associations of wild wheat, *Triticum dicoccoides*, in Israel. Theor Appl Genet1982; 62: 241-254

Nevo E., Korol A.B., Beiles A., Fahima T. *Evolution of Wild Emmer and Wheat Improvement. Population Genetics, Genetics Resources, and Genome Organization of Wheat's Progenitor*, *Triticum dicoccoides* Springer-Verlag, Heidelberg, Germany, 2002

Nevo E., Krugman T., Beiles A. Edaphic natural selection of allozyme polymorphisms in *Aegilops peregrina* at a Galilee microsite in Israel. Heredity 1994; 72: 109-112

Nevo E., Kirzhner V.M., Beiles A., Korol A.B. Selection versus random drift: Long-term polymophism persistence in small populations (evidence and modeling). Phil Trans Royal Soc B. 1997; 352: 381-389

Nevo E., Nishikawa K., Furuta Y., Gonokami Y., Beiles A. Genetic polymorphisms of alpha- and beta-amylase isozymes in wild emmer wheat, *Triticum dicoccoides*, Theor Appl Genet 1993; 85: 1029-1042

Nevo E., Noy-Meir I., Beiles A., Krugman R., Agami M. Natural selection of allozyme polymorphisms: A micro-geographical spatial and temporal ecological differentiations in wild emmer wheat. Israel Jour Bot 1991; 40: 419-449

Nevo E., Payne P.I. Wheat storage proteins: Diversity of HMW glutenin subunits in wild emmer from Israel. Geographical patterns and ecological predictability. Theor Appl Genet 1987; 74: 827-836

Owuor E.D., Beharav A., Fahima T., Kirzhner V.M., Korol A.B., Nevo E. Microscale ecological stress causes RAPD molecular selection in wild barley, Newe Ya'ar microsite, Israel. Genet Resou Crop Evol 2003; 50: 213-224

Owuor E.D., Fahima T., Beiles A., Korol A.B., Nevo E. Population genetics response to microsite ecological stress in wild barley, *Hordeum spontaneum.* Mol Ecol 1997; 6: 1177-1187

Owuor E.D., Fahima T., Beharav A., Korol A.B., Nevo E. RAPD divergence caused by microsite edaphic selection in wild barley. Genetica 1999; 105: 177-192

Pakniyat H., Powell W., Baird E., Handley L.L., Robinson D., Scrimgeour C.M., Nevo E., Hackett C.A., Caligari, P.D., Forster, B.P. AFLP variation in wild barley (*Hordeum spontaneum* C. Koch) with reference to salt tolerance and associated ecogeography. Genome 1997; 40: 332-341

Pavlíček T., Sharon D., Kravchenko V., Saaroni H., Nevo E. Microclimatic interslope differences underlying biodiversity contrasts in "Evolution Canyon", Mt. Carmel, Israel. Israel Jour Earth Sci 2003; 52: 1-9

Peng J.H., Korol A.B., Fahima T., Röder M.S., Ronin Y.I., Li Y.C., Nevo E. Molecular genetic maps in wild emmer wheat, *Triticum dicoccoides*: Genome-wide coverage, massive negative interference, and putative quasi-linkage. Genome Res. 2000; 10: 1509-1531

Peng J.H., Ronin Y. I., Fahima T., Röder M.S., Li Y.C., Nevo E., Korol A.B. Domestication quantitative trait loci in *Triticum dicoccoides,* the progenitor of wheat. Proc Natl Acad Sci USA 2003; 100: 2489-2494

Ryndin A., Kirzhner V.M., Nevo E., Korol A.B. Polymorphism maintenance in populations with mixed random mating and apomixis subjected to stabilizing and cyclical selection. Jour Theor Biol 2001; 212: 169-181

Salamini F., Ozkan H., Brandolini A., Schafer-Pregl R., Martin W. Genetics and geography of wild cereal domestication in the Near East. Nature Rev Genet 2002; 3: 429-441

Sharma S., Beharav A., Balyan H.S., Nevo E., Gupta P.K. Ribosomal DNA polymorphism and its association with geographical and climatic variables in 27 wild barley populations from Jordan. Plant Science 2004; 166:467-477

Somersalo S., Makela A., Rajala A., Nevo E., Peltonen-Sainio P. Morphophysiological traits characterizing environmental adaptation of *Avena barbata*. Euphytica 1998; 99: 213-220

Turpeinen T., Tenhola T., Manninen O., Nevo E., Nissial E. Microsatellite diversity associated with ecological factors in *Hordeum spontaneum* populations in Israel. Mol Ecol 2001; 10: 1577-1591

Turpeinen T., Vanhala T., Nevo E., Nissilä E. AFLP genetic polymorphism in wild barley (*Hordeum spontaneum*) populations in Israel. Theor Appl Genet 2003; 106:1333-1339

Volis S., Mendlinger S., Ward D. Differentation in populations of *Hordeum spontaneum* along a gradient of environmental productivity and predictability: life history and local adaptation. Biol Jour Linn Soc 2002; 77: 479-490

Volis S., Yakubov B., Shulgina E., Ward D., Zur V., Mendlinger S. Test for adaptive RAPD variation in population genetic structure of wild barley, *Hordeum spontaneum* Koch. Biol Jour Linn Soc 2001; 74: 289-303

Wright S. The genetical structure of populations. In: N.E. Morton (ed.) *Proc Conf, University Press of Hawaii*, Honolulu, Hawaii, 1973, pp. 313

Zohary D., Hopf J. *Domestication of Plants in the Old World*, 3rd edn. Oxford University Press, Oxford, UK, 2000

Chapter 7

GENE AND GENOME CHANGES DURING DOMESTICATION OF CEREALS

Carlo Pozzi[1], Laura Rossini[2], Alberto Vecchietti[1] and Francesco Salamini[3,*]

[1]Fondazione Parco Tecnologico Padano, via Haussmann 7, 26900 Lodi, Italy; [2]Università degli Studi di Milano, Dipartimento di Produzione Vegetale, via Celoria 2, 20133 Milan, Italy; and [3]Max-Planck Institute für Züchtungsforshung, Carl von Linnè weg 10, 50829 Cologne, Germany.
[*]Author for correspondence: salamini@mpiz-koeln.mpg.de

1. INTRODUCTION

Domestication is ".. a coevolutionary process by which human selection on the phenotypes of ... plant populations results in changes in the population's genotypes that makes them more useful to humans and better adapted to human intervention" (Clement, 1999). The complex of morphological, physiological and genetic changes that occurred during domestication was described as the "domestication syndrome" (Harlan, 1976). The domestication of grasses began during "the Neolithic revolution", about 12,000 years before present (BP), when a group of humans previously living as hunter-gatherers, became sedentary food-producers, pressed by a dry and cold climate episode (Diamond, 1997). The domestication process started with the conversion of wild grasses (characterized by small and naturally dispersed seeds) into cereals, which were day-length insensitive and produced large, non-shattering seeds necessitating human planting and harvesting. Several traits of the spikelet were modified. For example, reduced bracts and bristles and long involucral pedicel were selected, which contributed to the non-shedding habit. In addition, plants capable of maturing regardless of day-length (day-length insensitive plants) were selected to fully exploit incident solar radiation in the recently colonized temperate latitudes (Harper, 1977).

P.K. Gupta and R.K. Varshney (eds.), Cereal Genomics, 165-198.

In genetic terms, domestication of the grasses involved not only the selection against major genes that confer photoperiodic flowering, but also the accumulation of quantitative trait loci (QTLs) with small genetic effects that, collectively, confer substantial reductions in height, and smaller but significant changes in tillering, inflorescence and plant architecture. Domestication traits were probably selected as a unique set (Buckler *et al.*, 2001), thus allowing a genetically "convergent" domestication across grasses, as demonstrated for maize, rice and sorghum (Paterson *et al.*, 1995). This suggests that homologous, if not orthologous, genes/alleles have been involved in the evolution of domesticated grasses (Ladizinsky, 1985).

Classical genetics, molecular genetics, and genomics all contribute to our understanding of domestication. Genome mapping has often been carried out in the progeny of crosses between individual crop varieties and molecularly distant wild relatives, allowing the analysis of traits, which distinguish cultivated forms from their wild ancestors. Molecular maps, moreover, were exploited to compare some of the present-day crops with distantly related taxa, showing that only few major chromosomal rearrangements often distinguish among many crops. The possibility to map and clone QTLs (Doebley *et al.*, 1997; Frary *et al.*, 2000; Yano *et al.*, 2000) has allowed us to draw the following conclusions about domestication: (i) that the domestication syndrome originated by "sudden" genetic events, controlled by few major pleiotropic genes (Li *et al.* 1995; Lin *et al.*, 1995; Grandillo *et al*., 1999; Paterson, 2002), followed by the accumulation of minor mutations (Barton and Keightley, 2002); (ii) that a multifactorial, although not necessarily highly polygenic mode of inheritance characterizes domestication traits, which behave as "threshold traits": above-threshold changes in these traits are required to shift the developmental program. The latter hypothesis is vindicated by the observation that in natural populations, genetic variation sometimes exists without phenotypic variation (cryptic genetic variation; Lauter and Doebley, 2002).

2. THE TRAITS AND GENES THAT MADE DOMESTICATION POSSIBLE

A study of the levels and patterns of linkage disequilibrium (Morgante and Salamini, 2003), enables the detection of domestication related molecular footprints, suggesting that the genomic regions responsible for genetic variation are greatly reduced. Only a few hundred effective meiotic cycles (those resulting from hybridisation events taking place in nature or intentionally carried out by plant breeders) may have occurred during the

few-thousand-years of the domestication (Paterson, 2002). This is the basis to infer that chromosomal regions, which are 2-3 cM in length, are likely to remain in linkage disequilibrium (Paterson, 2002; Rafalski, 2002), suggesting that mutant loci with large effects on domestication traits may have been quickly fixed early in the domestication process. In such a situation, small DNA regions flanking domestication-related loci are now also characterized by low levels of diversity (linkage drag). This genetic phenomenon also explains the clustering of domestication genes. If domestication QTLs, in distant taxa, map to syntenic locations of the genome more often than would be expected by chance, such a finding would strongly suggest that corresponding genes were involved in the evolution of the phenotypes. The functional correspondence between non-syntenic genetic loci that have evolved in independent genomes, may in part be explained by chromosomal duplications within taxa (as hypothesized for maize; Paterson *et al.*, 1995).

Bioinformatic analyses of available data also indicate that domestication may confer a selective advantage to the sudden evolution of groups of tightly linked genes (Le Thierry d'Ennequin *et al.*, 1999), or "domestication complexes" (Lin *et al.*, 1995). The strong linkage of domestication factors facilitates the maintenance of the phenotypic identity of wild and cultivated populations coexisting in the same area, as observed for rice (Cai and Morishima, 2002). Alternatively, the clustering of genes into groups with relevance to plant adaptation can be explained by the effects of balancing selection in large natural populations; in this perspective, clustering of genes predisposed some crops to more rapid domestication, while the subsequent domestication fixed useful mutations within the cluster (Lin *et al.*, 1995). Examples of gene clustering are found in sorghum, rice and maize, where domestication loci in one crop largely correspond to those in the other crops (Paterson *et al.*, 1995): three QTLs affecting seed size correspond in the three species; for grain shattering, QTLs map to a single locus in sorghum, three loci in rice and 10 loci in maize; the sorghum single locus corresponds to a single rice QTL and to two maize QTLs located on duplicated chromosomal regions (Doebley and Stec, 1993).

The capacity of several domesticated grasses to flower in the long days of the temperate summer may be the result of mutations at a single ancestral locus, which is reflected in the observation that QTLs affecting flowering time and photoperiod also show correspondence between maize, sorghum, barley and wheat (Lin *et al.*, 1995), transcending 65 million years of reproductive isolation.

In pearl millet, factors involved in domestication syndrome concentrated mainly on two linkage groups. A crucial role of these factors in the developmental control of the spikelet and in the process of domestication has been demonstrated in several other related crops (Poncet *et al.*, 2000; Poncet *et al.*, 2002). Each domestication-related allele for non-shedding florets, nude seeds, and long spike pedicel can result through mutations in one or two genes so that these may be multiple events of domestication.

QTLs involved in the domestication syndrome appear to be located in gene-rich regions, which correspond to hot spots of recombination (Peng *et al.*, 2003). The current technical limitation in cloning QTL hinders the possibility to ascertain if the molecular basis of the domestication traits are the same across species, but some indications seem to confirm it: for example, plant height and flowering time are affected by mutations of the wheat *RhtD1a*, rice *SLR1* and maize *D8*, which are orthologous to each other (Peng *et al.*, 1999; Ikeda *et al.*, 2001; Thornsberry *et al.*, 2001). A similar situation is found in rice, where the *MOC1* gene (belonging to the *GRAS* family of transcription factors) promotes tillering and has a negative effect on plant height. In *moc1* rice mutants, the expression of the *OsTb1* (the rice orthologue of the maize *TB1*) is significantly reduced (Xueyong *et al.*, 2003). Table 1 provides a partial list of genes and QTLs involved in the domestication process.

2.1. Wheat

Wheat is a good example of the ploidy changes that occurred during cereal evolution: the ploidy levels of domesticated species range from diploid ($2n$=14) to hexaploid ($6n$=42), and all ploidy levels are available in domesticated wheats.

2.1.1. Non- Brittle Rachis

A non-brittle rachis in the diploid einkorn wheat (genome A^mA^m, where "m" indicates that the genome derives from *T. monococcum*), two genes (segregating with a 15 brittle: 1 tough rachis ratio in the F_2 progeny of wild × domesticated crosses) control the trait "*disarticulating* (brittle) *rachis*" (Sharma and Waines, 1980). The two dominant alleles derive from the einkorn progenitor *T. boeoticum*, which has a brittle rachis. Also the progenitor of tetraploid (A^uA^uBB, where "u" indicates the *T. urartu* origin), and hexaploid (A^uA^uBBDD) wheats, the wild emmer wheat (*T. dicoccoides*) (Dvoràk *et al.*, 1998), has brittle ears, which shatter into spikelets at

Table 1. A partial list of genetic loci involved in the control of domestication traits

Traits	Crop	Loci[1]	QTL[2]	Reference
Seed size (weight)	Millet	-	(LG 2, 4)	Devos *et al.* (2000), Poncet *et al.* (2002)
	Wheat	-	(1A, 2A, 3A, 4A, 7A, 5B, 7B)	Elias *et al.* (1996)
Glumes softness (free threshing)	Wheat	*Sog* (2A), *Tg* (2D), *Q* (5A)	-	Muramatsu (1963, 1986), Cao *et al.* (1997), Kato *et al.* (1998), Luo *et al.* (2000), Taenzler *et al.* (2002)
	Maize	-	(1, 2, 3, 4, 5)	Doebley *et al.* (1990), Doebley and Stec (1991, 1993)
Rachis stiffness (shattering)	Rice	-	(1, 3, 4, 6, 8, 9, 11)	Xiong *et al.* (1999), Cai and Morishima (2002)
	Sorghum	-	(LG c)	Paterson *et al.* (1995)
	Barley	*bt1* (1), *bt2* (2)	-	Takahashi (1972)
	Maize	-	(1, 2, 4, 5)	Doebley *et al.* (1990), Doebley and Stec (1991, 1993)
	Millet		(LG 6)	Poncet *et al.* (2002)
	Wheat	*Q* (5A)	-	Sears (1976), Cao *et al.* (1997), Chen *et al.* (1998)
Number of cupules lacking the pedicellate spikelet	Maize	-	(1, 3, 4)	Doebley *et al.* (1990), Doebley and Stec (1991, 1993)
Number of cupules per rank	Maize	-	(1, 5)	Doebley *et al.* (1990), Doebley and Stec (1991, 1993)
Number of rows per cupule	Maize	-	(2, 3, 4, 5)	Doebley *et al.* (1990), Doebley and Stec (1991, 1993)
Percentage of male spikelets in primary lateral inflorescence	Maize	-	(1, 3)	Doebley *et al.* (1990), Doebley and Stec (1991, 1993)
Secondary branches per panicle	Rice	-	(1, 7)	Xiong *et al.* (1999)
Anther length	Rice	-	(1, 2, 3, 5, 6, 8, 9)	Xiong *et al.* (1999)
Culm diameter	Rice	-	(1, 6, 7, 8, 11)	Xiong *et al.* (1999)
Spikelet density	Rice	-	(1, 3)	Xiong *et al.* (1999)

Table 1. Continued

Panicle length	Rice	-	(1, 7)	Xiong *et al.* (1999)
Panicle neck length	Rice	-	(1, 3, 6, 9)	Xiong *et al.* (1999)
	Maize	-	(1, 2, 3, 4)	Doebley *et al.* (1990), Doebley and Stec (1991, 1993)
Plant height	Rice	-	(1, 8, 9)	Xiong *et al.* (1999)
	Maize	-	(1, 2, 3, 4, 5, 6, 7, 8, 9, 10)	Lin *et al.* (1995)
	Sorghum	-	(LG a, g)	Lin *et al.* (1995)
	Millet	-	(LG 7)	Devos *et al.* (2000), Poncet *et al.* (2002)
Number of tiller	Rice	-	(1, 4)	Xiong *et al.* (1999)
	Sorghum	-	(LG c, d, h, g, j)	Paterson *et al.* (1995)
	Maize	*TB1*(1)	-	Doebley *et al.* (1997)
Heading date	Rice	-	(3, 6, 8, 11)	Xiong *et al.* (1999)
	Maize	-	(1, 2, 3, 5, 6, 7, 8, 9, 10)	Lin *et al.* (1995)
	Sorghum	-	(LG a, d, g)	Lin *et al.* (1995)
Dormancy	Rice	-	(11, 2, 3, 4, 5, 6, 8, 9, 11, 12)	Xiong *et al.* (1999)

1The numbers in parentheses refer to the chromosomes of the indicated species
[2]The numbers refer to chromosomes and LG to the linkage groups of the indicated species; for abbreviations, see text

maturity. In both tetraploid and hexaploid wheats, rachis fragility is reported to depend from two or more genes (Salamini *et al.*, 2002). The genetics of this trait is complex, since the loci influencing this trait are pleiotropic, influencing both the softness of the glume and the toughness of the ear rachis, thus linking the latter trait to the free-threshing habit.

Tetraploid hard wheat (*T. durum*) and hexaploid bread wheat (*T. aestivum*) represent the free-threshing wheats. It is believed that sometimes in the past, a tetraploid hybridised with the diploid species *Aegilops tauschii* (DD), and generated the spelt-like hulled hexaploid wheats (McFadden and Sears, 1946; Zohary and Hopf, 2000). As the natural distribution areas of *T. dicoccoides* and *A. tauschii* do not overlap, and wild hexaploids are unknown, the emmer partner in this hybridisation is assumed to have been a cultivated land race, implicating that all hexaploid wheats have an origin dating after the agriculture came into practice (Salamini *et al.*, 2002).

2.1.2. *Q* Locus for Square-Head

A complex locus, in hexaploid wheats, known as *Q* factor, is the characteristic of the evolution of the soft glume, free-threshing wheats (Kato *et al.*, 1998); only square-headed wheats (due to *Q* locus) combine a good threshability, a convenient grain size and shape, tough rachis and other characters, which facilitate harvesting (Snape *et al.*, 1985; Sourdille *et al.*, 2000). Mutagenic disruption of the *Q* gene in *QQ* hexaploid free-threshing wheat results in tenacious glumes (hulled), long, lax, fragile and awned or awnless ears, which have the *qq* genotype (MacKey, 1954). *Qq* heterozygotes have intermediate phenotypes due to the interaction of two active alleles (MacKey, 1954; Muramatsu, 1963): a square-headed hexaploid ear, in fact, can derive from either two doses of *Q* or five doses of *q*, demonstrating that *q* is functional but is hypomorphic to *Q*. Most probably the *Q* allele emerged by mutation from hulled domesticated tetraploids, and mutations at additional loci shaped the full expression of the free-threshing characteristic (Muramatsu, 1986).

2.1.3. The Gene *Tg* for Glume Tenacity

A later discovery established that in the hexaploid wheats the partially dominant *Tg* allele (Kerber *et al.*, 1974) - contributed by the D genome (Villareal *et al.*, 1995) - also affects glume tenacity. Thus, the free threshing hexaploids have the *tgtgQQ* genotype, while the *Ae. tauschii* lines have the dominant *Tg* allele. When the free-threshing habit was genetically dissected in tetraploid wild × domesticated crosses, the F_2 population showed a continuous distribution for the trait (Simonetti *et al.*, 1999). Four major QTLs were located on three chromosomes, and two of them in syntenic positions with the genes *Tg* and *Q* of the hexaploid wheats, whereby the *Tg*-like gene *Tg2* is located on another chromosome (2B). Thus, the polygenic control of the free-threshing character in tetraploid wheats is based on loci other than *Q* and *Tg*. In the A^m genome of einkorn, the free-threshing character is inherited as a recessive allele (soft glume, *sog*, with *Sog* conditioning tenacious glumes). This gene maps on chromosome 2 on a position syntenic to *Tg* (Taenzler *et al.*, 2002). In *sog* plants, glumes are soft but the ear is fairly compact, a negative pleiotropic effect that has hindered a wide use of free-threshing einkorn varieties (in other wheats, the effect of genes supporting the free-threshing habit on ear compactness was somehow alleviated by the polyploid state of the genome).

2.1.4. Genes for Seed Size

Genetics of seed size has been studied by generating *T. dicoccoides* substitution lines in *T. durum* (Elias *et al.*, 1996); the trait has been shown to be under polygenic control, with alleles contributing to the modulation of kernel size mapping to several chromosomes (Cantrell and Joppa, 1991). Synteny between wheat and rice indicates that eight seed-size QTLs detected in wheat correspond in chromosomal location to their rice counterparts (Lin *et al.*, 1995). In *T. dicoccoides*, seed size is also under the control of the major gene *Br*, which maps with eight other major domestication-related QTLs, making it tempting to speculate that the gene conditions pleiotropically a large part of the domestication syndrome (Peng *et al.*, 2003). An extensive QTL search in *T. dicoccoides* led to identification of 70 QTLs related to domestication, some with strong effects and mostly localized in gene-rich regions (Peng *et al.*, 2003).

Genetic data for domestication traits in wheats can be summarized as follows: (i) several genes contribute to the tough rachis and the free-threshing phenotypes; (ii) some of these genes control both glume softness and rachis strength; (iii) alleles of these genes are frequently semi-dominant (which may have extended the conscious selection, carried out during domestication, to heterozygous phenotypes); (iv) *Q*-like alleles shorten the ear, which contributes to the manifestation of a mutant phenotype, again supporting its conscious selection.

2.2. Maize

2.2.1. Differences Between Maize and Teosinte

An entire set of "domestication" QTLs were mapped in two crosses involving different races of maize and different subspecies of its ancestor, teosinte (*Zea mays* spp. *parviglumis*): five genomic regions (containing about 50 QTLs) were found to control most of the differences between maize and teosinte (Doebley and Stec, 1993; Doebley *et al.*, 1997; Westerbergh and Doebley, 2002). Maize and teosinte differ morphologically, the wild species having fully shattering ears, a single spikelet per cupule, highly indurate glumes and long lateral branches bearing staminate inflorescences. The QTLs controlling the number of tassel branches identified in crosses between teosinte species, map in the vicinity of the following loci: (Westerbergh and Doebley, 2002): (i) the maize *ramosa 1* locus, which produces a larger number of tassel branches (Neuffer *et al.*, 1997); (ii) the maize homologue of the Arabidopsis *Leafy* gene, which

controls inflorescence branching (Kyozuka *et al.*, 1998); and (iii) *fascicled ear* locus, which causes an increase in the number of tassel branches (Neuffer *et al.*, 1997).

2.2.2. The Genes *tru1* and *te1*

Despite the fact that we do not have a good estimate of the number of genes involved in the domestication of maize, the genes "*tassel replaces upper ear"* (*tru1*) and "*terminal ear1"* (*te1*) have been identified as candidates for one of the five chromosomal regions hosting most of the QTLs that modulate the morphological differences between maize and teosinte (Doebley *et al.*, 1990; Doebley and Stec, 1993; Doebley *et al.*, 1995). The *te1* is a regulatory gene that encodes a protein with conserved RNA-binding domains: it may function through RNA-binding activity, affecting the pattern of internode initiation, and causing an increase in the feminization of the terminal inflorescence on the main stalk (Veit *et al.*, 1998).

2.2.3. The Genes *tb1* and *tga1*

Monogenic loci have been identified which have supposedly played an important role in the domestication process of maize (Doebley *et al.*, 1997). The early maize domesticators concentrated on a small number of pleiotropic, and independently assorting mutations, most of which were recessive. Two loci have prominent roles in coordinating the morphological differences between wild and domesticated maize: *teosinte branched 1* (*tb1*) and *teosinte glume architecture 1* (*tga1*) (Doebley *et al.*, 1993; Dorweiler and Doebley, 1997). The *tb1* locus encodes a putative transcriptional regulatory protein, controlling lateral branch growth and probably acting as a repressor of apical growth (a phenotype typical of the teosinte ancestor). During domestication, a dominant *tb1* allelic variant of *tb1* was selected that is expressed at higher level than in teosinte in primary axillary meristems, such that these form ear shoots rather than the elongated tassel-tipped branches, typical of teosinte (Doebley *et al.*, 1997; Wang *et al.*, 1999). Thus, domestication proceeded also by altering gene regulation. The gene *tb1* is an orthologue of the *Antirrhinum majus* gene *CYCLOIDEA* (Luo *et al.*, 1999), which is involved in repression of axial growth in flower development and has been implicated in the evolution of floral morphology. In a similar way, effects on several developmental processes are correlated with the *tga1* locus, which controls in teosinte and maize the development of the cupulate fruitcase and the degree of glume induration (Dorweiler and Doebley, 1997). Studies conducted on maize-teosinte and teosinte-teosinte

F_2 populations show that, in the latter, the variation explained by QTL at any one domestication locus is much lower than in the former cross. The explanation is that breeders may have selected for domesticated plants stable in different environments, while natural stands often respond to local conditions by developmental modifications produced under natural selection (Westerbergh and Doebley, 2002). Thus, the maize allele of the *tb1* gene is believed to be less plastic in responding to environment than is the teosinte allele (Lukens and Doebley, 1999). A complex network of gene interactions seems to be at work in coordinating the action of *tb1*: epistatic effects have been detected with at least one another QTL (which may act as un upstream regulator of *tb1*; Lukens and Doebley, 1999; Lauter and Doebley, 2002), shedding new light on the interaction between cryptic variation and major effect QTLs.

Action of discrete loci and major QTLs is probably insufficient to give rise to the entire domestication syndrome in maize. Archaeological evidence show that early traits selected during domestication included the induration of the rachis of the glumes, shortened and rigid rachis internodes, and non-shattering phenotype (Doebley *et al.*, 1993). However, at the time when human cultivators transformed teosinte into modern maize, genetic modifiers did not yet accumulate to stabilize the maize-like phenotype with paired spikelets (Benz, 2001).

2.3. Barley

The genetic and morphological events that accompanied the development of cultivated barley from *Hordeum spontaneum* included the change from brittle to non-brittle rachis, the transition from distichous to polystichous spike, and the appearance of the *naked caryopsis* (*nud*) character (Søgaard and von Wettstein-Knowles, 1987; Harlan, 1976). Varieties with naked seeds, however, have been traditionally cultivated only in the Himalayan region (Badr *et al.*, 2000).

2.3.1. The Genes *btr1* and *btr2, Vrs1* and *int-c*

The two complementary genes that control rachis brittleness are *btr1* and *btr2,* the domesticated barley being homozygous recessive for either or both loci. This was earlier taken as evidence that barley was domesticated at least twice (Takahashi, 1955), but the monophyletic origin of the crop is now well established based on molecular marker studies (Badr *et al.*, 2000). Six–rowed plants carry a recessive allele of the *vrs1* (*Kernel rows*) gene and a

dominant allele of the interacting gene *Lateral floret fertility* (*int-c*) (Ayoub *et al.*, 2002). *Vrs1* has been proposed to condition plant morphology through epistatic interactions with a number of closely linked QTLs (Ayoub, 2002).

2.3.2. Genes for Seed Dormancy/ Vernalization, Photoperiod Response and Dwarfing

Domesticated and wild varieties can be distinguished by the presence of alternative alleles at several QTLs for decreased seed dormancy, a character that secures the persistence of wild lines in nature (Ullrich *et al.*, 1993). Relatively high levels of dormancy in wild barley often prevent immediate germination, while time of flowering depends on the satisfaction of a high vernalization requirement and short-day photoperiod response. During domestication, selection against these characteristics resulted in photoperiod insensitive cultivars with reduced seed dormancy. The control of flowering time and height is strictly correlated to the genetic mechanisms that control plant development and growth, and thus to cell division and size. This correlation is evident in cultivars characterized by distinct dwarfing genes, such as the semi-dwarf cultivars possessing the *ari-e.GP* and those possessing the *sdw1* genes (Börner *et al.*, 1998; see later in this Chapter).

2.4. Sorghum and Millet

In sorghum, shattering of the mature inflorescence has monogenic inheritance (Paterson *et al.*, 1995a; Paterson *et al.*, 1995b; Young, 1986). This suggests that the African sorghum was probably rapidly domesticated, because of the need of a single mutation in a critical step leading to shattering (*Sh1* gene). Similarly, in pearl millet (*Pennisetum glaucum*), the loss of shedding ability is controlled by a single locus (*al6*; Poncet *et al.*, 2000) while, in America, maize selectors had to put together several mutations, each with smaller effects on several distinct steps, to achieve the same results.

In sorghum, the factors involved in domestication syndrome are concentrated on four linkage groups, but independent and random occurrence of mutations in related genes may have formed alleles with very different phenotypic consequences (Draye *et al*., 2001).

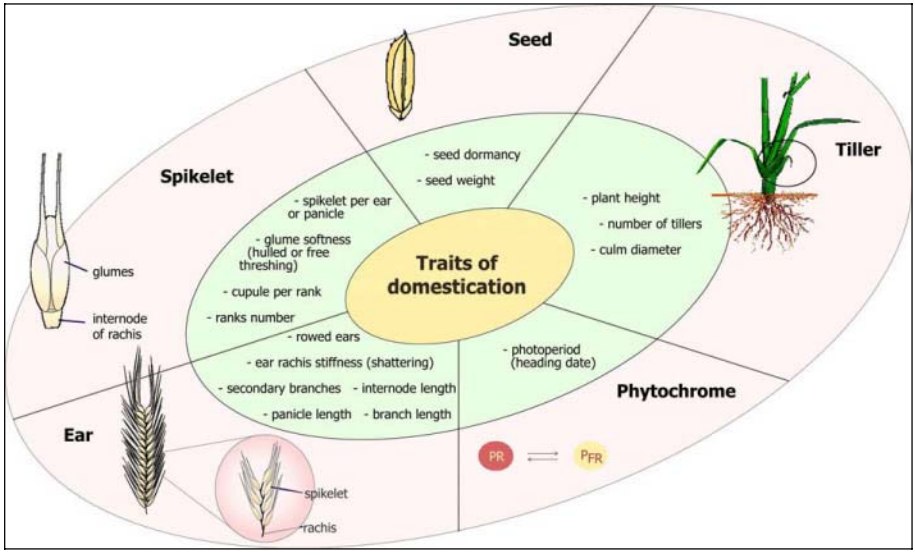


Figure 1. Some of the morphological differences that distinguish wild from domesticated cereals. Some of the major discriminating characters are listed below; refer to the Text for the others. Wild forms have small seeds, while domesticated forms have larger ones; the spikelet of wild ears fall apart at ripening through fragmentation of the rachis (shattering), whereas the domesticated forms have a tough rachis that holds the seeds together; the leaf-like glumes protecting the seed are attached tightly to the seed in wild forms, whereas they release the seed in the more advanced domesticated forms (naked forms).

2.5. Rice

Domestication-related differences between the wild *Oryza rufipogon* and cultivated rice (*Oryza sativa*) were investigated by Xiong *et al.* (1999). The authors demonstrated that 19 traits (including seven qualitative and 12 quantitative traits) are related to domestication and are concentrated in few chromosomal blocks. They investigated growth habit; plant height; panicle, anther and tiller number and culm circumference; heading date (photoperiodism); number of spikelets per panicle; spikelet density; shattering; number of secondary branches per panicle; panicle neck length; anthocyanin pigmentation (Figure 1). Each of the seven qualitative traits were controlled by a single Mendelian locus (for example the presence *vs.* absence of extruded stigmas, lax panicles and awns, were each accounted for by single loci). 72.4% of variation in plant height was explained by four QTLs, one of which coincided with the *sd1* locus previously reported to support dwarfism (Cho *et al.*, 1995), while *hd8* (Li *et al.*, 1995) explained more than half of the variation for heading date. Shattering was under the control of five QTLs. The Asian domesticators of rice consciously selected

for intermediate degrees of non-shattering that would reduce field losses but not preclude hand-threshing.

3. ORIGIN/LOSS OF DNA DIVERSITY DURING DOMESTICATION

Molecular genetics has the potential to assess precisely the genetic diversity between wild grasses and domesticated cereals, thus providing insights into the "monophyletic" *vs.* "polyphyletic" mode of domestication. In case of a single domestication event, the genetic diversity in the cultivated gene pool should be lower than that of wild populations, because of the inherent genetic bottleneck. Molecular analyses also help in identification of wild progenitors of domesticates and allow estimation of the time scale of domestication events.

An initial and fast domestication event should involve few linked genes which have pleiotropic effects and a strong selective advantage. Despite the initial population bottleneck, domesticated crops retain large proportions of the genetic diversity observed in their wild progenitors. The fact that grass seeds must have had to be consumed in large quantities, may suggest that several million plants must have constituted the first "domesticated" effective population (Tanksley and McCouch, 1997), and that the observed loss of diversity should involve genes for which alleles having a superior agronomic value were selected. For instance, the patterns of diversity at the maize domestication gene *tb1* are consistent with reasonably large effective populations that must have been maintained during domestication (Wang *et al.*, 1999). A second effect of the fixation of a particular domestication allele is a reduction of the nucleotide variability at linked loci. This reduction is a function of the selection level and the rate of recombination between the selected site and the linked loci being surveyed. This locus-specific bottleneck (hitchhiking effect), is witnessed as a skewed distribution of frequencies of molecular markers (Vigouroux *et al.*, 2002), when compared to the frequency distribution expected under an equilibrium-neutral model. In addition, the population bottleneck associated with domestication can cause a genome-wide loss of diversity that, for specific traits, could be misidentified as the signature of direct selection.

Grasses have undergone recent and strong selective sweeps targeted at phenotypes that improve agronomic performance, or qualities connected to their use as food. It should, thus, be possible to identify the genes controlling these phenotypes by mean of genomic scans for DNA regions showing the signature of selection (i.e. existence of discernible linkage

disequilibrium typical of genomic regions under selection) and highlighting a connected "hitchhiking" effect (Remington *et al.*, 2001). This procedure is often followed using association tests which identify genes that control agronomic traits, provided that functional variation of the investigated species (Thornsberry *et al.*, 2001) still exists in a wide collection of materials representing different eras of plant breeding. Such genomic scans help to create a more complete picture of how domestication and breeding have shaped crop genomes.

3.1. Wheat

The study of genetic diversity in wheat (*Triticum aestivum*) has proved useful in understanding the polyploidisation events that led to the formation of its hexaploid genome (AABBDD). This genome originated by allopolyploidisation, i.e. hybridisation, , between a tetraploid (AABB) and a diploid wheat (DD; most probably from *Ae. tauschii*) followed by chromosome doubling. Hexaploid wheats are much less diverse than their diploid progenitors, and substantial differences exist between the three genomes (Zohary and Feldman, 1962; Galili, 2000). No suitable explanation is, however available for the diversity in the A and B genomes of tetraploid although it may be speculated that external forces such as differential selection and interspecific introgression, and/or internal mechanisms involving processes like differential methylation, mutation or recombination might have played a role (Wendel, 2000). For example, the *A1* locus of the D genome has two very distinct haplotypes, which are found in both wheat and *Ae. tauschii* (Talbert *et al.*, 1998; Blake *et al.*, 1999), suggesting that the hexaploid wheat was derived from more than one polyploidisation events.

The importance of polyploidisation in domestication is still not fully understood. It has been speculated that mutation in genes critical for domestication in a diploid genome might have caused too extreme a phenotype to allow a fruitful use in domestication (Salamini *et al.*, 2002). It is also well established that allopolyploidisation is accompanied by rapid and non-random sequence elimination (Őzkan *et al.*, 2002). This genomic phenomenon seems to be involved in the increment of divergence of homoeologous chromosomes in order to provide the physical basis for the restriction of pairing to homologous chromosomes forming only bivalents in the newly formed allopolyploids (Shaked *et al.*, 2001; Őzkan *et al.*, 2002). This then perhaps contributed to disomic inheritance, full fertility and permanent heterosis between homoeoalleles, thus leading to the establishment of newly formed species in nature.

3.2. Maize

3.2.1. Reduced Diversity at Selected Loci

In maize, loci without significant allelic variation but which contribute to traits of human interest can be identified. For these loci, human selection during maize domestication and improvement has reduced variation in excess of the bottleneck effect. Molecular markers derived from expressed sequence tags (i.e. SSRs derived from EST) have been used in this crop, with the aim of identifying genomic regions that show evidence of selection during maize domestication or improvement (Vigouroux *et al.*, 2002). It is estimated that relative to wild species, there is an average of 30% drop in diversity, at any selected locus. The drop in diversity is substantially greater at genes involved in domestication, such as *tb1*, the promoter of which has 61-fold lower diversity in the crop than it does in the closest wild relative (Wang *et al.*, 1999). Interestingly, this drop in diversity does not extend for the entire length of the gene, as the coding region has levels of diversity similar (39%) to those at loci considered neutral for the process of domestication. These data allow a rough estimate of the selection coefficient associated with the fitness of the favourable allelic genotype during domestication, and with the duration needed to bring the maize allele to fixation: the fixation of *tb1*, for example, may have taken about 300-1000 years (depending on the population size considered; Eyre-Walker *et al.*, 1998).

Maize is an outcrossing species and the population on which selection was initially exercised, according to one of the hypotheses, consisted of about a million plants (Eyre-Walker *et al.*, 1998; Vigouroux *et al.*, 2002). This may explain the observed high level of recombination and the consequent low levels of linkage disequilibrium (Wang *et al.*, 1999). Moreover, in the presence of a relatively mild, domestication bottleneck, unselected genes should have retained high diversity and should be readily distinguished from those affected by selection (Eyre-Walker *et al.*, 1998).

The spread of maize from the highlands in central Mexico, where it was probably domesticated, to the lowlands (Matsuoka *et al.*, 2002) can be followed by analyzing the alleles derived from *Z. mays* spp. *parviglumis* (Matsuoka *et al.*, 2002) and those introgressed from ssp. *mexicana* (a highland teosinte species which forms frequent hybrids with maize; Wilkes, 1977). The complete interfertility of maize and teosinte suggests that time was not sufficient for substitution fixation at majority of the loci affecting yield and fitness, and only the loci strictly selected to confer agriculturally

important qualities to domesticated maize were fixed in this manner (Doebley and Stec, 1991).

The evolution of maize under artificial selection may imply that particular types of mutant alleles were favoured which may not be so favoured by natural selection in the wild (Haag and True, 2001). For other genes putatively involved in domestication processes, such as *terminal ear 1, te1* (Veit et al, 1998), no evidence of selection is available through haplotype analysis. However, the neutral pattern of evolution at *te1* was used to estimate the age of the maize gene pool (the time of divergence of the *Zea* and *Tripsacum* lineages) and the strength of the proposed domestication bottleneck (White and Doebley, 1999). It seems that *te1* was not involved in maize morphological evolution (White and Doebley, 1999), since there is no evidence for a reduction of nucleotide diversity in *te1* relative to other genes such as *adh1* and *adh2*.

Studies on the *dwarf8* gene suggest that it may affect the quantitative variation of maize flowering time and plant height (Thornsberry *et al.*, 2001). The low level polymorphisms in its sequence across several inbred lines also suggest that it has been a target of selection.

3.3. Barley

In barley, the patterns of diversity observed differ from that in other cereals. Studies conducted by Kahler and Allard (1981) indicate that there is a high level of parallel polymorphism between wild and cultivated barley. The persistence of wild alleles in the domesticated barley, may be interpreted as an indication of multiple domestication events, but could also be explained by gene flow (Ladizinsky and Genizi, 2001; even if outcrossing limits the possibility of describing this phenomenon). For instance, three major haplotypes for the homeotic *Barley Knotted-like3* gene (*Bkn-3;* Müller *et al.*, 1995) were exploited to follow the flow of germplasm from wild to domesticated lines of barley (Badr *et al.*, 2000). This study shows that haplotype I found for *Bkn-3* is rarely detected in wild populations except in the wild populations of Israel. The same haplotype is pervasive in domesticated western land races, which generated the cultivated Western gene pool. During the migration of domesticated barley eastwards from the Fertile Crescent towards the Himalayas, haplotype III replaced haplotype I and it is thus most prevalent in the Himalayas and in most Asian wild forms. A similar work was conducted in rice, where the homeotic gene *OSH3* (another *Knotted*-like gene) is putatively responsible for one of the traits

that are selectively introduced during the domestication of most of the *japonica* rice (Yutaka *et al.*, 2001).

3.4. Sorghum and Rice

Data on isozyme loci, which are generally not the target of human selection, suggest that *Sorghum bicolor* has about two-thirds of the diversity of its wild relative (Morden *et al.*, 1990). Surveys have been conducted also on the *Adh1* gene showing that the cultivated type has a modest diversity (0.24%), while this value rises to 0.36% in wild lines (Gaut and Clegg, 1993).

Research on the rice *Waxy* gene (which, when mutated, suppresses amylose formation and thus renders the caryopsis non-glutinous) shows a level of diversity that is lower than expected. This suggests that early domesticators of glutinous rice liked its adhesive quality and wanted to preserve that trait (Hirano *et al.*, 1998; Olsen and Purugganan, 2002). It is however, possible that male gametophytes recessive for the *waxy* allele have a disadvantage on an evolutionary timescale (Hirano *et al.*, 1998). The *Waxy* gene is an example of a gene with important enzymatic functions where a single mutation is able to alter its posttranscriptional regulation: this last study has clarified that a trait supposedly under quantitative genetic control may have molecular components allowing a simpler interpretation.

In rice, the partial-loss-of-function allele of *Heading date 1* (*Hd1*) is involved in the control of the photoperiodic response. Such an allele might have been selected in the process of breeding, because varieties that contain the weak *Hd1* allele show early heading and can be cultivated in a wide range of cultivation areas (Yano *et al.*, 2000).

4. WHEN AND WHERE: THE GENETICS AND THE GENOMICS OF THE TIMING AND LOCATION OF DOMESTICATION

Despite the fact that the common ancestor of major grass crops dates back to 55-70 million years (Ahn *et al.*, 1993; Kellogg, 2001), cereal domestication has a recent history of only about 5-10 thousand years. Domestication can be explained as the unintentional selection of favourable alleles *during* plant cultivation (Salamini *et al.*, 2002; Zohary and Hopf, 2000). Under this perspective, the apparently contrasting observations of the

clearly multi-regional harvest and use of wild species versus a putative restricted regionality of crop domestication (Salamini *et al*., 2002) can be reconciled: superior varieties of founder crops emerged in a core area (Lev-Yadun *et al*., 2000) and then moved throughout the region, displacing local wild or domesticated genotypes. This sort of considerations are supported by recent findings in pearl millet (Poncet *et al.*, 2002), where unconscious selection in the early evolution of the domesticates and gene flow from local wild populations account for the small number of differences observed between wild and domesticated gene pools.

The temporal extension of domestication is connected to the dimension of the founder population (Eyre-Walker *et al.*, 1998). For example, based on archaeological evidence, it is believed that einkorn domestication took place only a few centuries ago (Diamond, 1997). A similar rate of domestication *tempo* would suggest, in maize, a bottleneck population size of about 600 *Zea parviglumis* individuals, based on the genetic diversity at the gene *Adh1* (Eyre-Walker *et al.*, 1998).

The molecular data, while helping in determining the correct timing of domestication, allow also to trace - through genome-wide estimates of genetic similarity - a map of the domestication sites, which appear to be discrete and limited in number (Heun *et al.*, 1997; Badr *et al.*, 2000). Identification of the location can sometimes be very precise. For instance when the wild background contains several distinct genetic variants, and the domesticated forms have only one variant, it indicates a clear monophyletic origin of the domestication event (Zohary, 1999). Domestication locations (Figure 2) range from the Fertile Crescent (wheat, barley, oats and rye; Gopher *et al.*, 2002); Africa (for sorghum and millet); Asia (rice); and Central America (maize). The Fertile Crescent is the primary domestication centre for western agriculture which has produced the master copy of various crops. In the area, wild populations were harvested from natural stands before deliberate cultivation and domestication (Zohary and Hopf, 2000). The role of the founders of agriculture played by the populations leaving the Crescent is supported by two kind of evidence: (i) the region is the intersection of the current distribution of the wild progenitors of modern cereal species (i.e *Triticum urartu*, *boeoticum* and *dicoccoides*), wild barley (*Hordeum spontaneum*) and wild rye (*Secale vavilovii*) (Zohary and Hopf, 2000); (ii) the seeds of the wild species have been recovered from the early archaeological sites of the region, which suggested the presence of spikelets reoriented perpendicularly to the rachis, and that of opened cupulate fruitcases, which exposed the grain.

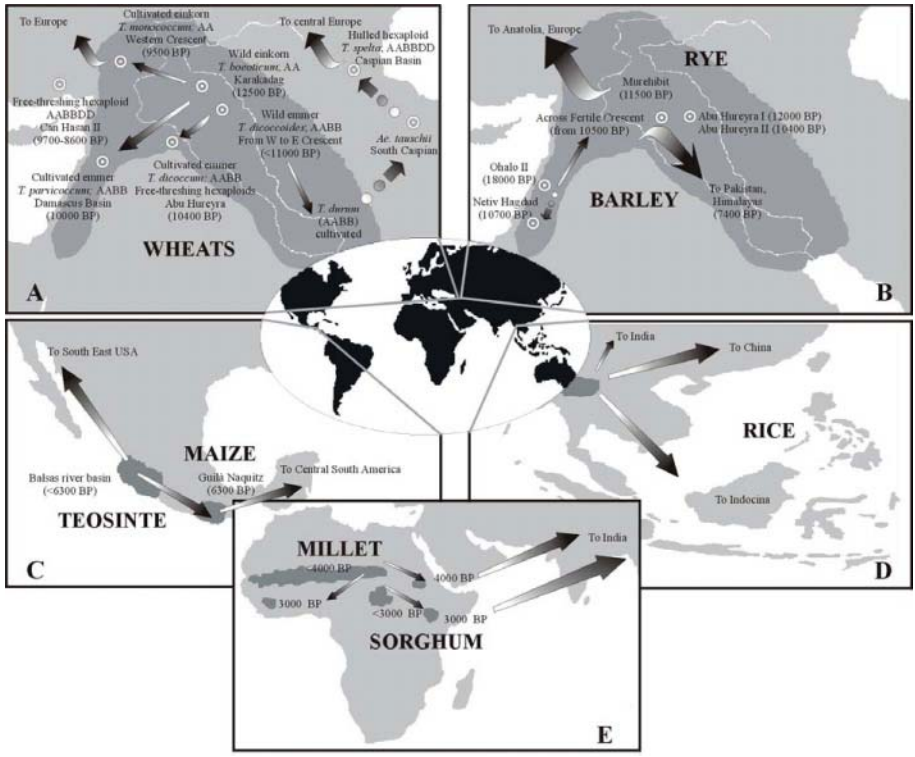


Figure 2. Worldwide centres of origin (location name or target symbol), time of domestication (in brackets, as years Before Present), diffusion and genome ploidy of the major cereals. Pathways of domestication are indicated with arrowed lines. Dotted lines indicate diffusion paths of uncertain definition. **A.** Wheats. The first wheat cultivated was the diploid einkorn whose core area of origin is in the Karacadag valley. It colonised western Crescent and then, as a feral form, reached the Balkans. The domestication from the wild species *T. boeoticum* occurred in the same area. The wild emmer *T. dicoccoides*, distributed from south-east Turkey to Iran and Iraq, has brittle ears that shatter at maturity into spikelets bearing larger seeds. On the contrary the domesticated form *T. dicoccum* has a non-brittle rachis. The wild emmer non free threshing forms *T. parvicoccum* diffused into Lebanon while the free threshing *T. durum* (hard wheat) is the only tetraploid wheat cultivated today. The final step of the wheat domestication is the occurrence of free threshing hexaploid *T. vulgare* (bread wheat), documented in south Turkey and east Iraq. See Text for further details. **B.** Barley originated in Israel-Lebanon from its wild progenitor *Hordeum spontaneum*. It has been domesticated across the Fertile Crescent and diffused to the Himalayas after introgression with the Asian *H. spontaneum* into an hybrid *H. agriocrithon*. The centres of domestication of wild rye (*Secale vavilovi*) are northern Syria (Murehibit) and surrounding areas (Abu Hureyra I, II). **C.** Centres of origin for both teosinte and maize are indicated by darker areas. The arrow lines indicate the pathway of the domestication of maize, that was initially brought under domestication in the Balsas river basin from teosinte before 6300 BP, and was later introduced in the Oaxaca area (Guilà Naquitz). From this region, *Zea mays* spreaded later over the Americas **D**. Rice (*Oryza sativa*) was domesticated around 10000 BP from *Oryza rufipogon*, whose centre of origin was across the Thailand and Myanmar border (darker area). Domesticated rice then spread towards South-East Asia around 4000 BP. **E.** The belt (darker area) stretching from western Sudan to Senegal is supposed to be the centre of origin of millet (*Pennisetum glaucum*). The arrow lines from its centre of origin show the two different domestication routes towards either western Africa or South-East Sahara (small darker areas). Sorghum (*Sorghum bicolor*) originated in an area across Sudan and Chad, and subsequently diffused in North-East Africa (Ethiopia). From Sahara and Ethiopia, millet and sorghum were then diffused in the Indian subcontinent.

4.1. Wheats

Genetic fingerprinting helps in the identification of the present-day location of the wild progenitor populations that gave rise to domesticated diploid einkorn wheat (Heun *et al.*, 1997). The site of domestication of einkorn was identified from the analysis of 288 AFLP marker loci (Heun *et al.*, 1997). According to molecular data the wild stands of *T. boeoticum*, the einkorn progenitor, are situated at the western foothills of the Karakadağ mountains, in southeast Turkey. In the western crescent (at the Abu Hureyra site), abundant remains of domesticated einkorn, as old as 9,500 years are available; presumably then wheat appeared in Cyprus, Greece and the Balkans at 8,000 years BP, and later in Yugoslavia, Bulgaria and Hungary. Einkorn was important for the early agriculture of Central Europe, but its cultivation started to decline in the Bronze Age. The location of domestication is, however, still a matter of discussion. Wild tetraploid emmer wheat (*T. dicoccoides*) is currently found in Israel, Jordan, Syria, Lebanon, south-east Turkey, north Iraq and west Iran. The domesticated form, *T. dicoccum*, first appeared in South-west Asia during the early part of the eighth millennium BC (Nesbitt and Samuel, 1998), and subsequently spread into Europe, reaching the Balkans by 8,500 years BP, and then following two routes across Europe, one westwards along the coast to Southern Italy, France and Spain, and the other through the river valleys of central Europe (reaching the Northern coast by 6,000 years BP; Barker, 1985). Emmer became the principal cereal of the European Neolithic and Bronze Age and was the progenitor of the free-threshing tetraploids (*T. durum*) and the hexaploids bread wheats that subsequently predominated (Zohary and Hopf, 2000). An analysis of AFLP data at 204 loci has indicated that domesticated tetraploid AABB wheats are most closely related to wild emmer populations from southeast Turkey. In fact, they are molecularly more similar to domesticated tetraploid wheats than are other populations sampled (Özkan *et al.*, 2002). The tetraploid *T. dicoccum*, 9,500-7,000 years ago, had spread from the Fertile Crescent to Mesopotamia, and then in Egypt, in the Mediterranean basin, in Ethiopia, Asia and India. It has been the most used cereal until the appearance of the naked tetraploid *T. turgidum* var. *durum* (the *pasta* wheat).

The origin of hexaploid wheats has represented a special event. Both *T. spelta* (hulled wheat) and *T. vulgare* (free-threshing) have an AABBDD hexaploid genome, which is not present in wild species. McFadden and Sears (1946) showed that the progenitors of the hexaploids are the tetraploid *T. turgidum* (AABB) and the wild grass *Ae. tauschii* (DD). At some unknown place, tetraploids hybridised with the diploid species and generated spelt-like hulled hexaploids. The involvement of a domesticated

tetraploid, *T.turgidum*, is suspected, because the distribution range of *Ae. tauschii* does not overlap with the distribution of the wild tetraploid *T. dicoccoides* (Nesbitt and Samuel, 1998). The AABBDD genome therefore probably stems from the anthropogenic expansion of tetraploid domesticated species into the distribution area of *Ae. tauschii* (Zohary and Hopf, 2000). The southern Caspian basin is a likely place for this hybridization, because the *strangulata* subspecies, a form of *Ae. tauschii* (the D-genome progenitor) still grows in this area and was considered to be the actual donor (discussed in Cox, 1998; Dvoràk *et al.*, 1998). However, newer data, which were obtained using molecular markers, raise doubts about the direct participation of the *strangulata* subspecies in this event (Lelley *et al*., 2000).

4.2. Maize

The morphological changes probably coincided with introduction of maize into habitats, which were never colonized by teosinte. The initial domestication of maize is believed to have taken place in close proximity to the Mexican city of Oaxaca, 400-500 km east of the present-day central Balsas river valley, which is the habitat of the subspecies of annual teosinte (Piperno and Flannery, 2001; Benz, 2001; Doebley, 1990). This finding is supported by phylogenetic data derived from the analysis of allozyme systems (Doebley, 1990). The domestication bottleneck must have been relatively short, according to the time lapse between domestication and the movement of human population from the core domestication area. This corresponds to an initial estimated population of roughly 6000 individuals (which corresponds to a different estimate as compared to the 600 derived from the analysis of *Adh1* locus, but still constitutes a very limited population; Eyre-Walker *et al.*, 1998). These data are in line with the findings derived from the analysis of the *tb1* gene (Wang *et al.*, 1999), suggesting that the process of domestication could have taken a few hundred years with only modest levels of selection.

4.3. Barley

It is believed that the event of first seed harvest from natural stands of two-rowed, brittle, wild *H. spontaneum* preceded the appearance of agriculture in the Fertile Crescent (Zohary and Hopf, 2000), as proven by remains found in Northern Syria and close to the Sea of Galilee dating as far back as 9,000-19,000 years BP. Non-brittle barley (i.e. domesticated) was first

grown in Tell Abu Ureyra, Tell Aswad, and Jarmo (Iraq), as inferred by Badr *et al.* (2000). Molecular data have been to some extent useful in the search for the location of barley domestication (Badr *et al.*, 2000). The southern part of the Fertile Crescent (Israel-Jordan area) is the most likely place for domestication of barley, since wild barley populations from this region are genetically very similar to domesticated forms. The once popular idea of a domestication center in the Himalayas has to be discarded because the local Himalayan wild form (*H. agriocrithon*) was found to be a hybrid between wild and domesticated types (Staudt, 1961; Badr *et al.* 2000). Wild populations found in the southern part of the Fertile Crescent in western Iran have also contributed germplasm to the cultivated barley on its way to the Himalayas. Badr *et al*. (2000) concludes that from its domestication in the western Fertile Crescent, barley moved eastwards and diversified locally in the Himalayas.

Traits which distinguish wild and cultivated barley must have appeared following a specific *tempo*: for example, tough and non-shattering ear, reduction in awn stiffness and an increase in grain size must have occurred very rapidly, while increases in grain yield and malting quality have been possibly accumulated over successive cycles of conscious selection (Ladizinsky and Genizi, 2001).

5. DOMESTICATION IN THE PRESENT DAY: THE GREEN REVOLUTION

As described in the above paragraphs, a key process in domestication was the selection of favourable alleles of the relatively few “domestication genes”. Also, modern plant breeding perhaps remained associated with novel variation in the same genes, but the variation at different genes/alleles acquired a major role, once the initial “domestication alleles” were fixed (Paterson *et al.*, 1995b; Xiong *et al.*, 1999). In barley, for example, after domestication eliminated the alleles causing brittleness of the rachis, additional loci conditioning rachis weakness were identified and became a target for breeding (reviewed in Kandemir *et al.*, 2000).

Flowering time, plant height and yield are other examples of domestication traits currently under selection for the genetic improvement of cereals. The cloning of genes supporting these traits is providing useful tools to optimise crop breeding. In rice, modern varieties carrying a photoperiod-insensitive allele of the *Se1/Hd-1* gene can be grown during any season and in most tropical and subtropical countries (Khush, 2001). *Se1/Hd-1* is closely related

to *CONSTANS*, a gene, which controls photoperiod response in Arabidopsis (Yano *et al.*, 2000). Recently, the *Hd3a* rice gene - also involved in the transition to flowering - has been isolated and shown to be orthologous to the *FLOWERING LOCUS T* (*FT*) gene, which promotes flowering in Arabidopsis (Kojima *et al.*, 2002). In the same crop, breeding for early maturity has taken advantage of the *ef* (early flowering) genes (Khush, 2001).

The reduction of wheat growth cycle was achieved by exploiting the *Ppd1* and *Ppd2* genes that cause photoperiod insensitivity (reviewed in Khush, 2001). An orthologue and syntenic gene in barley is also involved in photoperiod response (Börner *et al.*, 1998; reviewed in Griffiths *et al.*, 2003).

Selection of shorter plant stature began at the dawn of agriculture and continued during the *"Green Revolution"* (Borlaug, 1983). In the 1960s and 1970s, the application of large amounts of fertilizers caused traditional wheat and rice varieties to grow too tall and fall over, with consequent major yield losses. The problem was overcome by the deployment of new semi-dwarf lodging-resistant varieties which also partitioned a higher proportion of dry matter into the grain, leading to dramatic yield increases (reviewed in Hedden, 2003). In rice, the *Semidwarf1* (*Sd1*) gene encodes a GA20-oxidase enzyme controlling a key-step in gibberellin (GA) biosynthesis. The Green Revolution rice cultivars carry loss-of-function *sd1* alleles producing decreased GA synthesis and dwarf plants (Monna *et al.*, 2002; Sasaki *et al.*, 2002; Spielmeyer *et al.*, 2002). In wheat, semi-dominant mutations of the homoeologous/duplicated *Rht-B1a* and *Rht-D1a* genes conferring dwarfism and a reduced growth response to GA were used to develop new high yielding cultivars (Börner *et al.*, 1996; Peng *et al.*, 1999; reviewed in Hedden, 2003). The *Rht* gene encodes a repressor of GA signaling and is orthologous to the Arabidopsis *GAI*, maize *dwarf8 (d8)* and barley *Slender1 (Sln1)* genes, for which mutations that result in GA-insensitive dwarfs have also been described (Peng *et al*., 1997; Peng *et al.*, 1999; Chandler *et al.*, 2002). This class of genes - also present in rice (Ogawa *et al.*, 2000; Ikeda *et al.*, 2001) - encode GRAS transcription factors; the alleles, which have superior agronomic value, encode an altered protein insensitive to GA-dependent degradation and result in dwarf plants (Peng *et al.*, 1997; Peng *et al.*, 1999). Plant height and flowering time are partly controlled by common mechanisms, as demonstrated in several cereals. For instance, in wheat the *Rht* genes, which have pleiotropic effects on flowering time and tillering (Silverstone and Sun, 2000; Khush, 2001); in maize, where *d8* is a current target of selection for its adaptation to various flowering times (Thornsberry *et al*., 2001); and in barley, where the *ari-*

e.GP dwarfing allele found in the Scottish cultivar Golden Promise has been associated with early flowering (Ellis *et al.*, 1999). Pleiotropic effects are not surprising for genes controlling hormone action and may be of common occurrence for the traits targeted by domestication and breeding (Cai and Morishima, 2002).

6. HOW DURABLE ARE THE GENETIC CHANGES ESTABLISHED BY DOMESTICATION?

Weeds related to domesticated plants often appear in areas where no naturally occurring crop relatives are present. Some of these weeds may be feral forms, populations living in wild habitats that are derived either from crops or from crosses between wild and domesticated genotypes. These plants often exhibit intermediate characters between domesticated and wild forms and can help to analyze the stability of the genetic changes associated with domestication. For example, an einkorn feral form has arisen in the Balkans due to interbreeding of *T. monococcum* and *T. boeoticum,* indicating that the wild and domesticated gene pools have not been separated by speciation (Salamini *et al.*, 2002). 'Red rice' is an *Oryza* form that has recently spread in different regions of the world where no wild rice relatives are present (Bres-Patry *et al.*, 2001). Red rice has phenotipic traits intermediate between *Oryza rufipogon* and cultivated *indica* or *japonica* subspecies of *O. sativa* (Oka, 1988). Using a population derived from a 'red rice' × *O. sativa* cross, 12 out of the 29 identified QTLs underlying weediness (Bres-Patry *et al.*, 2001), corresponded to domestication QTLs previously identified in a cross between *O. sativa* and *O. rufipogon* (Xiong *et al*., 1999). Based on the concentration of QTLs in four regions of the genome, Paterson (2002) suggested that 'red rice' forms might be the result of 'reversion' of domestication genes to wild alleles at some of the loci. One implication of this hypothesis is that in the absence of human selection, the alleles involved in domestication might be rapidly lost.

7. A WAY TO THE FUTURE: EXOTIC GERMPLASM IN CEREAL BREEDING

The effect of selection during domestication and further breeding has led to the progressive limitation of the genetic variation in crop plants. For example, majority of hard red winter wheat cultivars in the United States have been derived from just two lines imported from Eastern Europe (Harlan, 1987). In rice, molecular analyses comparing modern varieties and

wild types reveal that the cultivated gene pool has a limited genetic variation compared to wild relatives (Wang *et al.*, 1992). A comparison of SSR allele frequencies in *H. spontaneum* and *H. vulgare* indicates a loss of rare alleles and a decrease in genetic diversity during domestication (Ellis *et al.*, 2000). Even in maize, which is considered to be a highly polymorphic species, genetic diversity at random loci has dropped by 30% on average (Buckler *et al.*, 2001). The narrow genetic basis of crops makes them more vulnerable to pests and represents a limitation for future genetic improvement (Harlan, 1987).

In contrast to the above, wild relatives of crops, carry agriculturally undesirable alleles together with few which are positive (e.g. monogenic disease resistances) and which were excluded from the cultivated gene pool due to domestication bottlenecks (Xiao *et al*., 1996; Tanksley and McCouch, 1997). In wild species, favourable QTL alleles often remain "cryptic" due to several factors including their low frequency, masking effects of deleterious alleles and the negative epistatic interactions (Xiao *et al.*, 1998; Gottlieb *et al.*, 2002; Lauter and Doebley, 2002; Peng *et al.*, 2003). However, QTL mapping with molecular markers in wild × cultivated crosses identified beneficial alleles derived from wild relatives (Xiao *et al.*, 1996, 1998; Poncet *et al.*, 2000, 2002; Peng *et al.*, 2003). These alleles have the potential to contribute to crop improvement when transferred to cultivated varieties.

Examples of the use of wild germplasm in traditional breeding have been reported for various cereal species. In wheat, almost 30 independent disease resistance genes have been identified, which were transferred using introgressions from the wild relatives (reviewed in Fedak, 1999, and Hoisington *et al.*, 1999). QTLs found in the exotic gene pool have also been exploited to enhance yield in cultivated varieties (Villareal *et al.*, 1995; Zamir, 2001). Yield increases in bread wheat have also been associated with a chromosome segment carrying a rust resistance gene from the wild species *Agropyron elongatum* (Hoisington *et al.*, 1999) and with a QTL from *T. dicoccoides* that improves the quality of pasta (Kovacs *et al.*, 1998). ''Synthetic'' hexaploid wheats obtained by crossing tetraploid wheat with diploid *Ae. tauschii* are used as a source of novel genetic variation for the improvement of agronomic characters, as demonstrated by segregating progenies derived from the hybridization of *T. durum* with *Ae. tauschii* (Hoisington *et al.*, 1999). In rice, genes for resistance to eight pathogens have been introgressed into cultivated germplasm from wild species (Brar and Khush, 1997). Barley genetic improvement has also benefited from the use of genes derived from wild barley and Middle Eastern landraces (Ellis *et al.*, 2000), as exemplified by the development of the *mlo*-mediated

resistance to powdery mildew (Büschges *et al.*, 1997; reviewed in Thomas *et al.*, 1998). In contrast, the potential of exotic germplasm for maize breeding is largely unexplored (Zamir, 2001), due to the high levels of polymorphism of the maize genome compared to other cereal species. However, raising concerns about the genetic vulnerability of maize have led to a recognition of the breeding value of exotic resources (Walsh, 1981; reviewed in Hosington *et al*., 1999), so that crosses to teosinte populations have already produced progenies incorporating traits with agricultural potential (Ray *et al.*, 1999). Genes useful for yield and disease resistance are also present in *Tripsacum,* a wild genus comprising several species related to maize (Cohen and Galinat, 1984; Hoisington *et al*., 1999).

Traditional breeding programmes have not always been successful in extracting useful traits from exotic germplasm. The use of beneficial alleles from wild species requires repeated back-crossing to recover most of the desirable agronomic traits and an efficient selection procedure is needed to retain the target allele from the exotic donor. Even when these are applicable, linkage drag may compromise the final result. These problems are exacerbated, in the case of complex agronomic traits, by the existence of numerous interacting QTLs, whose expression is also significantly influenced by the environment. These drawbacks can be in part solved using molecular tools (Xiao *et al.*, 1996, 1998; Zamir, 2001).

Two molecular-map-based studies have been conducted for simultaneous identification and transfer of wild QTLs into a cultivated genetic background. In the Advanced Back-cross (AB) QTL method (Tanksley and Nelson, 1996), molecular linkage maps are used to analyse populations obtained by repeatedly back-crossing a wild parent to a recurrent domesticated parent. The outcome of this procedure is a subset of alleles from the wild species that can be mapped and evaluated in a cultivated genetic background (Tanksley and Nelson, 1996). The rationale of this strategy is that beneficial wild alleles can be recovered from transgressive segregants that outperform the cultivated parent (Tanksley and Nelson, 1996; Tanksley and McCouch, 1997). Using this method, yield of a highly productive rice hybrid has been enhanced by the introduction of two QTL alleles from the low-yielding *O. rufipogon*, each increasing yield by about 17% compared with the original hybrid (Xiao *et al.*, 1996, 1998). A second AB experiment has shown that *O. rufipogon* can contribute favourable alleles for the improvement of rice cultivation under low-input cultural conditions (Moncada *et al.*, 2001). In both studies, about 50% of trait-enhancing QTL alleles were derived from the phenotypically inferior wild parent (Xiao *et al.*, 1998, Moncada *et al.*, 2001). Recently the AB method

has also been applied to barley and wheat with positive results for yield improvement (Huang *et al.*, 2003; Pillen *et al.*, 2003).

A refinement of the AB QTL method is the construction of exotic libraries (Zamir, 2001). An exotic library is a collection of homozygous lines each containing a single marker-defined chromosomal segment from a wild species inserted within a uniform cultivated background. In this strategy, backcrossing and marker-assisted selection are repeated for six generations and are followed by two more cycles of self-pollination to achieve homozygosity. Exotic libraries are a permanent genetic resource that can be tested independently by several investigators for multiple traits and can be directly used for breeding. The limited proportion of wild genome in the introgressed lines also reduces linkage drag effects. In grasses, a pioneering study anticipated the exotic library concept, when Kuspira and Unrau (1957) used whole-chromosome substitution lines to analyse quantitative traits in wheat. A collection of wheat introgression lines carrying chromosome segments from *Ae. tauschii* was recently characterised with microsatellite markers (Pestsova *et al.*, 2001). In rice, *O. glaberrima* (Doi *et al.*, 1998), *O. glumaepatula*, *O. meridionalis* and *O. eichingeri* (Yan et al., 2001) have already been used as donor parents to develop chromosome segment substitution lines (Yano, 2001).

Finally, the completion of the rice genome sequence and the availability of map positions for useful QTLs offer immense opportunities for the development of informative markers. These markers may be instrumental in isolating new useful alleles from wild germplasm and in eliminating the linkage drag (Yano, 2001).

8. REFERENCES

Ahn S., Anderson J.A., Sorrells M.F., Tanksley S.D. Homoeologous relationships of rice, wheat, and maize chromosomes. Mol Gen Genet 1993; 241: 483-490

Ayoub M., Symons S.J., Edney M.J., Mather D.E. QTLs affecting kernel size and shape in two-rowed by six-rowed barley cross. Theor Appl Genet 2002; 105: 237-247

Badr A., Müller K.J., Schäfer-Pregl R., El Rabey H., Effgen S., Ibrahim H.H., Pozzi C., Rohde W., Salamini F. On the origin and domestication history of barley (*Hordeum vulgare*). Mol Biol Evol 2000; 17: 499-510

Barker G. *Prehistoric farming in Europe*, Cambridge Univ Press, 1985

Barton N.H., Keightley P.D. Understanding quantitative genetic variation. Nature Rev Genet 2002; 3: 11-21

Benz B.F. Archaeological evidence of teosinte domestication from Guilà Naquitz, Oaxaca. Proc Natl Acad Sci USA 2001; 98: 2104-2106

Blake N.K., Lehfeldt B.R., Lavin M., Talbert L.E. Phylogenetic reconstruction based on low copy DNA sequence data in an allopolyploid: the B genome of wheat. Genome 1999; 42: 351-360

Borlaug N.E. Contributions of conventional plant breeding to food production. Science 1983; 219: 689-693

Börner A., Korzun V., Worland A.J. Comparative genetic mapping of loci affecting plant height and development in cereals. Euphytica 1998; 100: 245–48

Börner A., Plaschke J., Korzun V., Worland, A.J. The relationships between the dwarfing genes of wheat and rye. Euphytica 1996; 89: 69-75

Brar D.S., Khush G.S. Alien introgression in rice. Plant Mol Biol 1997; 35: 35–47

Bres-Patry C., Lorieux M., Clement G., Bangratz M., Ghesquirre A. Heredity and genetic mapping of domestication-related traits in a temperate *japonica* weedy rice. Theor Appl Genet 2001; 102: 118–126

Buckler E.S. IV, Thornsberry J.M., Kresovich S. Molecular diversity, structure and domestication of grasses. Genet Res Camb 2001; 77: 213-218

Büschges R., Hollricher K., Panstruga R., Simons G., Wolter M., Frijters A., van Daelen R., van der Lee T., Diergaarde P., Groenendijk J., Topsch S., Vos P., Salamini F., Schulze-Lefert P. The barley *Mlo* gene: a novel control element of plant pathogen resistance. Cell 1997; 88: 695-705

Cai W., Morishima H. QTL clusters reflect character associations in wild and cultivated rice. Theor Appl Genet 2002; 104: 1217-1228

Cantrell R.G., Joppa L.R. Genetic analysis of quantitative traits in wild emmer (*Triticum turgidum L.* var. *dicoccoides*). Crop Sci 1991; 31: 645-649

Cao W., Scoles G.J., Hucl P. The genetics of rachis fragility and glume tenacity in semi-wild wheat. Euphytica 1997; 94: 119-124

Chandler P.M., Marion-Poll A., Ellis M., Gubler F. Mutants at the *Slender1* locus of barley *cv*. Himalaya. Molecular and physiological characterization. Plant Physiol 2002; 129: 181–190

Chen Q.F., Yen C., Yang J.L. Chromosome location of the gene brittle rachis in the Tibetan weedrace of common wheat. Genet Res Crop Ev 1998; 45: 407-410

Cho Y.C., Chung T.Y., Suh H.S.D. Genetic characteristics of Korean weedy rice (*Oryza sativa* L.) by RFLP analysis. Euphytica 1995; 86: 103–110

Clement C.R. 1492 and the loss of Amazonian crop genetic resources. I. The relation between domestication and human population decline. Econ Bot 1999; 53: 188-202

Cohen J.I., Galinat W.C. Potential use of alien germplasm for maize improvement. Crop Sci 1984 24: 1011–1015

Cox T.S. Deepening the wheat gene pool. Jour Crop Prod 1998; 1**:** 1-25

Devos K., Pittaway T.S., Reynolds A., Gale M.D. Comparative mapping reveals a complex relationship between the pearl millet genome and those of foxtail millet and rice. Theor Appl Genet 2000; 100: 190–198

Diamond J. *Guns, Germs and Steel*. London: Random House, 1997

Doebley J. Molecular evidence and the evolution of maize. Econ Bot 199044 (Suppl); 6-27

Doebley J., Stec A. Genetic analysis of the morphological differences between maize and teosinte. Genetics 1991; 129: 285-295

Doebley J., Stec A. Inheritance of morphological differences between maize and teosinte: comparison of results of two F_2 populations. Genetics 1993; 141: 333-346

Doebley J., Stec A., Hubbard L. The evolution of apical dominance in maize. Nature 1997; 386: 485-488

Doebley J., Stec A., Wendel J., Edwards M. Genetic and morphological analysis of maize-teosinte F_2 population: implications for the origin of maize. Proc Natl Acad Sci USA 1990; 87: 9888-9892

Doebley J., Stec A., Gustus C. *Teosinte branched1* and the origin of maize: evidence for epistasis and the evolution of dominance. Genetics 1995; 141: 333-346

Doi K., Yoshimura A., Iwata N. RFLP mapping and QTL analysis of heading date and pollen sterility using backcross population between *Oryza sativa* L. and *Oryza glaberrima* Steud. Breed Sci 1998; 48:395-399

Dorweiler J., Stec A., Kermicle J., Doebley J. *Teosinte glume architecture 1*: a genetic locus controlling a key step in maize evolution. Science 1993; 262: 233-235

Dorweiler J.A., Doebley J. Developmental analysis of *Teosinte glume architecture 1*: a key locus in the evolution of maize (Poaceae). Am J Bot 1997; 84: 1313-1322

Draye X., Lin Y.-R.Q.X., Bowers J.E., Burow G.B., Morrell P.L., Peterson D.G., Presting G.G., Ren S.X., Wing R.A., Paterson A.H. Toward integration of comparative genetic, physical, diversity and cytomolecular maps for grasses and grains, using the *Sorghum* genome as a foundation. Plant Physiol 2001; 125: 1325-1341

Dvoràk J., Luo M.C., Yang Z.L., Zhang H.B. The structure of the *Aegilops tauschii* genepool and the evolution of hexaploid wheat. Theor Appl Genet 1998; 97: 657-670

Elias E.M., Steiger K.D., Cantrell R.G. Evaluation of lines derived from wild emmer chromosome substitutions. II. Agronomic traits. Crop Sci 1996; 36 : 228-233

Ellis R., Forster B.P., Robinson D., Handley L.L., Gordon D.C., Russell J.R., Powell W. Wild barley: a source of genes for crop improvement in the 21st century? Jour Exp Bot 2000; 51:9-17

Ellis R.P., Russell J., Ramsay L., Waugh R., Powell W. Barley domestication – *Hordeum spontaneum,* a source of new genes for crop improvement. Scottish Crop Research Institute Ann Rep 1998/99; 97-100 (http://www.scri.sari.ac.uk/)

Eyre-Walker A., Gaut R.L., Hilton H., Feldman D.L., Gaut B.S. Investigation of the bottleneck leading to the domestication of maize. Proc Natl Acad Sci USA 1998; 95: 4441-4446

Fedak G. Molecular aids for integration of alien chromatin through wide crosses. Genome 1999; 42: 584–591

Frary A., Nesbitt T., Frary A., Grandillo S., vanderKnaap E., Cong B., Liu J., Meller J., Elber R., Alpert K., Tanksley S. *fw2.2*: a quantitative trait locus key to the evolution of tomato fruit size. Science 2000; 289: 85-88

Galili S., Avivi Y., Millet E., Feldman M. RFLP-based analysis of three *RbcS* subfamilies in diploid and polyploid species of wheat. Mol Gen Genet 2000; 263: 674-680

Gaut B.S., Clegg M.T. Nucleotide polymorphism in the *Adh1* locus of pearl millet (*Pennisetum glaucum*) (Poaceae). Genetics 1993; 135: 1091-1097

Gopher A., Abbo S., Lev-Yadun S. The "when", the "where" and the "why" of the Neolithic revolution in the levant. Documenta Prehistorica 2002; 28: 49-62

Gottlieb T.M., Wade M.J., Rutherford S.L. Potential genetic variance and the domestication of maize. Bioassays 2002; 24: 685-689

Grandillo S., Ku H., Tanksley S. Identifying the loci responsible for natural variation in fruit size and shape in tomato. Theor Appl Genet 1999; 99: 978-987

Griffiths S., Dunford R.P., Coupland G., Laurie D. The evolution of *CONSTANS*-like gene families in barley, rice, and Arabidopsis. Plant Physiol 2003; 131: 1855-1867

Haag E.S., True J.R. Perspective: from mutants to mechanisms? assessing the candidate gene paradigm in evolutionary biology. Evolution 2001; 55: 1077-1084

Harlan J.R. Barley. In: Simmonds N.W. (ed.) *Evolution of Crop Plants*, Longman, London, UK, 1976, pp. 93-98

Harlan J.R. Gene centers and gene utilization in American agriculture. In: Yeatman C.W., Kafton D., Wilkes G. (eds.) *Plant Genetic Resources: A Conservation Imperative*, Westview, Boulder, CO, 1987, pp. 111–129

Harper J.L. *Plant Population Biology*. Academic Press, London, UK, 1977

Hedden P. The genes of the Green Revolution. Trends Genet 2003; 19: 5-9

Heun M., Schäfer-Pregl R., Klawan D., Castagna R., Accerbi M., Borghi B., Salamini F. Site of einkorn wheat domestication identified by DNA fingerprinting. Science 1997; 278: 1312-1314

Hirano H.Y., Eiguchi M., Sano Y. A single base change altered the regulation of the *Waxy* gene at the posttranscriptional level during the domestication of rice. Mol Biol Evol 1998; 15: 978-987

Hoisington D., Khairallah M., Reeves T., Ribaut J.M., Skovmand B., Taba S., Warburton M. Plant genetic resources: what can they contribute toward increased crop productivity? Proc Natl Acad Sci USA 1999; 96: 5937-43

Huang X.Q., Coster H., Ganal M.W., Roder M.S. Advanced backcross QTL analysis for the identification of quantitative trait loci alleles from wild relatives of wheat (*Triticum aestivum* L.). Theor Appl Genet 2003; 106: 1379-1389

Ikeda A., Ueguchi-Tanaka M., Sonoda Y., Kitano H., Koshioka M., Futsuhara Y., Matsuoka M., Yamaguchi J. *slender* rice, a constitutive gibberellin response mutant, is caused by a null mutation of the *SLR1* gene, an ortholog of the height-regulating gene *GAI/RGA/RHT/D8*. Plant Cell 2001; 13:999–1010

Kahler A.L., RW Allard R.W. Worldwide pattern of genetic variation among four esterase loci in barley (*Hordeum vulgare* L.). Theor Appl Genet 1981; 59: 101-110

Kandemir N., Kudrna D.A., Ullrich S.E., Kleinhofs A. Molecular marker assisted genetic analysis of head shattering in six-rowed barley. Theor Appl Genet 2000; 101:203-210

Kato K., Miura H., Akiyama M., Kuroshima M., Sawada S. RFLP mapping of the three major genes, *Vm1*, *Q* and *B1*, on the long arm of chromosome 5A of wheat. Euphytica 1998; 101: 91-95

Kellog E.A. Evolutionary history of the grasses. Plant Physiol 2001; 15: 133-138

Kerber E.R., Rowland G.G. Origin of the free threshing character in hexaploid wheat. Can Jour Genet Cytol 1974; 16: 145-154

Khush G.S. Green revolution: the way forward. Nature Rev Genet 2001; 2: 815-822

Kojima S., Takahashi Y., Kobayashi Y., Monna L., Sasaki T., Araki T., Yano M. *Hd3a*, a rice ortholog of the Arabidopsis *FT* gene, promotes transition to flowering downstream of *Hd1* under short-day conditions. Plant Cell Physiol 2002; 43: 1096-1105

Kovacs M.I.P., Howes N.K., Clarke J.M., Leisle D. Quality characteristics of durum wheat lines deriving high protein from *Triticum dicoccoides* (6b) substitution. Jour Cereal Sci 1998; 27: 47–51

Kuspira J., Unrau J. Genetic analysis of certain characters in common wheat using all chromosome substitution lines. Can Jour Plant Sci 1957; 37: 300–326

Kyozuka J., Konishi S., Nemoto K., Izawa T., Shimamaoto K. Down-regulation of *RFL*, the *FLO*/*LFY* homolog of rice, accompanied with panicle branch initiation. Proc Natl Acad Sci USA 1998; 95: 1979-1982

Ladizinsky G. Founder effect in crop plant evolution. Econ Bot 1985; 39: 191-199

Ladizinsky G., Genizi A. Could early gene flow have created similar allozyme-gene frequencies in cultivated and wild barley?. Genet Resour Crop Ev 2001; 48: 101-104

Lauter N., Doebley J. Genetic variation for phenotypically invariant traits detected in teosinte: implications for the evolution of novel forms. Genetics 2002; 160: 333-342

Le Thierry D'Ennequin M.L.T., Toupance B., Robert T., Godele B., Gouyon P. Plant domestication: a model for studying the selection of linkage. Jour Evol Biol 1999; 12: 1138-1147

Lelley T., Stachel M., Grausgruber H., Vollmann J. Analysis of relationships between *Aegilops tauschii* and the D genome of wheat utilizing microsatellites. Genome 2000 43: 661-668

Lev-Yadun S., Gopher A., Abbo S. The cradle of agriculture. Science 2000; 288: 1602-1603

Li Z., Pinson S.R.M., Stansel J.W., Park W.D. Identification of quantitative trait loci (QTLs) for heading date and plant height in cultivated rice (*Oryza sativa* L.). Theor Appl Genet 1995; 91: 374-381

Lin Y.-Y., Schertz K.F., Paterson A.H. Comparative analysis of QTLs affecting plant height and maturity across the Poaceae, in reference to an interspecific *sorghum* population. Genetics 1995; 141: 391-411

Lukens L.N., Doebley J. Epistatic and environmental interactions for quantitative trait loci involved in maize evolution. Genet Res Camb 1999; 74: 291-302

Luo D., Carpenter R., Copsey L., Vincent C., Clark J., Coen E. Control of organ asymmetry in flowers of *Antirrhinum* Cell 1999; 99: 367-376

Luo M.C., Yang Z.L., Dvoràk J. The *Q* locus of Iranian and European spelt wheat. Theor Appl Genet 2000; 100: 602-606

MacKey J. Neutron and X-ray experiments in wheat and a revision of the speltoid problem. Hereditas 1954; 40: 65-180

Matsuoka Y., Vigouroux Y., Goodman M.M., Sanchez G.J., Buckler E., Doebley J. A single domestication for maize shown by multilocus microsatellite genotyping. Proc Natl Acad Sci USA 2002; 99: 6080-6084

McFadden E.S., Sears E.R. The origin of *Triticum spelta* and its free-threshing hexaploid relatives. Jour Hered 1946; 37: 81-89

Moncada P., Martinez C.P., Borrero J., Chatel M., Gauch H. Jr, Guimaraes E., Tohme J., McCouch S.R. Quantitative trait loci for yield and yield components in an *Oryza sativa* × *O. rufipogon* BC_2F_2 population evaluated in an upland environment. Theor Appl Genet 2001*;* 102: 41–52

Monna L., Kitazawa N., Yoshino R., Suzuki J., Masuda H., Maehara Y., Tanji M., Sato M., Nasu S., Minobe Y. Positional cloning of rice semidwarfing gene, *sd-1*: rice "green revolution gene" encodes a mutant enzyme involved in gibberellin synthesis. DNA Res 2002; 9: 11-17

Morden C.W., Doebley J., Schertz K.F. Allozyme variation among the spontaneous species of *Sorghum* section *Sorghum* (Poaceae). Theor Appl Genet 1990; 80: 296-304

Morgante M., Salamini F. From plant genomics to breeding practice. Curr Opin Biotech 2003; 14: 214-219

Müller K.J., Romano N., Gerstner O., Garcia-Maroto F., Pozzi C., Salamini F., Rohde W. The barley *Hooded* mutation caused by a duplication in a homeobox gene intron. Nature 1995; 374: 727-730

Muramatsu M. Dosage effect of the spelta gene *q* of hexaploid wheat. Genetics 1963; 48: 469-482

Muramatsu M. The *vulgare* super gene *Q*: its universality in *durum* wheat and its phenotypic effects in tetrapolid and hexaploid wheat. Can Jour Genet Cytol 1986; 28, 30-41

Nesbitt M., Samuel D. Wheat domestication: archaeobotanical evidence. Science 1998; 279: 1433

Neuffer M.G., Coe E.H., Wessler S.R. *Mutants of Maize*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1997

Ogawa M., Kusano T., Katsumi M., Sano H. Rice gibberellin-insensitive gene homolog, *OsGAI*, encodes a nuclear-localized protein capable of gene activation at transcriptional level. Gene 2000; 245: 21–29

Oka H.I. Weedy forms of rice. In: Oka MI (ed.) *Origin of Cultivated Rice,* Japanese Science Society Press, Tokyo, Japan, 1988, pp. 107–114

Olsen K.M., Purugganan M.D. Molecular evidence on the origin and evolution of glutinous rice. Genetics 2002; 162: 941-950

Özkan H., Brandolini A., Schaefer-Pregl R., Salamini F. AFLP analysis of a collection of tetraploid wheats indicates the origin of emmer and hard wheat domestication in southeast turkey. Mol Biol Evol 2002; 19: 1797-1801

Paterson A.H., Schertz K., Lin Y., Liu S., Chang Y. The weedness of wild plants: molecular analysis of genes influencing dispersal and persistence of johnsongrass, *Sorghum halepense* (L.) *Pers.* Proc Natl Acad Sci USA 1995a; 92: 6127-6131

Paterson A.H., Lin Y.R., Li Z., Schertz K.F., Doebley J.F., Pinson S.R.M., Liu S.C., Stansel J.W., Irvine J.E. Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. Science 1995b; 269: 1714-1717

Paterson A.H. What has QTL mapping taught us about plant domestication? New Phytol 2002; 154: 591-608

Peng J., Carol P., Richards D.E., King K.E., Cowling R.J., Murphy G.P., Harberd N.P. The Arabidopsis *GAI* gene defines a signalling pathway that negatively regulates gibberellin responses. Gene Dev 1997; 11: 3194-3205

Peng J., Richards D.E., Hartley N.M., Murphy G.P., Devos K.M., Flintham J.E., Beales J., Fish L.J., Worland A.J., Pelica F., *et al.* 'Green revolution' genes encode mutant gibberellin response modulators. Nature 1999; 400: 256-261

Peng J., Ronin Y., Fahima T., Roeder M.S., Li Y., Nevo E., Korol A. Domestication quantitative trait loci in *Triticum dicoccoides* the progenitor of wheat. Proc Natl Acad Sci USA 2003; 100: 2489-2494

Pestsova E.G., Borner A., Roder M.S. Development of a set of *Triticum aestivum-Aegilops tauschii* introgression lines. Hereditas 2001; 135: 139-143

Pillen K., Zacharias A., Leon J. Advanced backcross QTL analysis in barley (*Hordeum vulgare* L.). Theor Appl Genet 2003; 107: 340-352

Piperno D.R., Flannery K.V. The earliest archaeological maize (*Zea mays* L.) from highland Mexico: new accelerator mass spectrometry dates and their implications. Proc Natl Acad Sci USA 2001; 98: 2010-2103

Poncet V., Lamy F., Devos K.M., Gale M.D., Sarr A., Robert T. Genetic control of domestication traits in pearl millet (*Pennisetum glaucum* L, Poaceae). Theor Appl Genet 2000; 100: 147–159

Poncet V., Martel E., Allouis S., Devos M., Lamy F., Sarr A., Robert T. Comparative analysis of QTLs affecting domestication traits between two domesticated × wild pearl millet (*Pennisetum glaucum* L., Poaceae) crosses. Theor Appl Genet 2002; 104: 965-75

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

Ray J. D., Kindiger B., Sinclair T. R. Introgressing root aerenchyma into maize. Maydica 1999; 44:113–117

Remington D.L., Thornsberry J.M., Matsuoka Y., Wilson L.M., Whitt S.R., Doebley J., Kresovich S., Goodman M.M., Buckler E.S. IV Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc Natl Acad Sci USA 2001; 98: 11479-11484

Salamini F., Őzkan H., Brandolini A., Schaefer-Pregl R., Martin W. Genetics and geography of wild cereal domestication in the near east. Nat Rev Genet 2002; 3: 429-441

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

Sears E.R. A synthetic hexaploid wheat with fragile rachis. Wheat Info Serv 1976; 41/42, 31-32

Shaked H., Kashkush K., Őzkan H., Feldmand M., Levy A.A. Sequence elimination and cytosine methylation are rapid and reproducible responses of the genome to wide hybridization and allopolyploidy in wheat. Plant Cell 2001; 13: 1749-1759

Sharma H.C., Waines J.G. Inheritance of tough rachis in crosses of *Triticum monococcum* and *T. boeoticum*. J Hered 1980; 7: 214-216

Silverstone A.L., Sun T.P. Gibberellins and the Green Revolution. Trends Plant Sci 2000; 5: 1-2

Simonetti M.C., Bellomo M.P., Laghetti G., Perrino P., Simeone R., Blanco A. Quantitative trait loci influencing free-threshing habit in tetraploid wheats. Genet Resources Crop Evol 1999; 46: 267-271

Snape J.W., Law C., Parker B.B., Worland A.J. Genetical analysis of chromosome 5A of wheat and its influence on important agronomic characters. Theor Appl Genet 1985; 71: 518-526

Søgaard B., von Wettstein-Knowles P. Barley: genes and chromosomes. Carlsberg Res. Comm 1987; 52: 123-196

Sourdille P., Tixier M.H., Charmet G., Gay G., Cadalen T., Bernard S., Bernard M. Location of genes involved in ear compactness in wheat (*Triticum aestivum*) by means of molecular markers. Mol Breed 2000; 6: 247-255

Spielmeyer W., Ellis M.H., Chandler P.M. *Semidwarf* (*sd-1*), "green revolution" rice, contains a defective gibberellin 20-oxidase gene. Proc Natl Acad Sci USA 2002; 99: 9043-9048

Staudt G. The origin of cultivated barleys: a discussion. Econ Bot 1961; 15: 205-212

Suh H.S., Sato Y.I., Morishima H. Genetic characterization of weedy rice (*Oryza sativa* L.) based on morpho-physiology, isozymes, and RAPD markers. Theor Appl Genet 1997; 94: 316–321

Taenzler B., Esposti R.F., Vaccino P., Brandolini A., Effgen S., Heun M., Schäfer-Pregl R., Borghi B., Salamini F. Molecular linkage map of Einkorn wheat: mapping of storage-protein and soft-glume genes and bread-making quality QTLs. Genet Res 2002; 80: 131-143

Takahashi R. The origin and evolution of cultivated barley. Adv Genet 1955; 7: 227-276

Takahashi R. *non brittle rachis 1* and *non brittle rachis 2*. Barley Genetic Newsletter 1972; 2: 181-182

Talbert L.E., Smith L.Y., Blake M.K. More than one origin of hexaploid wheat is indicated by sequence comparison of low-copy. Genome 1998; 41: 402-407

Tanksley S.D., Nelson J.C. Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. Theor Appl Genet 1996; 92: 191–203

Tanksley S.D., McCouch S.R. Seed banks and molecular maps: unlocking genetic potential from the wild. Science 1997; 277: 1063–1066

Thomas W.T.B., Baird E., Fuller J.D., Lawrence P., Young G.R., Russell J., Ramsay L., Waugh R., Powell W. Identification of a QTL decreasing yield in barley linked to *Mlo* powdery mildew resistance. Mol Breed 1998; 4: 381-393

Thornsberry J.M., Goodman M.M., Doebley J., Kresovich S., Nielsen D., Buckler E.S. IV *Dwarf8* polymorphisms associate with variation in flowering time. Nat Genet 2001; 28: 286-89

Ullrich S.E., Hayes P.M., Dyer W.E., Blake T.K., Clancy J.A. In: Walker-Simmonds M.K., Reid J.L. (eds.) *Pre-harvest Sprouting in Cereals*, American Association of cereal chemists, Inc., Saint-Paul, Minnesota, 1993, pp.136-145

Veit B., Briggs S.P., Schmidt R.J., Yanofsky M.F., Hake S. Regulation of leaf initiation by the *terminal ear 1* gene of maize. Nature 1998; 393: 166-168

Vigouroux Y., McMullen M., Hittinger C.T., Houchins K., Schulz L., Kresovich S., Matsuoka Y., Doebley J. Identifying genes of agronomic importance in maize by screening microsatellites for evidence of selection during domestication. Proc Natl Acad Sci USA 2002; 99: 9650-9655

Villareal R.L., Del Toro E., Mujeeb-Kazi A., Rajaram S. The 1BL/1RS chromosome translocation effect on yield characteristics in a *Triticum aestivum* L. cross. Plant Breed 1996; 114: 497–500

Walsh J. Genetic vulnerability down on the farm. Science 1981; 214: 161-164

Wang Z.Y., Second G., Tanksley S.D. Polymorphism and phylogenetic relationships among species in the genus *Oryza* as determined by analysis of nuclear RFLPs. Theor Appl Genet 1992; 83: 565-81

Wang R.L., Stec A., Hey J., Lukens L., Doebley J. The limits of selection during maize domestication. Nature 1999; 398: 236-239.

Wendel J.F. Genome evolution in polyploids. Plant Mol Biol 2000; 42: 225-249

Westerbergh A., Doebley J. Morphological traits defining species differences in wild relatives of maize are controlled by multiple quantitative trait loci. Evolution 2002; 56: 273-283

White S.E., Doebley J.F. The molecular evolution of *terminal ear1*, a regulatory gene in the genus *Zea*. Genetics 1999; 153: 1455-1462

Wilkes H.G. Hybridization of maize and teosinte, in Mexico and Guatemala and the improvement of maize. Econ Bot 1977; 31: 254-293

Xiao J., Grandillo S., Ahn S.N., McCouch S.R., Tanksley SD, Jiming L., Yuan L. Genes from wild rice improve yield. Nature 1996; 384: 223-224

Xiao J., Li J., Grandillo S., Ahn S.N., Yuan L., Tanksley S.D., McCouch S.R. Identification of trait-improving quantitative trait loci alleles from a wild rice relative, *Oryza rufipogon*. Genetics 1998; 150: 899-909

Xiong L.Z., Liu K.D., Dai X.K., Xu C.G., Zhang C. Identification of genetic factors controlling domestication-related traits of rice using an F_2 population of a cross between *Oryza sativa* and *O. rufipogon*. Theor Appl Genet 1999; 98: 243-51

Xueyong L., Qian Q., Zhiming F., Yonghong W., Guosheng X., Dall Z., Xiaoqun W., Xinfang L., Sheng T., Fujimoto H., Ming Y., Da L., Bin H., Jlayang L. Control of tillering in rice. Nature 2003; 422: 618-621

Yan H., Liu G., Cheng Z., Min S., Zhu L. Characterization of euploid backcross progenies derived from interspecific hybrids between *Oryza sativa* and *O. eichingeri* by restriction fragment length polymorphism (RFLP) analysis and genomic in situ hybridization (GISH). Genome 2001; 44: 86-95

Yano M., Katayose Y., Ashikari M., Yamanouchi U., Monna L., Fuse T., Baba T., Yamamoto K., Umehara Y., Nagamura Y., Sasaki T. *Hd1*, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the Arabidopsis flowering time gene *CONSTANS* Plant Cell 2000; 12: 2473-2484

Yano M. Genetic and molecular dissection of naturally occurring variation. Curr Opin Plant Biol 2001; 4: 130–35

Young B.A. Seed shattering in forage grasses - inheritance and possibilities for improvement. Proc Grass Breed Work Planning Conf, 1986

Yutaka S., Yoshimichi F., Hiro-Yuki H. Mutations that cause amino acid substitutions at the invariant positions in homeodomain of *OSH3* KNOX protein suggest artificial selection during rice domestication. Genes Genet Syst 2001; 76: 381-392

Zamir D. Improving plant breeding with exotic genetic libraries. Nature Rev Genet 2001; 2: 983-989

Zohary D., Feldman M. Hybridization between amphydiploids and the evolution of polyploids in the wheat group (*Aegilops-Triticum*). Evolution 1962; 16: 44-61

Zohary D. Monophyletic *vs.* polyphyletic origin of the crops on which agriculture was founded in the Near East. Genet Resour Crop Ev 1999; 46: 133-142

Zohary D., Hopf M. *Domestication of Plants in the Old World*, 3rd edn, Oxford Univ Press, Oxford, UK, 2000

Chapter 8

QTLs AND GENES FOR DISEASE RESISTANCE IN BARLEY AND WHEAT

Ahmed Jahoor[1,*], Lars Eriksen[2] and Gunter Backes[1]

[1]*Risø National Laboratory, Plant Research Department, Frederiksborgvej 399, DK–4000 Roskilde, Denmark;* [2]*Sejet Plantbreeding, Nørremarksvej 67, DK–8700 Horsens, Denmark*
[*]Author for correspondence: a.jahoor@risoe.dk

1. INTRODUCTION

The diseases caused by fungal pathogens are the economically most important foliar diseases in cereals not only in the temperate climate zone but also in the subtropical and tropical zones. These diseases can reduce the kernel yield and kernel quality of the affected plants dramatically (Griffey *et al.*, 1994; Conry and Dunne 2001). To control these diseases, fungicides are often used. The chemical protection is costly and not always effective and in addition it can cause environmental risk. Therefore, genetic resistance is the most economic sustainable way to control these diseases in cereals. Two kinds of resistance—qualitative or major race–specific and quantitative or partial resistance—have been described and analyzed in the crop plants (van der Plank, 1968). The major race–specific resistance is frequently controlled by the interaction between a dominant resistance (R) gene in the host plant and a dominant avirulance (Avr) gene of the pathogen (Hammond–Kosack and Jones, 1997; Collins *et al*., 2003). This system is in agreement with Flor´s gene–for–gene hypothesis that was confirmed for several crop plant–pathogen systems including cereals. In contrast, quantitative resistance should be effective against most of the pathogen populations. Due to this fact, quantitative resistance does not follow Flor`s gene–for–gene model. Based on the gene–for–gene hypothesis (Flor, 1955), different genes for diseases resistance have been identified (Jahoor and Fischbeck, 1987b). However, the restricted availability of new resistance genes in the gene pool of cultivated

P.K. Gupta and R.K. Varshney (eds.), Cereal Genomics, 199-251.

cereals and a rapid development of the pathogen toward new virulence (Hovmøller *et al.,* 2001) forces the breeders to look for effective resistance genes in the wild relatives or exotic material (Jahoor and Fischbeck, 1987a; b; Garvin *et al.*, 2000). With an increased number of resistance genes that can be incorporated in a breeding line or in a cultivar, the use of host–pathogen interaction to identify different resistance genes for a disease will become very difficult, since the disproportionate increase in the number of pathogen isolates needed will become prohibitive (Schönfeld *et al.,* 1996). This difficulty can be overcome using the new methods involving DNA markers to identify resistance genes in different genetic backgrounds. Mapping with molecular markers, such as RFLPs (*R*estriction *F*ragment *L*ength *P*olymorphism), AFLPs (*A*mplified *F*ragment *L*ength *P*olymorphism), RAPDs (*R*andom *A*mplified *P*olymorphism *D*NA) and microsatellites is a powerful tool to localize resistance genes in cereal genomes without the knowledge of their function or their sequences. The development of molecular markers and construction of genetic maps of barley and wheat (Graner *et al.*, 1991; Ramsay *et al.*, 2000; Röder *et al.*, 1998) made it possible to localize genomic regions controlling the expression of disease resistance genes.

By the availability of dense linkage maps of the barley and wheat genome and progress in the field of molecular genetics, two major advance were made in the field of resistance of cereals against diseases: individual gene loci (quantitative trait loci, QTLs) have been identified and mapped as the cause for quantitatively inherited resistance (Backes *et al.,* 1995; Backes *et al.*, 2003; Eriksen *et al.*, 2003). Quantitative resistance is characterized by a more or less continuous transition from susceptible to resistant genotypes in a segregating population, while clear groups of resistant and susceptible lines can be defined for qualitative genes. For quantitative resistance against powdery mildew, leaf rust and scald in barley, QTLs were localized in different genetic backgrounds (Heun 1992; Backes *et al.,* 1996; Chen *et al.,* 1994; Thomas *et al.,* 1995; Qi *et al.,* 1998; Spaner *et al.,* 1998; Falak *et al.,* 1999; Kicherer *et al.,* 2000; Toojinda *et al.,* 2000).

Pyramiding of different highly effective resistance genes for a disease will probably provide a more durable resistance. Further combination of highly effective major race specific genes with genes underlying quantitative resistance will certainly lead to more durable resistance. However, such combinations of different kinds of resistance cannot be identified with the help of pathogen isolates. DNA markers offer an opportunity not only to detect combinations of qualitative resistance genes to pyramid in a cultivar but they can also be used to combine qualitative and quantitative inherited resistance in a variety.

2. GENES FOR DISEASE RESISTANCE IN BARLEY (*HORDEUM VULGARE* L.)

Improvement of resistance against diseases caused by fungal pathogens, such as powdery mildew, different rusts, scald and barley leaf stripe are causing huge reductions in the realized kernel yield compared with the yield potential of barley varieties. Several genes for resistance against these pathogens are known and have been localised, either acting qualitatively leading to absolute resistance, or quantitatively, resulting in partial resistance. A few of these genes have also been isolated, giving deeper insight into the mechanisms of disease resistance (Büschges, *et al.*, 1997; Zhou, *et al.*, 2001; Brueggeman, *et al.*, 2002).

In this section, we discuss resistance genes for several fungal pathogens, each causing a different disease. The location of these genes is summarized in Fig. 1, with the qualitatively acting genes on the right side of the respective chromosome and the quantitatively acting gene on the left side. This is based on the joined map published by Qi *et al.* (1996) and other information published thereafter. This other information includes the joined map of 1H published by Jensen (2002), the map of morphological markers published by Franckowiak (1997), the map of the cross '1B–87' × 'Vada' (Graner, *et al.*, 1991), the joined map published by Mohler (1997), the map from 'Proctor' × 'Nudinka' combining AFLP and RFLP markers (Becker, *et al.*, 1995), the maps published by Liu *et al.* (1996), combining RFLP markers and microsatellites and the map of the cross 'Vogelsanger Gold' × 'Tystofte Prentice' (Kjær, *et al.*, 1995). The positions, of the centromeres were derived from the consensus map published by Langridge *et al.* (1995).

2.1. Qualitatively Inherited Genes

2.1.1. Powdery Mildew (Caused by *Blumeria graminis* f.sp. *hordei*)

On chromosome 1H, five powdery mildew resistance genes have been localised: *Mlra*, *Mla*, *Mlk* and *Mlnn* on 1HS, and *MlGa* on 1HL. *Mla* has been shown to be the most polymorphic of all known barley *Ml*–genes identified in the line 'Algerian' (Briggs & Stanford, 1938) and was localised with molecular markers by Graner *et al.* (1991) near the RFLP marker cMWG645. Schüller *et al.* (1992) found a close linkage between *Mla* and the RFLP marker MWG036 and finally, Schwarz *et al.* (1999) detected co–segregation between the RFLP marker MWG2197 and the resistance gene. The gene is one of the few barley resistance genes with known sequence

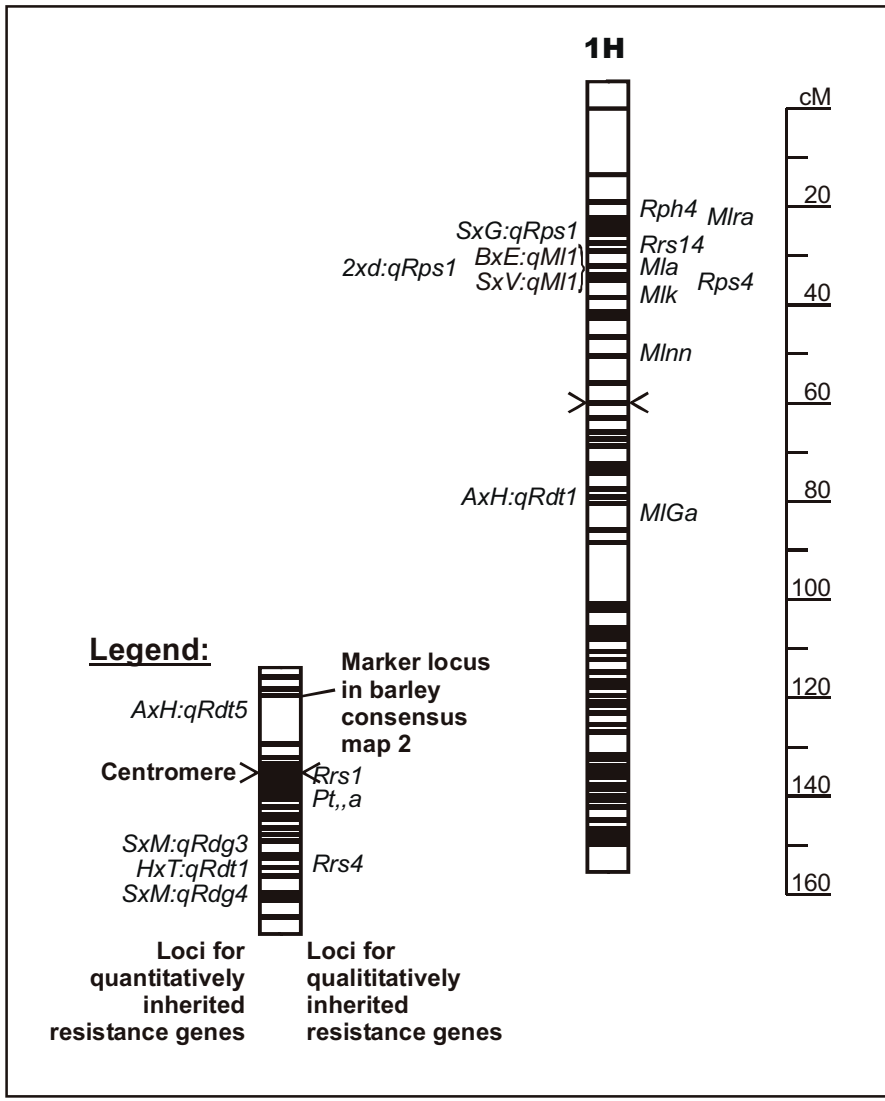


Figure 1. Quantitatively and qualitatively inherited resistance genes localised on the different barley chromosomes based on a joined linkage map (Qi *et al.*, 1996).

information; it was isolated by Zhou *et al.* (2001) and encoded a 108–kDa protein with a nucleotide–binding region (NBR) and a leucine–rich repeat (LRR) region.

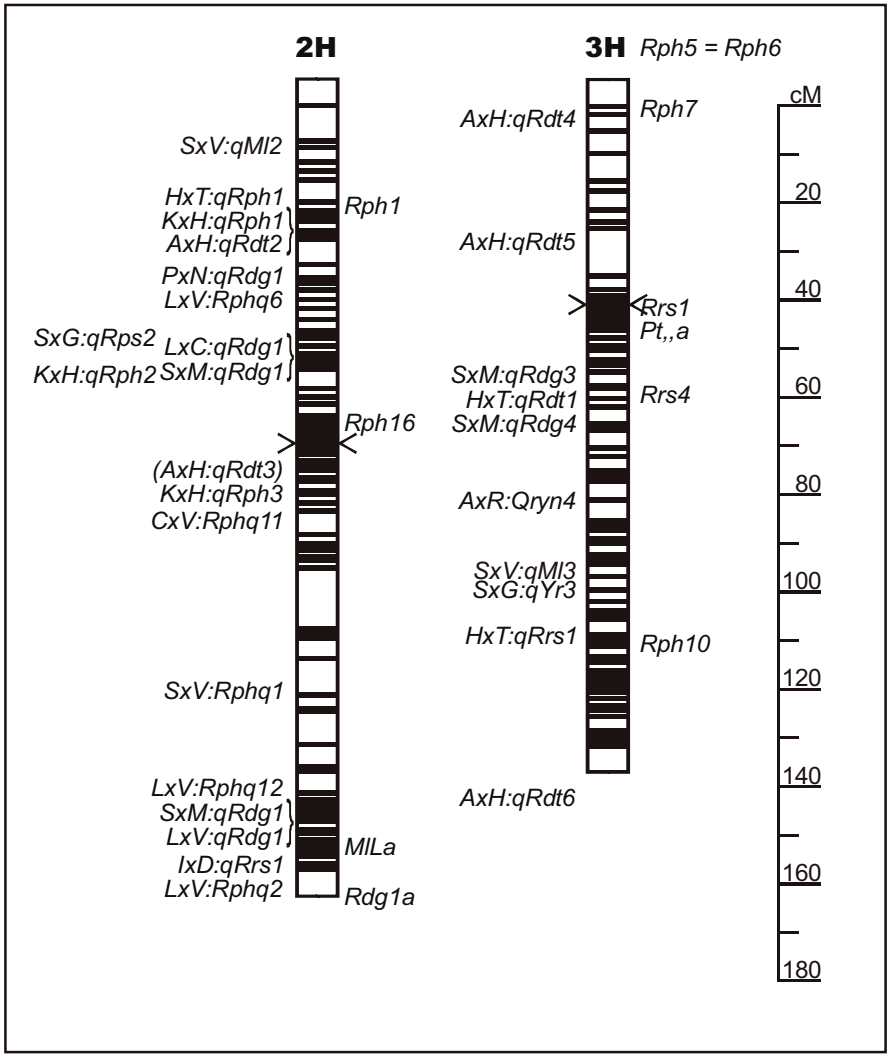


Figure 1. Continued

More proximal on the same chromosome arm is located, the *Mlra*–gene, originating from 'Ragusa b'. It was first described by Wiberg (1974) in the line 'Weihenstephan 41/ 145' and localized by Doll and Jensen (1986) between the hordein genes *Hor1* and *Hor2*. In contradiction to that, Jensen (2002) placed the gene distal from *Hor2* and near the leaf rust resistance gene *Rph4* (*Pa4*). *Mlk* and *Mlnn* are also localized on the short arm of this chromosome, but more proximal than *Mla*. *Mlk* was described for the first time together with *Mla* by Briggs and Stanford (1938). Giese (1981) localized the gene on 1H and according to Jensen (2002), the distance between *Mla*

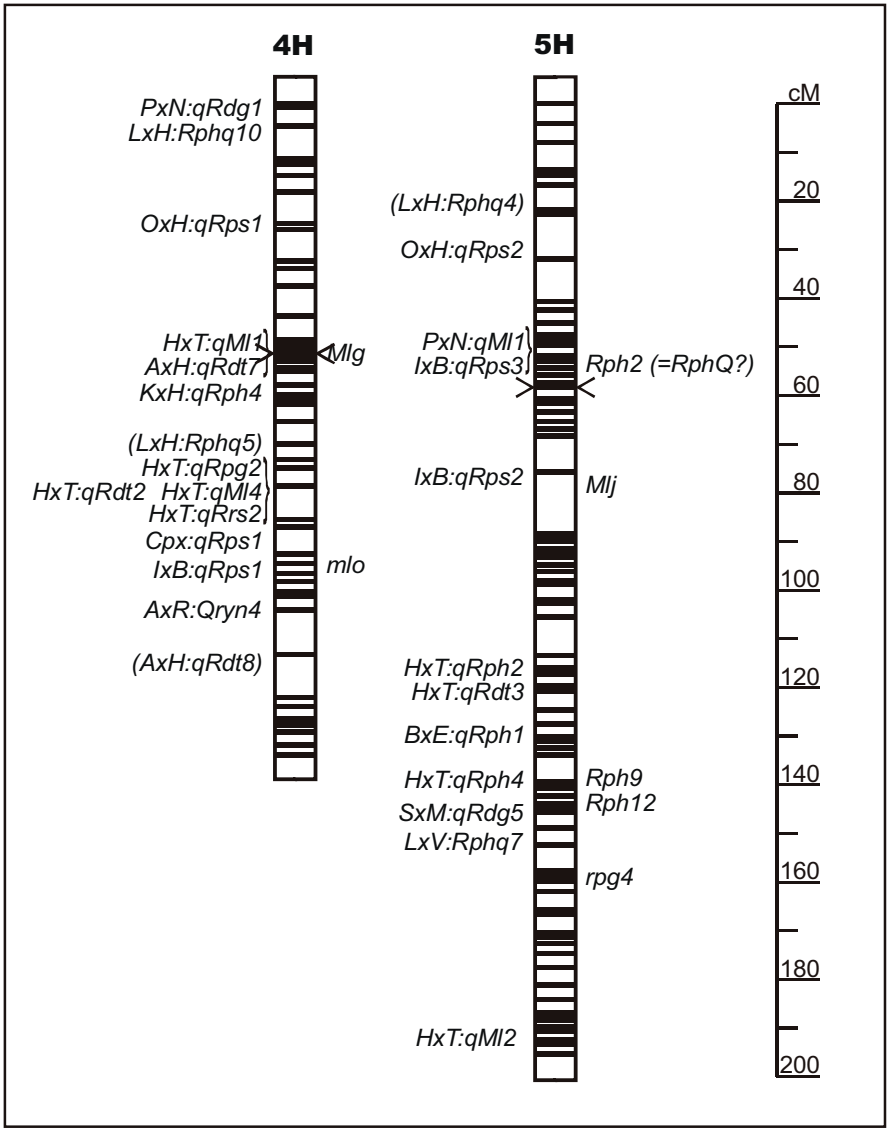


Figure 1. Continued

and this gene locus is only 0.5 cM. *Mlnn* was detected by Hiura (1960) from the line 'Nigrinudum' (C.I. 11549). The gene was localized by Jensen and Jørgensen (1975) and more precisely in a joined map by Jensen (2002) between the RFLP markers CDO99 and ABG053. The last powdery mildew resistance gene on this chromosome is *MlGa* that was detected and localized

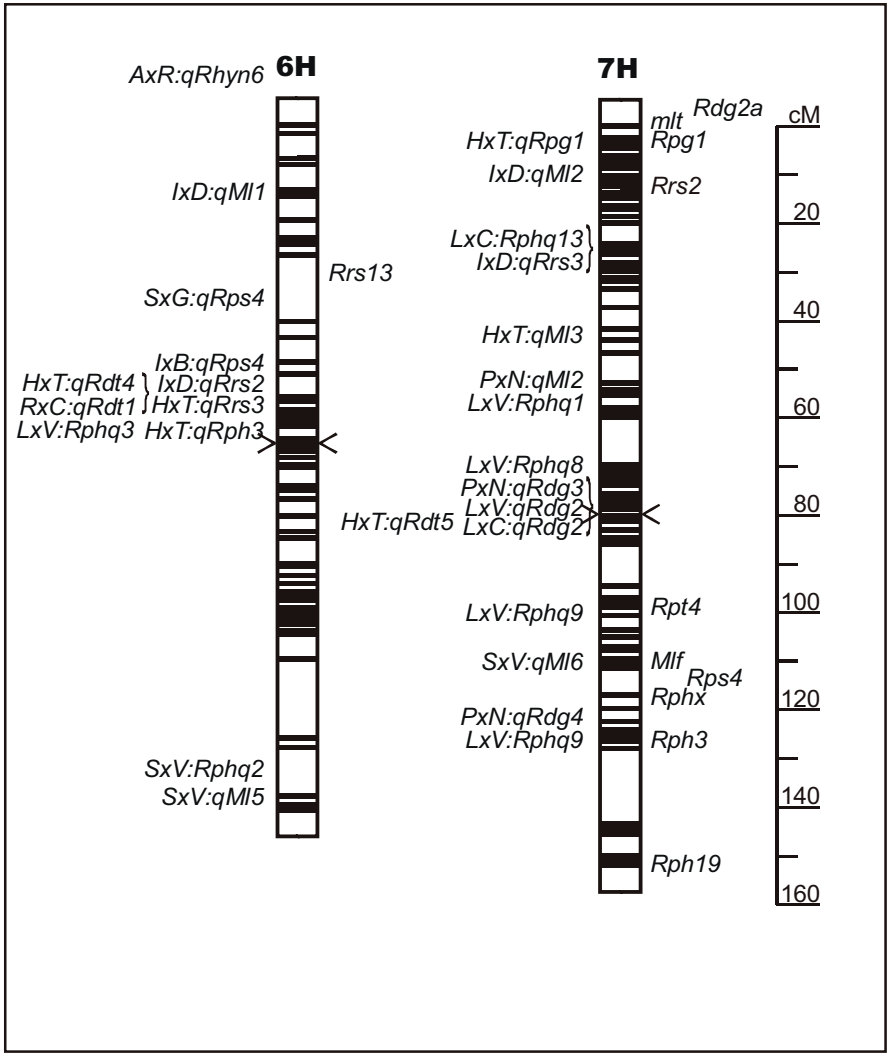


Figure 1. Continued

by Hossain and Sparrow (1991a; 1991b). Jensen (2002) placed the gene in tight linkage with the RFLP marker ABR377.

On chromosome 2H, only one powdery mildew resistance gene is known in barley. The *MlLa* locus, with a resistance originating from *Hordeum laevigatum* was localized in a cross 'Alf' × 'Vogelsanger Gold' (Hilbers, *et al.*, 1992), with 'Alf' bearing the resistant allele. The locus is flanked by the RFLP loci cMWG660 and MWG97.

Two resistance genes are localized on chromosome 4H: *Mlg* and *mlo*. *Mlg* was identified by Briggs and Stanford (1943) in the variety 'Goldfoil'. Fifty years later, Görg *et al.* (1993) found the gene to co–segregate with the RFLP marker MWG032 near the centromere on chromosome 4H. However, the more prominent resistance gene on 4H is *mlo*. It confers leaf lesion phenotype and broad–spectrum resistance. The resistance was found both in artificial mutants (Nover & Schwarzbach, 1971) and barley landraces from Ethiopia (Negassa, 1985). Hinze *et al.* (1991) found that the RFLP marker bAL88/2 was cosegregating with this resistance locus. The *mlo* gene was isolated in 1997 (Büschges, *et al.*). The corresponding 60–kDa protein is membrane–anchored and the resistance is based on a loss–of–function of this protein.

The powdery mildew resistance gene on 5HL, *Mlj* and the two genes on 7H, *mlt* and *Mlf* have been localized by Schönfeld *et al.* (1996) from wild barley lines (*H. vulgare* ssp. *spontaneum*). *Mlj* is flanked by the RFLP markers MWG592 and MWG999. On 5HL, the *mlt* locus is flanked by the RFLP loci MWG035 and MWG555a, while *Mlf* is mapped on 7HL, between the RFLP loci MWG053 and MWG539.

2.1.2. Leaf Rust (syn. Brown Rust, Caused by *Puccinia hordei*)

The leaf rust resistance gene on chromosome 1H was described *Rph4* (*Pa4*) by Roane and Starling (1967). They detected it in the varieties 'Gold' and 'Lechtaler'and localized it at a distance of 17 cM from the *Mla*–locus in a cross Gold × Cepada Capa (McDaniel & Hathcock, 1969).

On chromosome 2H, two leaf rust resistance genes were localised; (i) *Rph1* (*Pa*, *Pa1*) was found in the varieties 'Oderbrucker', 'Speciale' and 'Sudan' (Roane and Starling, 1967) and was localised by Tullen and McDaniel (1971). (ii) *Rph16* was localized from two wild barley (*H. vulgare* ssp. *spontaneum*) lines in a cross with the susceptible line 'L94' (Ivandic *et al*., 1998). The gene *Rph16* co–segregated with the RFLP markers MWG874 and MWG2133.

Three to four leaf rust resistance genes were reported on chromosome 3H. Zhong *et al.* (2003) localized *Rph6* on 3HS flanked by the RFLP markers and concluded, that the gene was allelic to the previously localized resistance gene *Rph5* (Mammadow, *et al.*, 2003) and closely linked to *Rph7*, 1.3 cM distal to the RFLP marker MWG691 (Graner *et al*., 2000; Brunner *et al*., 2000). The gene was first detected in the cultivar 'Cebada Capa' (Parlevliet, 1976). Another leaf rust resistance gene, *Rph10* on 3HL and linked with the isozyme locus *Est2* (Feuerstein *et al.*, 1990) was detected in near–isogenic

lines of wild barley (*H. vulgare* ssp. *spontaneum*) in a background of the variety 'Clipper'. It was mapped to the interval between the RFLP loci ABG495b and MWG838 (Collins *et al.*, 2001).

The two loci *RphQ* and *Rph2*, on 5HS, near the centromere, are either closely linked or allelic. They are localized in an interval between the RFLP markers CDO749 and ITS1 (Borovkova *et al.*, 1997). On 5HL, two more leaf rust resistance genes are localized; *Rph9* is flanked by the marker loci *HVDHN9* and *Est9*, and *Rph12* is flanked by the marker loci *va* and *ABC155* (Borovkova *et al.*, 1998).

Feuerstein *et al.* (1990) localized, together with the *Rph10* locus, another resistance gene originating from *H. vulgare* ssp. *spontaneum* and designated it as *Rph11*. It is linked to the isozyme loci *Acp3* and *Dip2*.

Three leaf rust resistance loci *Rphx*, *Rph3* and *Rph19* are all localized on 7H. *Rphx* was localized between the RFLP loci ABC310a and ABC461 (Hayes *et al*., 1996). *Rph3* was detected in the cultivar 'Estate' (Roane and Starling, 1967) at a distance of 9.7 cM from *Rphx* in a more distal position (Jin *et al.*, 1993), and *Rph19* was described in the cultivar 'Reka 1' between the SSR markers HVM49 and HVM11 (Park and Karakousis, 2002).

2.1.3. Stem Rust (syn. Black Rust, Caused by *Puccinia graminis*)

Two genes for resistance against stem rust are known, *Rpg1* on chromosome 7H between the two RFLP markers ABG704 and ABG312 (Kleinhofs *et al.*, 1993) and *rpg4* on chromosome 5H between the RFLP markers MWG740 and ABG390 (Borovkova *et al.*, 1995). *Rpg1* was isolated and identified as a receptor–kinase like protein (Brueggeman *et al.*, 2002). A transformation of the susceptible cultivar 'Golden Promise' with a construct containing the respective sequence resulted in a highly resistant genetically modified barley line, thus proving the function of the isolated gene (Horvath *et al.*, 2003).

2.1.4. Stripe Rust (syn. Yellow Rust, Caused by *Puccinia striiformis*)

Jensen and Jørgensen (1997) localised *Rps4* (*Yr4*) conferring resistance to stripe rust in the variety 'Abed Deba' on chromosome 1H close to the *Mla* locus. A more precise localisation was published later (Jensen, 2002). In an attempt to pyramid quantitatively and qualitatively inherited resistance genes against stripe rust, Castro *et al.* (2003a) mapped the qualitative gene *Rpsx* originating from the line 'C.I. 10587' to 7HL.

2.1.5. Scald (Caused by *Rhynchosporium secalis*)

The nomenclature of the scald resistance genes in barley was recently revised by Bjørnstad *et al.* (2002) as several symbols were earlier used for different alleles of the same locus. In this review, both old and new symbols are included.

Garvin *et al.* (1997) backcrossed different Iranian and Turkish wild barley lines and detected the scald resistance genes *Rrs14*. Later they localized the gene on chromosome 1H between the hordein loci *Hor1* and *Hor2* (Garvin, *et al.*, 2000).

Two genes conferring resistance to scald were localized on chromosome 3H: the complex locus *Rrs1* near the centromere and *Rrs4* about 20 cM more distal on the long arm of the chromosome. *Rrs1* was earlier described as *Rh*, *Rh1*, *Rh3* or *Rh4* (Dyck & Schaller, 1961; Habgood & Hayes, 1971) and *rh6* (Baker & Larter, 1963). The gene was mapped in the cross 'Igri' (susceptible) × 'Trition' (resistant) and co–segregated with the RFLP markers cMWG680, MWG582, ABG462, BCD263 and BCD828. *Rrs4* was earlier described as *Rhy* (Barua *et al.*, 1993) and *rrsx* (Patil, 2001). Barua *et al.* (1993) linked the gene to the RFLP–marker CDO1174 (6.8 cM) in a cross between the scald–susceptible cultivar 'Blenheim' and a scald–resistant SCRI breeding line. In the population 'Ingrid' (susceptible) × 'Nigrinudum' (C.I. 11549, resistant), Patil *et al.* (2003) mapped the gene between the two SSR loci HVM60 and HVM36b.

The scald resistance gene *Rrs13* was identified in a backcross line carrying resistance genes obtained from wild barley (*H. vulgare* ssp. *spontaneum*) and was mapped on chromosome 6H between the marker loci *Cxp3* and ABG458 (Abbott *et al*., 1995).

The scald resistance gene *Rrs2* (formerly *Rh2*) was described by Dyck and Schaller (1961) and localized by Schweizer *et al.* (1995) in 'Atlas' (C.I. 4118), crossed with the susceptible cultivar "Steffi". It co–segregated with the RFLP marker CDO545.

2.1.6. Barley (Leaf) Stripe (Caused by *Pyrenophora graminea*/ *Drechslera graminea*)

Barley stripe resistance gene *Rdg1a*, which is also known as 'Vada–resistance' was localized between the RFLP loci MSU21 and *Xris45b* on 2HL using the cross 'Alf' (*Rdg1a*) × 'Vogelsanger Gold' (*rdg1a*) (Thomsen *et al,*

1997). The other gene *Rdg2a* for resistance against the same disease was identified in the cultivar 'Thibaut' and mapped 2.5 cM distally tothe RFLP marker MWG2018 on 7HS (Tacconi *et al*., 2001).

2.1.7. Net Blotch (Caused by *Pyrenophora teres/ Drechslera teres*)

Two resistance genes acting against *Pyrenophora teres* have been localized in barley until now. One, preliminarily named *Pt,,a* was mapped to chromosome 3H, between the RFLP markers BCD828 and MWG2138, using the cross 'Igri' (resistant) × 'Franka' (susceptible) (Graner *et al.*, 1996). The other gene, *Rpt4,* originated from the cultivar 'Galleon' and provided resistance against the sport form of net blotch. This was localized on chromosome 7H using a DH–population derived from a cross with the susceptible cultivar 'Haruna Nijo' (Williams *et al.,* 1999).

2.2. Quantitative Inherited Genes

In contrast to the qualitatively inherited genes, the quantitatively inherited genes are presented in the historical order of the QTL analyses, starting with the earliest publications. The QTLs are presented in the map (Fig. 1) in the form cross:QTL. If a name of the QTL is specified in the respective publication, it is used; else, a name is composed by 'q', the symbol of the disease as used for qualitatively inherited traits and a running number.

2.2.1. Powdery Mildew (Caused by *Blumeria graminis* f.sp. *hordei*)

The first QTL localisation for powdery mildew resistance in barley was performed by Heun (1992) in a doubled haploid (DH–) population from the cross 'Proctor' × 'Nudinka'. He localised one QTL on chromosome 5H (*P*x*N:qMl1*) and the other on chromosome 7H (*P*×*N:qMl2*). Together, they accounted for 19.8% of the observed variance.

Backes *et al.* localised QTLs in a DH–population resulting from the cross 'Igri' × 'Danilo' based on field data (1995) and compared them with data form detached primary leaves (1996). One QTL on 6H (*I*×*D:qMl1*) was only detected only in the field, while the other QTL on 7H (*I*×*D:qMl2*) was detected both on detached leaves and in the field.

Thomas *et al.* (1995) detected a QTL at the position of the *Mla* locus on chromosome 1H (*B×E:qMl1*) in a DH–population from the cross 'Blenheim' × 'E224/3' and explained it by the action of *Mla13*, present in the SCRI line 'E224/3'.

Three QTLs for powdery mildew resistance were detected in a DH–population based on the cross 'Harrington' × 'TR306' (Spaner *et al.*, 1998; Falak *et al.*, 1999). One was localised on chromosome 4H at the position of the qualitative resistance locus *Mlg* (*H×T:qMl1*), another on chromosome 5H (*H×T:qMl2*), and the last, weaker one, on chromosome 6H (*H×T: qMl3*).

In a segregating population of recombinant inbred (RI–) lines from a cross of the wild barley (*H. vulgare* ssp. *spontaneum*) line '1B–87' and the barley cultivar 'Vada', Backes *et al.* (2003) localised six different QTLs for resistance against powdery mildew. Two of those QTLs were localised on positions of qualitatively inherited powdery mildew resistance loci: one QTL on chromosome 1H (*S×V:qMl1*) at the position of the *Mla* locus and one on 7H (*S×V:qMl5*) at the position of the *Mlf* locus. Two QTLs were detected at positions where resistance–gene analogs (RGAs) were localised in the same cross: one on chromosome 2H (*S×V:qMl2*) and one on 3H (*S×V:qMl3*). Two additional QTL were detected on chromosome 4H (*S×V:qMl4*) and 6H (*S×V:qMl5*) .

2.2.2. Leaf Rust (Brown Rust, Caused by *Puccinia hordei*)

Thomas *et al.* (1995) detected one QTL for resistance against leaf rust (brown rust) in the cross 'Blenheim' × 'E224/3' on chromosome 5H (*B×E:qRph1*). In the already mentioned DH–population from 'Harrington' × 'TR306', Spaner *et al.* (1998) detected three QTLs for resistance against leaf rust. For all of them, the 'TR306'–allele conferred resistance. One of the QTLs (*H×T:qRph1*) was localised on chromosome 2H, one on 5H (*H×T:qRph2*) and a final one on chromosome 6H (*H×T:qRph3*). They detected a significant interaction between the QTL on chromosome 2H and the one on 6H. According to the LOD–curve published, there is a second peak on chromosome 6H, nearly as high as the first one. The respective QTL (*H×T:qRph4*) would be at the same position as the qualitative leaf rust resistance genes *Rph9* and *Rph12*.

Qi *et al.* (1998a, 1999) localised ten QTLs for resistance against leaf rust in an RI–population derived from the cross between 'L94', an extremely susceptible Ethiopian landrace, and the cultivar 'Vada'. The linkage map was

based exclusively on AFLP markers (Qi *et al.*, 1998b). On chromosome 7H, they localised three QTLs (*L×V:Rphq1*, *L×V:Rphq8*; *L×V:Rphq9*). On the chromosomes 2H, 4H and 5H, two QTLs were detected, respectively (2H: *L×V:Rphq6*, *L×V:*Rphq2; 4H: *L×V:Rphq10*, *L×V:Rphq5*; 5H: *L×V:Rphq4*, *L×V:Rphq7*). Finally, one QTL was mapped to chromosome 6H (*L×V:Rphq3*). Caused by lack of common markers and inconsistencies between the maps, an unambiguous placement of *L×V:Rphq4* and *L×V:Rphq5* had not been possible in Figure 1. Therefore, these markers are set in parentheses. Disease resistance was tested both in seedling stage and on adult plants and with two different isolates. Only one QTL (*L×v:Rphq2*) was effective both in seedling stage and in adult plants. Three QTLs were effective only in seedling stage (*L×V:Rphq1; LV:Rphq3; L×V:Rphq7*), the rest only in adult plants. Expression of the resistance showed quantitative isolate specificity, spanning from nearly no difference (*L:V:Rphq2* in adult plants) to very strong differences (*L:V:Rphq7* and *L:V:Rphq8* in adult plants). The same group also examined a second cross, 'L94' × 'Cebada Capa'. Here they confirmed two of the QTLs detected in 'L94' × 'Vada' (*L×V:Rphq3, L×V: Rphq10*), but detected three additional QTLs, two on chromosome 2H (*L×C:Rphq11*, *L×C:Rphq12*) and one on 7H (*L×C:Rphq13*). Only *Rphq3* was effective both in seedlings and adult plants. For some of the QTLs in 'L94' × 'Vada', near–isogenic lines (NILs) were produced by marker–assisted backcrossing (van Berloo *et al.*, 2001). Two NILs for *Rphq2* were tested and showed a degree of resistance significant higher than 'L94'.

In a DH–population from the cross 'Krona' × 'HOR1063', Kicherer *et al.* (2000) localised four QTL for resistance against leaf rust. Three of these four QTLs were mapped to chromosome 2H (*K×H:Rphq1*, *K×H:Rphq2*, *K×H:Rphq3*), one to 4H (*K×H:Rphq4*). About one third of the explained variance was contributed by QTL × QTL interactions in this experiment.

Backes *et al.* (2003) localised two QTL for resistance against leaf rust in the above mentioned RI–population from the cross '1B–87' × 'Vada'. One of the QTLs was mapped to 2HL (*S×V:qRph1*) and the other on 6HL (*S×V:qRph2*). Both QTLs were localised close to RGAs in this cross.

2.2.3. Stem Rust (Black Rust, Caused by *Puccinia graminis*)

In a DH–population from 'Harrington' × 'TR306', Spaner *et al.* (1998) detected two QTLs for resistance against stem rust. One of these QTLs (*H×T:qRph1*) was mapped to the position of the qualitative stem rust

resistance gene *Rpgq1* on chromosome 7H, and the other one on chromosome 4H (*H×T:qRpg2*).

2.2.4. Stripe Rust (Yellow Rust, Caused by *Puccinia striiformis*)

Chen *et al.* (1994) localised two QTLs for resistance against barley stripe rust in a DH–population from a cross between a resistant ICARDA/CIMMYT line and the cultivar 'Bowman'. A minor QTL was detected on chromosome 4H (*I×B:qRps1*), a major one on 5H (*I×B:qRps2*). In the same population, Hayes *et al.* (1996) detected—based on field results—the major QTL on 5H, but not the minor QTL on 4H, as the respective peak fell below detection thresholds. Nevertheless, based on seedling data under controlled conditions, the QTL on 4H was approved. Additionally, they detected a further QTL on chromosome 5H (*I×B:qRps3*) based on field data and an additional QTL on chromosome 6H based on seedlings data (*I×B:qRps4*). The QTLs on 4H and on 5H were introgressed into the cultivar 'Steptoe' by backcrossing and production of DH–lines from the BC_1–lines (Toojinda *et al.*, 1998). The effect of the respective QTLs could be confirmed in those DH–lines.

Toojinda *et al.* (2000) detected QTLs for stripe rust resistance in a DH–population from the cross 'Shyri' × 'Galena' on the chromosomes 1H, 2H, 3H and 6H (*S×G:qRps1*, *S×G:qRps2*, *S×G:qRps3* and *S×G:qRps4*, respectively). The disease was tested in the field. The same population was tested with three different isolates (Castro *et al.*, 2002) in seedling stage and the QTLs on 1H and 6H were confirmed with all of them.

In an attempt to pyramid quantitative resistance genes, Castro *et al.* (2003b), crossed the cultivars 'Orca' and 'Harrington' and then again with the line 'DI–72'. In the resulting DH–population, they localised three different QTLs for resistance against stripe rust on the chromosomes 1H (*2×D:qRps1*), 4H (*O×H:qRps1*) and 5H (*O×H:qRps2*). The QTLs on 4H and 7H segregated in the first cross, 'DI–72' contributed the QTL on 1H. They further crossed lines with the qualitative resistance gene *Rpsx* with lines resulting from the QTL pyramiding to combine qualitative and quantitative resistance genes in one line (Castro *et al.*, 2003a). A further QTL on 4H was mapped in the resulting populations (*Cpx:qRps1*).

2.2.5. Scald (Caused by *Rhynchosporium secalis*)

Three QTLs for resistance against scald were localised by Backes *et al.* (1995), on chromosome 2H (*IxD:qRrs1*), 6H (*IxD:qRrs2*) and 7H

(*IxD:qRrs3*), respectively. The QTL on chromosome 2H had the largest effect of those three. In the above mentioned DH–population from the cross 'Harrington' × 'TR306', Spaner *et al.* (1998) detected three QTLs. They were localised on chromosome 3H (*HxT:qRrs1*), 4H (*HxT:qRrs2*) and 6H (*HxT:qRrs3*). Jensen *et al.* (2002) mapped three QTLs in a DH–population of the cross 'Alexis' × 'Regatta' to the chromosomes 3H (AxR:Qryn3), 4H (AxR:Qryn4) and 6H (Qryn6), respectively.

2.2.6. Barley (Leaf) Stripe (Caused by *Pyrenophora graminea/ Drechslera graminea*)

In a DH–population derived from the cross 'Proctor' × 'Nudinka', Pecchioni *et al.* (1996) localised two major and two minor QTLs. The two major QTLs were detected on chromosome 2H (*PxN:qRdg1*) and 7H (*PxN:qRdg3*); the two minor QTLs on chromosome 4H (*PxN:qRdg2*) and 7H (*PxN:qRdg4*). They also tried to analyse, if known PR genes or qualitative resistance genes co–localised with those QTLs (Pecchioni, *et al.*, 1999). They found the *RsmMx* gene acting against BSMV to be at or near the position of the major QTL on chromosome 7H (*PxN:qRdg3*).

Arru *et al.* (2002) localized three QTLs for resistance against barley leaf stripe in two RI–populations, one derived from the cross 'L94' × 'Vada' and one derived from the cross 'L94' × 'C123'. In the cross with 'Vada', one major QTL was localised on chromosome 2H (LxV:qRdg1) and a minor one on 7H (LxV: qRdg2). In the population with 'C123', a QTL was found on 7H at the same position as in the population with the 'Vada'–parent (LxC:qRdg2), and an additional QTL was found on 2H (LxC:qRdg1), different from the one in the 'Vada'–population.

A DH–population from the cross 'Steptoe' × 'Morex', was examined by Arru *et al.* (2003) with two different isolates in order to localise QTLs for resistance against barley leaf stripe. One QTL on chromosome 2H (*SxM:qRdg2*) was common for both isolates, two QTLs on 3H were linked (*SxM:qRdg3*, *SxM:qRdg4*), a further QTL on chromosome 2H (*SxM:qRdg1*) and a QTL on 5H (*SxM:qRdg5*) were isolate–specific.

2.2.7. Net Blotch (Caused by *Pyrenophora teres/ Drechslera teres*)

Five QTLs for resistance against net blotch were detected by Spaner *et al.* (1998) using a DH–population from the cross 'Harrington' × 'TR306'. The

QTLs were mapped to the chromosomes 3H (*HxT:qRdt1*), 4H (*HxT:qRdt2*), 5H (*HxT:qRdt3*), 6H (*HxT:qRdt4*) and 7H (*HxT:qRdt5*).

Richter *et al.* (1998) also detected 12 QTLs in an F_2–population derived from a cross between the susceptible cultivar 'Arena' and the resistant Ethiopian lancrace 'HOR 9088'. They scored the disease on the first and second leaves, in each case twice, once 7 days and second 9 days after inoculation. For each of the four sets of data, they detected three QTLs as follows: (i) for the scoring of the first leaf 7 days after infection (d.a.i.), 2 QTLs on chromosome 6H (*AxH:qRdt9, AxH:qRdt12*) and one QTL on 4H (*AxH:qRdt8*) were detected; (ii) for the scoring of the second leaf 7 d.a.i., 2 QTLs on chromosome 6H (*AxH:qRdt10, AxH:qRdt11*) and 1 QTL on 3H (*AxH:qRdt6*) were localised; (iii) for the scoring of the first leaf 9 d.a.i., they mapped 2 QTLs to chromosome 2H (*AxH:Rdt2, AxH:qRdt3*) and one to 1H (*AxH:qRdt1*); and (iv) finally, for the second leaf 9 d.a.i., two QTLs on chromosome 3H (*AxH:qRdt4, AxH:qRdt5*) and one on 4H (*AxH:qRdt7*) were found. Due to non–availability of common markers, the placement of *AxH:qRdt3* and *AxH:qRdt7* on the map in Figure 1 cannot be precise. Therefore, the names are set in parentheses. The situation was even worse for the four QTLs on 6H. Therefore, *AxH:qRdt8*, *AxH:qRdt9*, *AxH:qRdt10* and *AxH:qRdt11* were not placed on the map. In the original publication, all the four loci are placed tightly linked in the centre of the linkage map.

A solitary major QTL was also localised using a DH population derived from a cross between the susceptible cultivar 'Rolfi' and the resistant line 'C.I.8919'. The QTL was mapped to 6HL (*RxC:qRdt1*) and explained 65% of the phenotypic variance for this trait (Manninen *et al.*, 2000).

3. GENES FOR DISEASE RESISTANCE IN WHEAT (*TRITICUM AESTIVUM* L.)

Wheat (*Triticum aestivum* L.) is influenced by a number of diseases, mainly caused by fungal pathogens. Important pathogens are the biotrophic fungi causing the leaf and stripe rust diseases and powdery mildew and the necrotrophic fungi causing a number of blotches on leaves and ear *e.g.* septoria tritici blotch and fusarium. In the following, the emphasis is put on markers for genes and QTLs for resistance to septoria tritici blotch (Table 1), and fusarium head blight (Table 2). In addition, markers for powdery mildew (Table 3), stripe rust (Table 4) and leaf rust (Table 5) resistance genes and QTLs are listed. Both single genes and genes located by QTL mapping methods are shown in the tables. The closest marker(s) with the linkage

distance between gene and marker in parentheses is shown. For QTLs, when possible, markers flanking the QTL with the interval length in cM are shown. When the linkage distance between marker and gene is given as 0 cM, either no recombinants between the gene and the marker was detected in a mapping population, or the marker and the gene are located on an alien chromosome segment that is not expected to recombine with the wheat genome chromatin.

Due to the hexaploid nature of wheat with three genomes, many essential genes are triplicated, and consequently wheat tolerates the loss and gain of whole chromosomes or parts of chromosomes (Law *et al*., 1987). This has made possible the introgression of genes, especially resistance genes, from many alien sources into the wheat genome (see Tables 3, 4 and 5; Rajaram *et al*., 2001). Recombination between alien translocated segments and wheat chromosomes is mostly absent or severely restricted (Gale and Miller, 1987). This simplifies the task of finding linked markers for resistance genes residing on these translocations in wheat, as a marker in the translocation will remain associated with the gene(s) of interest through consecutive crosses even if the physical distance from the gene to marker is large (Procunier *et al*., 1995). Further, the high polymorphism between the introgressed segment and the wheat genome should make markers easier to identify. However, the introgression of a piece of alien DNA often introduces undesirable agronomic characters from the alien source (Knott, 1989). For example the widely used 1BL/1RS wheat – rye translocation is associated with reduced bread making quality (Graybosch *et al*., 1993). These disadvantages have to be eliminated through breeding and obviously the absence or reduction in recombination between alien and wheat DNA makes this task difficult.

3.1. Septoria Tritici Blotch

During the past 20–30 years septoria tritici blotch caused by the fungus *Mycosphaerella graminicola* has grown from being a minor disease to a dominating disease in Europe (Hardwick *et al*., 2001). The disease is favoured by a wet environment, which facilitates sporulation from pycnidia, the rain increasing the chances of dispersal of the pycnidiospores (Shaw and Royle, 1993). The fungus also has sexual structures in the form of pseudothecia producing airborne ascospores; these are believed to have greatest importance as primary inoculum in newly established wheat crops (Eriksen *et al*., 2001; Eriksen and Munk, 2003). Field resistance is expressed quantitatively, however when inoculating with specific isolates in the field, specific resistance has clearly been shown (Kema and van Silfhout, 1997; Brown *et al*., 2001). Eight resistance genes have been designated *Stb1*, *Stb2*, *Stb3*, *Stb4*, *Stb5, Stb6, Stb7* and *Stb8* (Wilson, 1985; Somasco *et al*., 1996;

Arraiano *et al* 2001; Brading *et al*., 2002; McCartney *et al*., 2003; Adhikari *et al*., 2003). The genes *Stb5, Stb6, Stb7* and *Stb8* have been mapped with molecular markers (Table 1). The relationship between *Stb4* and *Stb6* as well as the relationship between *Stb1, Stb2* and *Stb3* and the other designated genes is unclear. The *Stb4* gene was derived from the spring wheat cultivar 'Tadinia' and has been found to be effective against the Californian Stb population for a number of years (Somasco *et al*., 1996). However, a recent report suggests that this resistance has become ineffective in the Willamette valley of Oregon, as *Stb4* is believed to be present in the cultivar 'Gene' that suffered increasing levels of Stb from its introduction until it was finally considered susceptible (Cowger *et al*., 2000). Based on avirulence to isolate IPO323 the gene *Stb6* appears to be present in some European cultivars (L. Eriksen, unpublished). However, virulence was frequent in the pathogen population in a sample of isolates collected in Denmark during 1997 (Eriksen *et al*., 2003a); in this case, the resistance has to be considered broken. The *Stb5* gene is derived from the wild D genome progenitor of wheat, *Aegilops tauschii*, through the synthetic hexaploid wheat 'Synthetic 6*x*'. This gene is thus unlikely to be present in any commercial cultivars. Few attempts have been made to map genes responsible for resistance to Stb using QTL mapping methods. In a recent study, a doubled haploid population from a cross between the winter wheat cultivars 'Savannah' and 'Senat' was used to map QTLs in 'Senat' for resistance to *M. graminicola* (Eriksen *et al*., 2003a). Composite interval mapping was performed on AUDPC (Area Under Disease Progress Curve) from two field trials. Four QTLs namely, *QStb.risø–2B*, *QStb.risø–3A.2*, *QStb.risø–6B.2* and *QStb.risø–7B* provided resistance in the field over two environments. The QTL on 3A was either pleiotropic for plant height or was closely linked to a QTL for plant height. The QTL on 6B explained a very high proportion of the phenotypic variance and was flanked by the same two SSR markers in both trials. A genetic distance of only 3 cM separated these markers. Some of these QTLs (e.g. the 6B QTL) showed an effect in the field as well as against specific isolates in growth chamber, whereas others did not. Other QTLs that were detected only in growth chamber included *QStb.risø–3A.1*, *QStb.risø–3B* and *QStb.risø–6B.1*.

3.2. Fusarium Head Blight

Fusarium head blight (Fhb) is caused by a number of fungal species in the group *Fusarium*. The most important pathogens causing Fhb of wheat is *Fusarium avenaceum*, *F. culmorum*, *F. graminearum*, *F. poa* and *Microdochium nivale* (Parry *et al*., 1995). Resistance to Fhb is believed to be equally effective to these species (van Eeuwijk *et al*., 1995). Recently, molecular markers for Fhb were reviewed (Kolb *et al*., 2001). Since then a

Table 1. Genes and QTLs for resistance to septoria tritici blotch mapped using molecular markers in bread wheat

Gene / QTL	Chr. arm	Nearest marker(s)/ marker interval[a]	Marker type[b]	R^2 (%)[c]	Material used for mapping/ marker development	Reference
Stb5	7DS	*Xgwm44* (7.2 cM)	SSR	–	'Synthetic 6*x*' (Synthetic hexaploid)	Arraiano *et al*. (2001)
Stb6	3AS	*Xgwm369* (2 cM)	SSR	–	'Flame'	Brading *et al*. (2002)
Stb7	4AL	*Xwmc219* (1.2 cM)	SSR	–	'Estanzuela Federal'	McCartney *et al*. (2003)
Stb8	7BL	*Xgwm146* (3.5 cM)	SSR	–	'W7984' (Synthetic hexaploid)	Adhikari *et al*.(2003)
QStb.risø–2B	2BL	*Xwmc175a – Xwmc175a* (16 cM)	SSR	25.9	'Senat'	Eriksen *et al*. (2003a)
QStb.risø–3A.1	3AS	*Xgwm369*	SSR	–	'Senat'	Eriksen *et al*. (2003a)
QStb.risø–3A.2	3AS	*Xwmc489 – Xwmc505* (5 cM)	SSR	18.2	'Senat'	Eriksen *et al*.(2003a)
QStb.risø–3B	3BL	*M62/P38–373*	AFLP	–	'Senat'	Eriksen *et al*. (2003a)
QStb.risø–6B.1	6B	*M48/P32–112*	AFLP	–	'Senat'	Eriksen *et al*. (2003a)
QStb.risø–6B.2	6B	*Xwmc397 – Xwmc341* (3 cM)	SSR	67.9	'Senat'	Eriksen *et al*. (2003a)
QStb.risø–7B	7B	*M49/P38–229 – M49/P11–229* (22 cM)	AFLP	12.2	'Senat'	Eriksen *et al*. (2003a)

[a] genetic distance marker to gene or distance between markers in marker intervals is shown in parentheses
[b] AFLP = Amplified Fragment Length Polymorphism, SSR = Simple Sequence Repeat
[c] proportion of phenotypic variance explained by the investigated QTL locus

number of reports of markers for resistance to this disease have been published. Resistance to fusarium head blight is of a quantitative nature and genes for Fhb resistance have been mapped using QTL analysis. Most studies mapping resistance to Fhb used the spring wheat 'Sumai 3', of Chinese origin, or 'Sumai 3' derived material. 'Sumai 3' was selected from a cross

between the moderately susceptible Italian cultivar 'Funo' and the moderately susceptible Chinese cultivar 'Taiwan Wheat' (Bai and Shaner, 1994; Shen *et al*., 2003b). In studies with materials involving 'Sumai 3', a very consistent QTL on chromosome 3BS has been detected (*QFhs.ndsu–3B*), explaining 15.4% – 60.1% of the phenotypic variance, in the various studies (Table 2). In 'Sumai 3', the most likely source of *QFhs.ndsu–3B* is 'Taiwan Wheat' (Shen *et al*., 2003b). Waldron *et al*. (1999) and Anderson *et al*. (2001) used mapping populations involving 'Sumai 3' directly. Others have used mapping populations where the resistant parent had 'Sumai 3' in their pedigree (Anderson *et al*., 2001; Buerstmayr *et al*., 2002, 2003; Zhou *et al*., 2002; Boudoncle and Ohm, 2003; Guo *et al*., 2003). The Chinese line 'Ning 894037' apparently did not have 'Sumai 3' in its pedigree (Shen *et al*., 2003b) and the pedigree of the Chinese line 'Huapei 57–2' is not known (Bourdoncle and Ohm, 2003); however, based on SSR data it was concluded that both lines probably carry *QFhs.ndsu–3B* (Bourdoncle and Ohm, 2003; Shen *et al*., 2003b). Consequently, the effect of *QFhs.ndsu–3B* from 'Sumai 3' is well investigated in different genetic backgrounds, and a number of easy to use SSR and STS markers are available for this QTL (Table 2). The resistance of 'Sumai 3' is apparently the result of transgressive segregation in the progeny of two moderately susceptible parents, and may be the result of a combination of QTLs. Therefore, the presence of *QFhs.ndsu–3B* is not necessarily enough to ensure resistance in the progeny of a cross, and other QTLs should be selected simultaneously (Zhou *et al*., 2003). Additional QTLs in 'Sumai 3' or its derivatives, and in other material carrying *QFhs.ndsu–3B* have been mapped on chromosomes 1B and 5A, and the chromosome arms 2BL, 3AL, 3AS, 6AS and 6BS (Table 2). It should thus be possible to select several QTLs with markers when using these lines in breeding. Not surprisingly several QTLs have been detected where the susceptible parent of the mapping population contributes the resistance allele, *i.e*. 'Stoa' in 'Sumai 3'/'Stoa' (Waldron *et al*., 1999; Anderson *et al*., 2001), 'Alondra' in 'Ning 894037'/'Alondra' (Shen *et al*., 2003b) and 'Patterson' in 'Huapei 57–2'/'Patterson' (Bourdoncle and Ohm, 2003). A number of QTLs for Fhb resistance have been mapped in cultivars/lines unrelated to 'Sumai 3'; this is in the cross between the French cultivars 'Renan' and 'Récital' (Gervais *et al*., 2003), in the Romanian wheat cultivar 'Fundulea 201R' (Shen *et al*., 2003a) and in tetraploid wheat (Otto *et al*., 2002). Finally, Somers *et al*. (2003) mapped QTL for both resistance and fusarium toxin (Deoxynivalenol, DON) accumulation in a cross between the Chinese wheat 'Wuhan–1' and 'Maringa'. 'Maringa' has the Fhb resistant Brasilian wheat 'Frontana' in its pedigree. QTLs were found on 3BS and 5AS at the same location as *QFhs.ndsu–3B* and *QFhs.ifa–5A*, respectively.

Table 2. **QTLs for resistance to Fusarium mapped using molecular markers**

QTL[a]	Chromosome	Nearest marker(s)/ marker interval[b]	Marker type[c]	R^2 (%)[d]	Material used for mapping/ marker development	Reference
–	1B	*XgluB1*	Glutenin	9.0–9.7	'CM–82036' ('Sumai 3'/'Thornbird S')	Buerstmayr *et al*. (2002)
–	1B	*Xbarc8* – *Xgwm131* (14.6cM)	SSR	18.7	'Fundulea 201R'	Shen *et al*. (2003)
QFhs.ndsu–2A	2AL	*XksuH16*	RFLP	14.3	'Stoa'	Waldron *et al*. (1999)
QFhs.inra–2A	2A	*Xgwm311b* *Xgwm382c*	SSR	6.4–14.4	'Renan'	Gervais *et al*. (2003)
–	2BL	*Xgwm120*	SSR	4–7	'Ning 7840' ('Aurora'/'Anhui 11'//'Sumai 3')	Zhou *et al*. (2002)
QFhs.inra–2B	2B	*Xgwm374*	SSR	8.5–12	'Renan'	Gervais *et al*. (2003)
–	2DS	*Xgwm261*– *Xgwm296* (8.3cM)	SSR	12.1	'Alondra'	Shen *et al*. (2003b)
–	2DL	*Xgwm539*	SSR	9	'Wuhan'	Somers *et al*. (2003)
–	3AL	*Xbcd941*	RFLP	9.1	'ND2603' ('Sumai 3'/'Wheaton')	Anderson *et al*. (2001)
–	3AS	*Xgwm5*	SSR	8.1	'Huapei 57–2'	Bourdoncle and Ohm (2003)
QFhs.ndsu–3AS	3AS	*Xgwm2*	SSR	37	*T. dicoccoides* (tetraploid wheat relative)	Otto *et al*. (2002)
QFhs.inra–3A	3A	*Xbcd0372*	RFLP	4–6.2	'Récital'	Gervais *et al*. (2003)

Table 2. Continued

–	3AS	*Xgwm674* – *Xbarc67* (2.8cM)	SSR	13	'Fundulea 201R'	Shen *et al*. (2003a)
QFhs.ndsu–3B	3BS	*Xcdo981*	RFLP	15.4	'Sumai 3'	Waldron *et al*. (1999); Liu and Anderson (2003)
	3BS	*Xgwm493*	SSR	24.8	'Sumai 3'	Anderson *et al*. (2001); Liu and Anderson (2003)
	3BS	*Xgwm533*	SSR	41.6	'ND2603' ('Sumai 3'/ 'Wheaton')	Anderson *et al*. (2001)
	3BS	*Xgwm533*– *Xgwm493* (ca. 8cM)	SSR	28.6–60.1	'CM–82036' ('Sumai 3'/ 'Thornbird S')	Buerstmayr *et al*. (2002, 2003)
	3BS	*Xgwm533* *Xbarc147*	SSR	21–52 18–52	'Ning 7840' ('Aurora'/'Anhui 11'//'Sumai 3')	Zhou *et al*. (2002)
	3BS	*Xbarc133*	SSR	23.6	'Huapei 57–2' (probably same QTL as in 'Sumai 3')	Bourdoncle and Ohm (2003)
	3BS	*SRST.3B1*	STS	50	'Ning 7840' ('Aurora'/'Anhui 11'//'Sumai 3')	Guo *et al*. (2003)
	3BS	*Xbarc133*– *Xgwm493* (10.6cM)	SSR	42.5	'Ning 894037', probably the same QTL as in 'Sumai 3'	Shen *et al*. (2003b)
–	3BS	*Xgwm533*	SSR	13	'Maringa'	Somers *et al*. (2003)

Table 2. Continued

– (DON accumulation)	3BS	*Xgwm533*	SSR	11	'Maringa'	Somers *et al*. (2003)
–	3BS	*Xgwm566*	SSR	4	'Maringa'	Somers *et al*. (2003)
–	3BL	*Xgwm247*	SSR	10.7	'Huapei 57–2'	Bourdoncle and Ohm (2003)
QFhs.inra–3B	3B	*Xtam61, Xgwm131b, Xgwm383b*	SSR	5.4–10.5	'Renan'	Gervais *et al*. (2003)
–	4BS	*Xwg909*	RFLP	7.2	'Stoa'	Anderson *et al*. (2001)
–	4BS	*Xwmc238*	SSR	12	'Wuhan'	Somers *et al*. (2003)
QFhs.ifa–5A	5A	*Xgwm293–Xgwm304* (ca. 3cM)	SSR	7.5–23.2	'CM–82036' ('Sumai 3'/ 'Thornbird S')	Buerstmayr *et al*. (2002; 2003)
QFhs.inra–5A.1	5A	*Xgwm443, Xpsr0170a*	SSR	5–6.2	'Renan'	(Gervais *et al*., 2003)
QFhs.inra–5A.2	5A	*Xgwm639b*	SSR	14–19.2	'Renan'	Gervais *et al*. (2003)
QFhs.inra–5A.3	5A	*B1*	Awns	8.5–19.2	'Renan'	Gervais *et al*. (2003)
– (DON accumulation)	5AS	*Xgwm96*	SSR	6	'Maringa'	Somers *et al*. (2003)
–	5BL	*Xbarc59*	SSR	7.1	'Patterson'	Bourdoncle and Ohm (2003)
QFhs.inra–5D	5D	*Xcfd0029*		–	'Renan'	Gervais *et al*. (2003)
–	6AS	*XksuH4*	RFLP	11.6	'ND2603' ('Sumai 3'/ 'Wheaton')	Anderson *et al*. (2001)
–	6BS	*Xbcd331*	RFLP	6	'Sumai 3'	Waldron *et al*. (1999)

Table 2. Continued

–	6BS	*Xcdo524*	RFLP	3.9	'Sumai 3'	Waldron *et al*. (1999)
–	6BS	*Xbarc101*	SSR	9.2	'Sumai 3'	Anderson *et al*. (2001)
–	6BS	*Xbcd1383*	RFLP	4.9	'ND2603' ('Sumai 3'/'Wheaton')	Anderson *et al*. (2001)
–	6BS	*Xgwm644* – *Xgwm88* (6.4cM)	SSR	4.4	'Ning 894037'	Shen *et al*. (2003b)
QFhs.inra–6D	6D	*Xcfd0042*		6.6	'Renan'	Gervais *et al*. (2003)

[a] – no name provided for gene/QTL in reference
[b] genetic distance marker to gene or distance between markers in marker intervals is shown in parentheses
[c] RFLP = Restriction Fragment Length Polymorphism, STS = Sequence Tagged Site, SSR = Simple Sequence Repeat
[d] proportion of phenotypic variance explained by the investigated QTL locus

Laboratory protocols for MAS of the 'Sumai 3' derived QTLs *QFhs.ndsu – 3B* and the *T. dicoccoides* derived QTL *QFhs.ndsu–3AS* can be found at http://maswheat.ucdavis.edu.

3.3. Powdery Mildew

As many as 27 genes/loci (five of them with multiple alleles) for resistance to wheat powdery mildew, caused by *Blumeria graminis* f.sp. *tritici*, have been described and were recently reviewed by Hsam and Zeller (2002). For powdery mildew resistance, QTLs have also been described. Keller *et al*. (1999) found 18 QTLs in a cross between the wheat 'Forno' and the spelt 'Oberkulmer'; two of these QTLs(Table 3) were consistent over five environments and over two years. One of these consistent QTLs coincided with the *Pm5* gene on chromosome 7BL in the wheat parent, even though isolates virulent on *Pm5* were present in the trial (Keller *et al*., 1999). Chantret *et al*. (2001) obtained similar results at the *MlRE* locus; a QTL providing resistance against a virulent mildew population was detected at this position. The most consistent QTL over environments and growth stages detected by Chantret *et al*. (2001) was located on chromosome 5D.

Table 3. Genes and QTLs for resistance to powdery mildew mapped using molecular markers

Gene/ QTL[a]	Chromo-some	Nearest marker(s)/ marker interval [b]	Marker type[c]	R^2 (%)[d]	Material used for mapping/ marker development	Reference
Pm1	7AL	*Xcdo347* (0 cM)	RFLP	–	'Axminster'/8* 'Chancellor'	Ma *et al*. (1994)
	7AL	*Whs178* (3 cM)	RFLP	–	'Axminster'/8* 'Chancellor'	Hartl *et al*. (1995)
	7AL	*C320* (0 cM) *C638* (0 cM)	STS	–	'Zhengzhou 871124'	Hu *et al*. (1997)
Pm1c	7AL	*18M2* (0.9 cM) specific for *Pm1c*	AFLP	–	'Weihenstephan Stamm M1N'	Hartl *et al*. (1999)
Pm1e (formerly Pm22)	7AL	*Xgwm344* (0.9 cM)	SSR	–	'Virest'	Singrün *et al*. (2003)
Pm2	5DS	*Xbcd1871* (3.5 cM)	RFLP	–	'Ulka'/8* 'Chancellor'	Ma *et al*. (1994)
	5DS	*Xws295* (2.7 cM)	RFLP	–	'Ulka'/8* 'Chancellor'	Hartl *et al*. (1995)
Pm3a, b, c	1AS	*Xwhs179* (3.3 cM)	RFLP	–	'Chul' (*Pm3b*)	Hartl *et al*. (1993)
Pm3b	1AS	*Xbcd1434* (1.3 cM)	RFLP	–	'Chul'/8* 'Chancellor'	Ma *et al*. (1994)
Pm3g (Mlar)	1AS	*XGli–A5* (5.2 cM)	Gliadin	–	'Courtot'	Sourdille *et al*. (1999)
Pm3	1AS	*PSP2999* (co–segregating)	SSR	–	'Courtot' (*Pm3g*)	Bougot *et al*. (2002)

Table 3. Continued

Pm4a	2AL	*Xbcd1231, Xcdo678* (both co–segregating)	RFLP	–	'Khapli'/8* 'Chancellor'	Ma *et al*. (1994)
	2AL	*4aM1, 4aM2* (3.5 cM)	AFLP	–	'Khapli'/8* 'Chancellor'	Hartl *et al*. (1999)
Pm5e	7BL	*Xgwm1267* (6.6 cM) *Xgwm783* (11.0 cM)	SSR	–	'Fuzhang 30'	Huang *et al*. (2003)
Pm6	2BL	*Xbcd135* (1.6 cM) *Xbcd266* (4.8 cM)	RFLP	–	6 NILs with 'Prins' as recurrent parent and different doners of *Pm6*. Mapped in 'IGV1–463'	Tao *et al*. (2000)
Pm8/Pm17 (allelic)	1BL/1RS 1AL/1RS	*aIAG95* (specific detection of *Pm8* and *Pm17*)	STS	–	Varios wheats carrying the 1BL/1RS (*Pm8*) and 1AL/1RS (*Pm17*) wheat rye translocations originating from 'Petkus' and 'Insave' rye	Mohler *et al*. (2001)
Pm13	3DS	*Xutv14* (0 cM)	STS	–	Several wheat *Ae. longissima* recombinant lines	Cenci *et al*. (1999)
Pm18	7A	*Xwhs178* (4.4 cM)	RFLP	–	'Weihenstephan M1N'	Hartl *et al*. (1995)
Pm21	6AL/6VS	OPH17–1900 (0 cM)	RAPD	–	Translocation line with 6VS chromosome arm introgressed from *Haynaldia villosa*	Qi *et al*. (1996)
	6AL/6VS	$SCAR_{1265}$ $SCAR_{1400}$ (0 cM)	SCAR	–	*H. villosa* and various *H. villosa* translocation and substitution lines	Liu *et al*. (1999)

Table 3. Continued

Pm24	1DS	*Xgwm337* (2.4 cM) *XAAT/CCA–346* (0 cM) *XACA/CTA–407* (0 cM)	SSR AFLP AFLP	–	'Chiyacao'	Huang *et al*. (2000)
Pm25	1A	$OPAG04_{950}$ (12.8 cM) $OPX06_{1050}$ (17.2 cM) $OPAI14_{600}$ (21.6 cM)	RAPD	–	'NC96BGTA5', *Pm25* was introgressed from *T. monococcum*	Shi *et al*. (1998)
Pm26	2BS	*Xwg516* (0 cM)	RFLP	–	Chromosome substitution line 2BS of *T. turgidum* var. *dicoccoides* (TTD140) into 'Bethlehem'	Rong *et al*. (2000)
Pm29	7DL	*Xwg341* (2.2 cM) and several AFLP markers	RFLP	–	'Pova' derived from a 'Poros' (*T. aestivum*)–*Ae. ovata* alien addition line	Zeller *et al*. (2002)
Pm30	5BS	*Xgwm159* (5–6 cM)	SSR	–	'87–1'/'C20'//2–3*'8866', *Pm30* was introgressed from *T. dicoccoides* ('C20') to wheat line '87–1'	Liu *et al*. (002)
MlRE (3 linked loci)	6AL	*XksuD27* (14.3–33.0 cM)	RFLP	–	'RE714', *MlRE* was introgressed from *T. dicoccum*	Chantret *et al*. (2000)
–	5D	*Xcfd26* *XgbxG083c* *Xcfd8B9*	SSR RFLP SSR	28.1–37.9	'RE714'	Chantret *et al*. (2001)
Qpm.vt–1B	1B	*Xgwm259* – *Xwg241* (ca. 20 cM)	SSR RFLP	17	'Massey'	Liu *et al*. (2001)

Table 3. Continued

Qpm.vt–2A	2A	*Xgwm304a –Xgwm312* (ca. 27 cM)	SSR	29	'Massey'	Liu *et al*. (2001)
Qpm.vt–2B	2B	*Xwg338– Xgwm526a* (ca. 5cM)	RFLP SSR	11	'Massey'	(Liu *et al*. (2001)
–	5A	*Xpsr644a– Xpsr945a* (10 cM)	RFLP	22.9	'Oberkulmer' (spelt)	Keller *et al*. (1999)
–	7B	*Xglk750– Xmwg710a* (13 cM)	RFLP	31.8	'Forno'	Keller *et al*. (1999)

[a] – no name provided for gene/QTL in reference
[b] genetic distance marker to gene or distance between markers in marker intervals is shown in parentheses
[c] AFLP = Amplified Fragment Length Polymorphism, RAPD = Random Amplified Polymorphic DNA, RFLP = Restriction Fragment Length Polymorphism, SCAR = Sequence Characterised Amplified Region, STS = Sequence Tagged Site, SSR = Simple Sequence Repeat
[d] proportion of phenotypic variance explained by the investigated QTL locus

3.4. Stripe Rust

Stripe rust (yellow rust) of wheat is caused by the fungus *Puccinia striiformis* f.sp. *tritici*. In total, 32 *Yr* loci for resistance to stripe rust have been designated (McIntosh *et al*., 1998). Markers have been developed for at least 12 of these loci and for a number of QTLs and temporarily designated genes (Table 4). The resistance provided by these genes is often isolate specific and can be short lived. Several examples of resistances becoming ineffective are known (*e.g.* Hovmøller, 2001; Chen *et al*., 2002). One strategy to prolong the life span of these resistance genes is the pyramiding of several genes into one cultivar. As *P. striiformis* has no known sexual stage this could be an effective strategy, as virulence genes are not readily recombined in the fungus. However, in practice mutations to virulence seem to occur frequently enough so that even gene pyramids are eventually overcome by the pathogen (Wellings and McIntosh, 1990). The accumulation of several resistance genes in one wheat cultivar is considerably facilitated by the availability of molecular markers for such genes.

Laboratory protocols for the MAS of genes *Yr5*, *Yr15* and *Yr17* can be found at http://maswheat.ucdavis.edu. For *Yr5* STS markers developed from the RGAP markers (Table 4) of Yan *et al*. (2003) are described.

Table 4. Genes and QTLs for resistance to stripe rust mapped using molecular markers

Gene/ QTL[a]	Chromo-some arm	Nearest marker(s)/ marker interval[b]	Marker type[c]	R^2 (%)[d]	Material used for mapping/ marker development	Reference
Yr5	2BL	*Xwgp17* (0 cM) *Xwgp18* (0 cM) *Xwgp19* (0 cM) *Xwgp20* (0 cM)	RGAP	–	NIL from backcrossing *T. spelta album* (*Yr5*) to 'Avocet S'	Yan *et al*. (2003)
Yr7	2BL	*P36/M49–4* *P31/M50–1* *P36/M43–5* *P37/M53–1*	AFLP	–	'Cranbrook' and 'CD87'	Bariana *et al*. (2001)
Yr9	1BL/ 1RS	*Xgwp4* (0 cM) *Xgwp7* (0 cM) *Xgwp8* (0 cM) *Xgwp 9* (0 cM)	RGAP	–	*Yr9* NIL in 'Avocet S' background.	(Shi *et al*. (2001)
	1BL/ 1RS	*iag95* (0 cM)	SCAR	–	'Petkus' 1RS chromosome in wheat cultivar 'Pavon' and wheat – rye 1RS – 1BS recombinant lines	Mago *et al*. (2002)
Yr10/ Yr10vav (allelic)	1BS	*Xpsp3000* (1.2 / 2.7 cM)	SSR	–	Wheat lines 'P.I.178383' and 'QLD709'. *Yr10vav* originates from *T. vavilovii*	Wang *et al*. (2002) Bariana *et al*. (2002) Bariana *et al*. (2002)
YrMoro (probably identical to *Yr10*)	1BS	*S26M47*	STS	–	'Moro'	Smith *et al*. (2002)

Table 4. Continued

Yr15	1BS	*Nor1* (11.0 cM)	RFLP	–	NIL from backcrossing *T. dicoccoides* (*Yr15*) to *T. durum* line 'D447'	Sun *et al*. (1997)
		Xgwm33 (5.0 cM)	SSR			Chague *et al*. (1999)
		Xgwm413 (5.5 cM)	SSR			Peng *et al*. (2000)
Yr17	2AS	*SC–Y15* (0.8 cM)	SCAR	–	'VPM1', *Yr17* was introgressed from *Ae. ventricosa*	Robert *et al*. (1999)
	2AS	*csVrga13* (diagnostic for Yr17)	STS	–	NILs developed from 'VPM1'	Seah *et al*. (2001)
Yr18	7DS	*Xbcd1438* – *Xwg834* (ca.3 cM)	RFLP	12–36	'Opata 85' the gene is believed to be *Yr18*	Singh *et al*. (2000)
	7DS	*Xwmc405b* – *P36/M41–2* (ca. 10 cM)	SSR AFLP		'CD87'	Bariana *et al*. (2001)
	7DS	*Xgwm295.1*	SSR	11–24	'Fukuho–komugi'	Suenaga *et al*. (2003)
Yr26	1BS	*Xgwm11* (1.9 cM)	SSR	–	Wheat line 'R55', *Yr26* originates from *T. turgidum*	Ma *et al*. (2001)
		Xgwm18 (1.9 cM)				
		Xgwm413 (5.1 cM)				
Yr28	4DS	*Xbcd265* – *Xmwg634* (ca. 13 cM)	RFLP	9–31	*Ae. tauschii* via synthetic hexaploid	Singh *et al*. (2000)
Yr29	1BL	*Xpsr305* – *P39/M38–2* (ca. 5 cM)	RFLP AFLP	–	'CD87'	Bariana *et al*. (2001)
	1BL	*PstAAG MseCTA–1*	AFLP	31.1	'Pavon 76'	William *et al*. (2003)
		PstAAG MseCGA–1		24.6		

Table 4. Continued

Yr30	3BS	*Xgwm389*	SSR	0.2–24.5	'Oligoculm'	Suenaga *et al*. (2003)
Yr32	2AL	*M62/P19–156* *Xwmc198*	AFLP SSR	–	'Carstens V' and 'Senat', present in many European cultivars	Eriksen *et al*. (2003b)
YrKat	2DS	*Xwmc111–Xwmc025a* (ca. 10 cM)	SSR	–	'Katepwa'	Bariana *et al*. (2001)
Yrns–B1	3BS	*Xgwm493* (20.5–21.7 cM)	SSR	–	'Lgst.79–74'	Börner *et al*. (2000)
YrH52	1BS	*Xgwm413* (1.3 cM) *Xgwm934* (1.4 cM) *Xgwm903* (1.4 cM)	SSR	–	*T. dicoccoides* accession 'Hermon H52'	Peng *et al*. (2000)
–	3BS	*Xfba190 – XksuG53* (ca. 2 cM)	RFLP	16–28	'Opata 85'	Singh *et al*. (2000)
–	3DS	*Xfba241 – Xfba91* (ca. 6 cM)	RFLP	14	'Opata 85'	Singh *et al*. (2000)
–	5DS	*Xfbb238 – Xfba114*	RFLP	8	'Opata 85'	Singh *et al*. (2000)
QYR1	2BL	*Xgwm47 – Xgwm501* (3.4 cM)	SSR	46	'Camp Remy'	Boukhatem *et al*. (2002)
QYR2	2AL	*Xgwm356 – Xgwm382* (20.0 cM)	SSR	10.7–15.4	'Camp Remy'	Boukhatem *et al*. (2002)
QYR3	2BS	*Xcdo405 – Xbcd152* (6.7 cM)	RFLP	30.7	'Opata 85'	Boukhatem *et al*. (2002)
QYR4	7DS	*Xwg834 – Xbcd1438* (6.4 cM)	RFLP	13.9	'Opata 85'	Boukhatem *et al*. (2002)

[a] – no name provided for gene/QTL in reference
[b] genetic distance marker to gene or distance between markers in marker intervals is shown in parentheses
[c] AFLP = Amplified Fragment Length Polymorphism, RFLP = Restriction Fragment Length Polymorphism, RGAP = Resistance Gene Analog Polymorphism, SCAR = Sequence Characterised Site, STS = Sequence Tagged Site, SSR = Simple Sequence Repeat
[b] proportion of phenotypic variance explained by the investigated QTL locus

3.5. Leaf Rust

Leaf rust of wheat is caused by the biotrophic fungus *Puccinia triticina* (synonym *Puccinia recondita* f.sp. *tritici*). The total number of designated loci is 51 (McIntosh *et al*., 1998). Like stripe rust, leaf rust resistance genes are typically isolate specific and usually a new gene only provides protection for a limited period before the virulence spectrum of the pathogen population changes and the resistance loses its effectiveness (Kolmer, 1996).

A number of stripe– and leaf rust resistance genes have been found to be closely linked *e.g. Lr34/Yr18*, *Lr37/Yr17*, *Lr26/Yr9*, *Lr46/Yr29* (Singh *et al*., 1990; Bariana and McIntosh, 1993; Suenaga *et al*., 2003; William *et al*., 2003). This can be due to the presence of these genes on an alien segment, as it is the case for *Lr37/Yr17* in the *Ae. ventricosa* segment on chromosome 2AS (Bariana and McIntosh, 1993). The genes *Lr26/Yr9* in the rye segment of the 1BL/1RS wheat – rye translocation have been shown to be closely linked within the rye segment (Singh *et al*., 1990). This has the advantage that selection for both leaf and stripe rust resistance genes is possible while performing one marker test.

Laboratory protocols for the MAS of leaf rust resistance genes *Lr21*, *Lr25*, *Lr29*, *Lr35*, *Lr37*, *Lr39*, *Lr47*, and *Lr50* with some of the markers listed in Table 5 can be found at http://maswheat.ucdavis.edu.

4. QUALITATIVE VS. QUANTITATIVE NATURE OF DISEASE RESISTANCE

QTLs for disease resistance have often been detected in the chromosomal regions in which previously major race–specific resistance genes were located. The RFLP markers MWG851a and MWG855a, which detected QTLs for mildew resistance in barley on 7H (Backes *et al.*, 1996), are involved in the same chromosomal fragment that contains the *mlt*, a race specific mildew resistance gene (Schönfeld *et al.*, 1996). For powdery mildew resistance, again a QTL locus for mildew resistance was detected on the chromosome 1H in a RI population originating from a cross between 'Vada' and a wild barley line '1B–87' from Israel (Backes *et al.*, 2003). The highly polymorphic *Mla* locus for race–specific resistance gene has also been found to be located on the same chromosomal region of 1H (Schüller *et al.*, 1992). As for leaf rust resistance, the *Rph16*, major race–specific resistance gene and the *LR2* locus, a quantitative trait locus (QTL), have been found in

Table 5. Genes and QTLs for resistance to leaf rust mapped using molecular markers

Gene/ QTL[a]	Chromo-some	Nearest marker(s)/ marker interval[b]	Marker type[c]	R^2 (%)[d]	Material used for mapping/ marker development	Reference
Lr1	5DL	*Xpsr567* (0.04 cM) *Xabc718* (0.12 cM) *Xgwm272* (0.12 cM)	RFLP RFLP SSR	–	NIL 'Thatcher*Lr1*'	Ling *et al*. (2003)
Lr9	6B	*XksuD27* (0 cM)	RFLP	–	'RL6010', *Lr9* originates from *Ae. umbellulata*)	Autrique *et al*. (1995)
Lr10	1AS	*Xcdo426*	RFLP	–	'Opata'	Nelson et al. (1997)
	1AS	*STSLrk10–6* (0 cM)	STS	–	NIL 'Thatcher*Lr10*'	Schacher–mayr *et al*. (1997)
Lr19	7DL	*XksuG39* (0 cM) and 7 more	RFLP	–	'Agatha', *Lr19* originates from *Agropyron elongatum*	Autrique *et al*. (1995)
	7DL	$STSLr19_{130}$ (0 cM)	STS	–	'Indis *Lr19*'	Prins *et al*. (2001)
	7DL	$SCS73_{719}$ (6.4 cM)	SCAR	–	Wheat line 'HW2046'	Cherukuri *et al*. (2003)
Lr21/ Lr40	1DS	*KSUD14–STS* (0–0.2 cM)	STS	–	Wheat lines 'KS86WGRC2' and 'KS89WGRC7', *Lr21* originates from *Ae. tauschii*	Huang and Gill (2001)
Lr23	2BS	*Xtam72 – Xrz69* (ca. 9 cM)	RFLP	12	Synthetic wheat ('Altar 84'/*Ae. tauschii* 'W–219')	Nelson *et al.* (1997)
Lr24	3DL	*Xcdo482* (0 cM) and 7 more	RFLP	–	'Agent', *Lr24* originates from *A. elongatum*	Autrique *et al*. (1995)

Table 5. Continued

Lr25	4A/ 2R	*Lr25F20/ Lr25R20*	SCAR	–	NIL 'RL6084' in 'Thatcher' background, *Lr25* is on 2R of the 4A/2R wheat – rye translocation	Procunier *et al*. (1995)
Lr26	1BL/ 1RS	*iag95 (0cM)*	SCAR	–	'Petkus' 1RS chromosome in wheat cultivar 'Pavon' and wheat – rye 1RS – 1BS recombinant lines	Mago *et al*. (2002)
Lr27	3BS	*XksuG53* – *Xbcd907* (ca. 10 cM)	RFLP	11–17	'Opata'	Nelson et al. (1997)
Lr28	4AL	*OPJ–01*$_{378}$	STS	–	Seven NILs with *Lr28* introgressed from *Ae. speltoides*	Naik *et al*. (1998)
Lr29	7DS	*Lr29F18/Lr29R18* (0 cM) *Lr29F24/Lr29R24* (0 cM)	SCAR	–	NIL 'RL6080' in 'Thatcher' background, *Lr29* originates from *A. elongatum*	Procunier *et al*. (1995)
Lr31	4BL	*Xcdo20* – *Xbcd1265* (ca. 12 cM)	RFLP	–	'Opata'	Nelson *et al.* (1997)
Lr32	3DS	*Xbcd1278* (3.6 cM) *Xcdo395* (6.9 cM)	RFLP	–	Wheat line 'RL5713', *Lr*32 originates from *Ae. tauschii*	Autrique *et al*. (1995)
Lr34	7DS	*Xwg834* – *Xbcd1872* (ca. 12cM)	RFLP	16–45	'Opata'	Nelson et al. (1997)
	7DS	*Xgwm295.1*	SSR	36.4–45.2	'Fukuhu–komugi'	Suenaga *et al*. (2003)

Table 5. Continued

Lr35	2B	*BCD260F1/35R2* (0 cM)	STS	–	NIL Thatcher*Lr35*', *Lr35* originates from *T. speltoides*	Seyfarth *et al*. (1999)
	2B	*Sr39F/Sr39R* (0 cM)	SCAR	–		Gold *et al*. (1999)
Lr37	2AS	*SC–Y15* (0.8 cM)	SCAR	–	'VPM1', *Lr37* introgressed from *Ae. ventricosa*	Robert *et al*. (1999)
	2AS	*csVrga13* (0 cM)	STS	–	NILs developed from 'VPM1'	Seah *et al*. (2001)
Lr39	2DS	*Xgwm210* (10.7 cM)	SSR	–	Wheat line 'TA4186', i.e. *Lr39* in the background of 'Wichita', *Lr39* originates from *Ae. tauschii*	Raupp *et al*. (2001)
Lr46	1BL	*Xwmc44*	SSR	12.9–17.4	'Oligoculm'	Suenaga *et al*. (2003)
	1BL	*PstAAGMseCTA–1*	AFLP	49	'Pavon 76'	William *et al*. (2003)
		PstAAGMseCGA–1		45.7		
Lr47	7AS	*PS10R/PS10L* (0 cM)	STS	–	Wheat line 'T7AS–7S#1S–7AS.7AL' carrying *Lr47* on a *T. speltoides* translocation	Helguera *et al*. (2000)
		PS10R/PS10L2 (0 cM)	CAPS			
Lr50	2BL	*Xgwm382* (6.7 cM) *Xgdm87* (9.4 cM)	SSR	–	Various wheat lines with *Lr50* introgressed from *T. timopheevii* ssp. *armeniacum*	Brown–Guedira *et al*. (2003)
–	7BL	*Xrz508 – Xfbb189 (ca. 5cM)*	RAPD	20–24	'Opata'	(Nelson et al., 1997)

[a] – no name provided for gene/QTL in reference
[b] genetic distance marker to gene or distance between markers in marker intervals is shown in parentheses
[c] AFLP = Amplified Fragment Length Polymorphism, CAPS = Cleaved Amplified Polymorphic Site, RFLP = Restriction Fragment Length Polymorphism, RGAP = Resistance Gene Analog Polymorphism, SCAR = Sequence Characterised Site, SSR = Simple Sequence Repeat, STS = Sequence Tagged Site
[d] proportion of phenotypic variance explained by the investigated QTL locus

the same confidence interval on chromosome 2H in the cross between a land race, 'Hor 1063' from Turkey and cultivar 'Krona' from Germany (Kicherer *et al.*, 2000). For scald resistance, major race–specific resistance genes and QTLs have been detected particularly on chromosomes 2H, 4H and 6H in a barley DH population developed from the cross between "Alexis" and "Regatta" (Jensen *et al.*, 2002). Such observations are in agreement with other host–pathogen interactions. In barley, QTLs for resistance against net blotch disease (Richter *et al.*, 1998), yellow rust and also for leaf rust (Thomas *et al.*, 1995) mapped approximately on the same position as earlier localised qualitative resistance genes. Keller *et al.* (1999) localized a QTL for resistance to wheat powdery mildew at or near the position of the qualitative *Pm5* locus. In addition, for example Leonard–Schipper *et al.* (1994) localized a QTL for resistance to late blight in potato to chromosomal segments to which race –specific alleles of the *R1* locus for resistance to the same pathogen had been localized. This tight relation between "qualitative" and "quantitative" resistance is further supported by the fact that the effect of resistance–QTLs can differ significantly using different defined isolates of a pathogen (Danesh and Young, 1994; Leonards–Schippers *et al.*, 1994; Backes *et al.*, 1996; Caranta *et al.,* 1997; Li *et al.,* 1999; Qi *et al.,* 1999) and that often resistance is determined genetically by one or two major QTL/genes together with several minor QTLs/genes (Pecchioni *et al.,* 1996; van der Voort *et al.,* 1998; Kicherer *et al.,* 2000).

There are several arguments for co–localization of qualitative and quantitative resistance:

(i) A major race specific gene at a given QTL results in complete susceptibility against the isolates with virulence, and complete resistance against the avirulent isolates, however the inoculum is a mixture of both isolates and therefore, only a part of the pathogen population is affected. (Keller *et al.,* 1999).

(ii) Combination of qualitative resistance genes can result in higher resistance levels than the individual genes, as demonstrated for leaf rust in wheat by Kloppers and Pretorius (1997). This could produce results like the ones for the leaf rust QTLs, where the effect of the combination of resistant alleles is lower than expected by the assumption of an additive model (Backes *et al.*, 2003; Kicherer *et al.*, 2000).

(iii) A "defeated" qualitative resistance allele at a locus acts as a quantitative gene. An example of this kind of behavior has been found in the investigation of Li *et al.* (1999), where one locus was identified as qualitatively acting after

infection with one isolate and as a QTL after infection with two other isolates.

(iv) A qualitative resistance locus represents an "extreme" allele of a quantitative locus.

(v) Frequently, resistance genes seem to have a clear tendency to cluster in multigene families (Jahoor *et al.*, 1993; Yu *et al.,* 1996; Wei *et al.,* 1999). Therefore, some QTLs could be "mild" members of these families.

(vi) Since regulatory elements and genes, which are necessary as complementary factors in resistance reaction, can be located in close linkage to the qualitative resistance genes (Kalavacharla *et al.,* 2000), these genes can be localized as QTLs near the respective qualitative genes.

Looking at all these arguments, it might be concluded, that the terms "qualitative resistance" and "quantitative resistance" need not be used to define two different kinds of resistance as originally intended by van der Plank (1968). Basically, the term "quantitative" and "qualitative" describe only the way one is treating the data: if we cannot see clear resistance and susceptible groups, we have to use quantitative methods. The additional benefit of these quantitative methods is that we can find additional loci, interactions between loci and interactions between loci and environmental factors as shown in wheat for the powdery mildew resistance gene *MlRE* (Chantret *et al.,* 2000) and in barley for powdery mildew and leaf rust genes (Backes *et al.,* 2003).

5. CO–LOCALIZATION OF RESISTANCE GENES AND *R*ESISTANCE *G*ENE *A*NALOGUES (RGA)

The plant disease resistance genes cloned to date can be divided into a few major groups based on the characteristics of the proteins they encode (Hammond–Kosack and Jones, 1997; Dangl and Jones, 2001; Fluhr, 2001). The majority of these genes code for proteins with a nucleotide binding site (NBS) motif followed by long C–terminal regions containing more or less well–organised Leucine Rich Repeats (LRRs). These NBS–LRR disease resistance genes belong to a large and diverse superfamily of genes, which can be subdivided based on characteristic N–terminal features of their products. One class of NBS-LRR resistance proteins contains an N–terminal TIR (Toll–Interleukin–Receptor homology) domain, another contains a predicted N–terminal coiled coil, while other NBS–LRR resistance proteins

contain neither of these N–terminal motifs (Hammond–Kosack and Jones, 1997; Dangl and Jones, 2001; Fluhr, 2001).

Five disease resistance genes cloned from monocots belong to the NBS–LRR class: the gene *Xa1* for bacterial blight resistance in rice, *Pib* and *Pi–ta* for rice blast resistance, *Rp1–D* for rust resistance in maize, and *Mla* for powdery mildew resistance in barley (Yoshimura *et al.,* 1998; Collins *et al.,* 1999; Wang *et al.,* 1999; Wei *et al.*, 1999; Bryan *et al.,* 2000; Halterman *et al.,* 2001; Zhou *et al.,* 2001). Some of the NBS–LRR genes to which no function has yet been assigned probably represent disease resistance genes, some may contribute quantitatively to disease resistance, while others may represent defeated disease resistance genes or genes that perform functions unrelated to disease resistance (Dangl and Jones, 2001; Backes *et al.*, 2003; Wang *et al.,* 1999; Yan *et al.,* 2003).

The discovery of common motifs in cloned resistance genes, such as leucin–rich–repeats (LRR), nucleotide–binding–sites (NBS), serine/threonine kinase domains, offer opportunity for isolation of similar sequences in other plant species. Specific genomic DNA amplified by degenerated primers based on these motifs is known as resistance gene analogue (RGA). The RGA approach has been used to map disease resistance genes and to isolates resistance genes (Leister *et al.*, 1998; Collins *et al.*, 1998, 2001). In barley, RGAs have been mapped to all seven barley chromosomes (Backes et al, 2003; Madsen *et al.*, 2003). On chromosomes 2H and 6H, close linkage between RGAs and quantitative powdery mildew as well as leaf rust resistance has been detected, whereas on chromosome 3H and 7H seedling resistance to powdery mildew and RGAs are linked (Backes *et al.*, 2003). In wheat, RGAs have been found that are linked with the stripe resistance gene *Yr5* and *Yr9* (Shi *et al.*, 2001; Yan *et al.*, 2003). Therefore, RGAs offer an ideal marker type to localize and probably isolate disease resistance genes. This marker type can also be used in marker–assisted breeding for resistance, since this marker type is based on PCR.

6. MARKER ASSISTED RESISTANCE BREEDING

Frequently, resistance genes incorporated in the commercial cultivars are overcome by the new pathotypes that possess corresponding virulence genes. Therefore, it is important to look for new sources of resistance to specific pathogens. New sources of resistance are generally found in exotic material or wild relatives of cultivated crop plants. With the increased number of resistance genes, a large number of isolates of a pathogen has to be employed to identify resistance genes. Therefore, DNA based markers can be used to

identify different resistance genes and/or different alleles of a resistant locus (Jahoor and Fischbeck, 1993; Hartl *et al.*, 1993, 1995; Mohler and Jahoor 1995). A large number of markers linked with a variety of resistance genes have been identified in cereals. These markers offer an opportunity not only to identify resistance genes but also to transfer resistance genes successfully into the cultivars via marker–assisted selection (MAS). With the application of co–dominant markers, it is possible to identify genotypes that are heterozygous and homozygous for resistance genes. This will allow selecting homozygous resistant individuals in early generations that will not segregate for resistance. Further, with the combined application of DH–techniques and MAS techniques, the time for developing new resistant cultivars can be drastically reduced. This will help breeders to develop resistant cultivars faster. Pyramiding of resistance genes will certainly prolong the effective lifetime of resistance genes, since the pathogen has to develop more than one virulence gene through mutation and/or recombination to over come the resistance genes in the host cultivar. However, newly identified resistance genes are often highly effective against all existing pathotypes of a given pathogen (Jahoor and Fischbeck, 1987b). They can only be identified in separate lines. If two highly effective resistance genes are to be combined in a line, only linked molecular markers can be used to identify such combinations. Combinations of different highly effective resistance genes in a cultivar will certainly provide a more durable resistance. By the application of MAS, different resistance genes for different diseases can be selected. Such genotypes possessing a combination of resistance genes not only for one disease but also for different diseases will certainly reduce the application of fungicide drastically.

To identify quantitative resistance, field experiments with replications under different environments are required. Therefore, successful identification of quantitative resistance takes several years. Hence, it is very hard or impossible to transfer QTLs for resistance from exotic material to adapted material via conventional breeding. Therefore, breeders are often hesitating to use quantitative resistance in their breeding programs. However, quantitative resistance often provides more durable resistance than qualitative resistance genes. Further, combination of qualitative and quantitative inherited resistance genes provide the ideal way to control diseases. Such combination can only be identified by the application of DNA markers. Before DNA marker for quantitative resistance can be employed in MAS, they have to be identified and validated in different genetic backgrounds. For example, *Fusarium* resistance in wheat is mostly controlled by QTLs and has been identified particularly in “Sumai 3”. This source of resistance has been mapped with DNA markers, which have been validated in different genetic

backgrounds. Therefore, DNA markers can be employed to transfer 'Sumai 3' resistance into adapted material.

Many genes for resistance to different diseases have been identified in wild relatives of cultivated cereals. Several backcrosses between a donor and recurrent parents are necessary to transfer resistance genes from wild relatives. Therefore, it takes several years to select adapted material from such backcrosses. Tanksley and Nelson (1996) proposed a method for simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines by the application of DNA markers. By selecting genotypes for backcrossing with the help of DNA markers, the extent of genome of the donor/unadapted germplasm can be reduced in a few backcrossing steps. Since backcrossing with a single recurrent parent will lead to the development of more or less near–isogenic lines for disease resistance. This procedure required further crosses to develop varieties. Therefore, it is important to use different varieties in crosses when transferring resistance from unadapted material. From such types of crosses, several varieties have been released in Germany and in Denmark, in which powdery mildew resistance genes have been transferred from wild barley (Backes and Jahoor, 2001).

7. SUMMARY AND OUTLOOK

Yield losses caused by disease can be overcome by the introduction of resistance cultivars. However, the rapid development of pathogens can overcome the newly introduced resistance gene in a cultivar. Therefore, it is important to incorporate new resistance genes or combination of resistance genes. However newly identified resistance genes are highly effective and cannot be identified with the help of host pathogen interaction. The combination of several highly effective resistance genes is impossible to identify with the help of gene-for gene model. Therefore, DNA markers linked with resistance genes provide a tool to identify and select highly effective as well combination of resistance genes in a cultivar via marker assisted selection. In this chapter, a comprehensive literature survey concerning disease resistance genes mapped or cloned so far have been summarized. It appears that disease resistance genes against most important diseases have been mapped with the molecular markers in wheat and barley. Therefore, DNA markers can be selected to conduct MAS for disease resistance. However, only few markers have been or are being used in practical resistance breeding. To employ DNA markers in plant breeding, they have to be breeder friendly and they should be linked very closely. Simultaneous application of several DNA markers linked not only to different

resistance genes for a disease but also for different diseases, breeding lines can be selected that will posses durable resistance. With the fast development of DNA chips technology, DNA chips will be available that would contain specific primers for different resistance genes. These DNA chips will allow identification of not only different resistance loci but also different alleles of a resistance locus. RGA have been found to be linked with the qualitative as well as quantitative disease resistance genes. A large number of ESTs is available for wheat and barley. The RGA can be searched in the EST data base and with the help of these data base perhaps DNAchips can be developed to conduct MAS. Further, the linked DNA markers have also been employed to isolate disease resistance genes.

8. REFERENCES

Abbott D.C., Lagudah E.S., Brown A.H. D. Identification of RFLPs flanking a scald resistance gene on barley chromosome 6. Jour Heredity 1995; 86: 152–154

Adhikari T.B., Anderson J.M., Goodwin S.B. Identification and molecular mapping of a gene in wheat conferring resistance to *Mycosphaerella graminicola*. Phytopathology 2003; 93: 1158–1164

Anderson J.A., Stack R.W., Liu S., Waldron B.L., Fjeld A.D., Coyne C., Moreno–Sevilla B., Fetch J.M., Song Q. J., Cregan P.B., *et al.* DNA markers for fusarium head blight resistance QTLs in two wheat populations. Theor Appl Genet 2001; 102: 1164–1168

Arraiano L.S., Worland A.J., Ellerbrook C., Brown J.K.M. Chromosomal location of a gene for resistance to septoria tritici blotch (*Mycosphaerella graminicola*) in the hexaploid wheat 'Synthetic 6*x*'. Theor Appl Genet 2001; 103: 758–764

Arru L., Niks R.E., Lindhout P., Vale G., Francia E., Pecchioni N. Genomic regions determining resistance to leaf stripe (*Pyrenophora graminea*) in barley. Genome 2002; 45: 460–466.

Arru L., Francia E., Pecchioni N. Isolate specific QTLs of resistance to leaf stripe (*Pyrenophora graminea*) in the 'Steptoe' x 'Morex' spring barley cross. Theor Appl Genet 2003; 106:668–675

Autrique E., Singh R.P., Tanksley S.D., Sorrells M.E. Molecular markers for four leaf rust resistance genes introgressed into wheat from wild relatives. Genome 1995; 38: 75–83

Backes G., Graner A., Foroughi–Wehr B., Fischbeck G., Wenzel G., Jahoor A. Localization of quantitative trait loci (QTL) for agronomic important characters by the use of a RFLP map in barley (*Hordeum vulgare* L.). Theor Appl Genet 1995; 90: 294–302

Backes G., Schwarz G., Wenzel G., Jahoor A. Comparison between QTL analysis of powdery mildew resistance in barley based on detached primary leaves and on field data. Plant Breed 1996; 115: 419–421

Backes G., Jahoor A. Wege zur Nutzung der genetischen Potentials der Wild gerste – mit molecularen Markern zur neuen Resistenzgenen. Bericht über die 52. Tagung 2001 der Vereinigung der Pflanzenzüchter und Saatgutkaufleute Östereichs BAL Gumpenstein, 2001; 81–85

Backes G., Madsen L.H., Jaiser H., Stougaard J., Herz M., Mohler V., Jahoor A. Localisation of genes for resistance against *Blumeria graminis* f.sp *hordei* and *Puccinia graminis* in a cross between a barley cultivar and, a wild barley (*Hordeum vulgare* ssp. *spontaneum*) line. Theor Appl Genet 2003; 106: 353–362

Bai G. and Shaner G. Scab of wheat: Prospects for control. Plant Disease 1994; 78: 760–766

Baker R. J., Larter R. N. The inheritance of scald resistance in barley. Can Jour Genet Cytol 1963; 5: 445–449

Bariana H.S., McIntosh R.A. Cytogenetic studies in wheat. XV. Location of rust resistance genes in VPM1 and their genetic linkage with other disease resistance genes in chromosome 2A. Genome 1993; 36: 476–482

Bariana H.S., Brown G.N., Ahmed N.U., Khatkar S., Conner R.L., Wellings C.R., Haley S., Sharp P.J., Laroche A. Characterisation of *Triticum vavilovii*–derived stripe rust resistance using, cytogenetic and molecular analyses and its marker–assisted selection. Theor Appl Genet 2002; 104: 315–320

Bariana H.S., Hayden M.J., Ahmed N.U., Bell J.A., Sharp P.J., McIntosh R.A. Mapping of durable adult plant and seedling resistances to stripe rust and stem rust diseases in wheat. Aust Jour Agric Res 2001; 52: 1247–1255

Barua U. M., Chalmers K. J., Thomas W. T. B., Hackett C. A. Molecular mapping of genes determining height, time to heading, and groth habit in barley (*Hordeum vulgare*). Genome 1993; 36: 1080–1087

Becker J., Vos P., Kuiper M., Salamini F., Heun M. Combined mapping of AFLP and RFLP markers in barley. Mol Gen Genet 1995; 249: 65–73

Bjørnstad.Å, Patil V., Tekauz A., Marøy A.G., Skinnes H., Jensen A., MacKey J. Resistance to scald (*Rhynchosporium secalis*) in barley (*Hordeum vulgare*) studied by near–isogenic lines: I. Markers and differential isolates. Phytopathology 2002; 92 :710–720

Börner A., Röder M.S., Unger O., Meinel A. The detection and molecular mapping of a major gene for non–specific adult–plant resistance against stripe rust (*Puccinia striiformis*) in wheat. Theor Appl Genet 2000; 100: 1095–1099

Borovkova I.G., Steffenson B.J., Jin Y., Rasmussen J.B., Kilian A., Kleinhofs A., Rossnagel B. G., Kao K.–N. Identification of molecular markers linked to the stem rust resistance gene *rpg4* in barley. Phytopathology 1995; 85: 181–185

Borovkova I.G., Jin Y., Steffenson B., Kilian A., Blake T. K., Kleinhofs A. Identification and mapping of a leaf rust resistance gene in barley line Q21861. Genome 1997; 40: 236–241

Borovkova I.G., Jin Y., Steffenson B.J. Chromosomal location and genetic relationship of leaf rust resistance genes *Rph9* and *Rph12* in barley. Phytopathology 1998; 88: 76–80

Bougot Y., Lemoine J., Pavoine M.T., Barloy D., Doussinault G. Identification of a microsatellite marker associated with *Pm3* resistance alleles to powdery mildew in wheat. Plant Breed 2002; 121, 325–329

Boukhatem N., Baret P.V., Mingeot D., Jacquemin J.M. Quantitative trait loci for resistance against yellow rust in two wheat–derived recombinant inbred populations. Theor Appl Genet 2002; 104: 111–118

Bourdoncle W., Ohm H.W. Quantitative trait loci for resistance to fusarium head blight in recombinant inbred lines from the cross Huapai 57–2 / Patterson. Euphytica 2003; 131: 131–136

Brading P.A., Verstappen E.C.P., Kema G.H.J., Brown J.K.M. A gene–for–gene relationship between wheat and *Mycosphaerella graminicola,* the septoria tritici blotch pathogen. Phytopathology 2002; 92: 439–445

Briggs F.N., Stanford E. H. Linkage of factors for resistance to mildew in barley. Jour Genet 1938; 37: 107–117

Briggs F. N., Stanford E. H. Linkage relations to the Goldfoil factor for resistance to mildew in barley. J. Agric. Res. 1943; 66: 1–5

Brown J.K.M., Kema G.H.J., Forrer H.–R., Verstappen E.C.P., Arraiano L.S., Brading P.A., Foster E. M., Fried P.M., Jenny E. Resistance of wheat cultivars and breeding lines to septoria tritici blotch caused by isolates of *Mycosphaerella graminicola* in field trials. Plant Pathol 2001; 50: 325–338

Brown–Guedira G.L., Singh S., Fritz A.K. 2003. Performance and mapping of leaf rust resistance transferred to wheat from *Triticum timopheevii* subsp. *armeniacum*. Phytopathology 2003; 93: 784–789

Brueggeman R., Rostoks N., Kudrna D., Kilian A., Han F., Chen J., Druka A., Steffenson B., Kleinhofs A. The barley stem rust–resistance gene *Rpg1* is a novel disease–resistance gene with homology to receptor kinases. Proc Natl Acad Sci USA 2002; 99: 9328–9333

Brunner S., Keller B., Feuillet C. Molecular mapping of the *Rph7.g* leaf rust resistance gene in barley (*Hordeum vulgare* L.). Theor Appl Genet 2000; 101:783–788

Bryan G.T., Wu K., Farrall L., Jia Y., Hershey H.P., McAdams S.A., Faulk K.N., Donaldson G.K., Tarchini R., Valent B. A single amino acid difference distinguishes resistant and susceptible alleles of the rice blast resistance gene Pi–ta. Plant Cell 2000; 12: 2033–2045

Buerstmayr H., Lemmens M., Hartl L., Doldi L., Steiner B., Stierschneider M., Ruckenbauer P. Molecular mapping of QTLs for fusarium head blight resistance in spring wheat. I. Resistance to fungal spread (type II resistance). Theor Appl Genet 2002; 104: 84–91

Buerstmayr H., Steiner B., Hartl L., Griesser M., Angerer N., Lengauer D., Miedaner T., Schneider B., Lemmens M. Molecular mapping of QTLs for Fusarium head blight resistance in spring wheat. II. Resistance to fungal penetration and spread. Theor Appl Genet 2003; *in press*.

Büschges R., Hollricher K., Panstruga R., Simons G., Wolter M., Frijters A., van Daelen R., van der Lee T., Diergaarde P., Groenendijk J., *et al.* The barley *mlo* gene: A novel control element of plant pathogen resistance. Cell 1997; 88:695–705.

Caranta C., Letesure V., Palloix A. Polygenic resistance of pepper to poly viruses consists of a combination of isolate–specific and broad – spectrum quantitative trait loci. Mol Plant Microbe Interact 1997; 10:372–378

Castro A. J., Chen X., Hayes P. M., Knapp S. J., Line R. F., Toojinda T., Vivar H. Coincident QTL which determine seedling and adult plant resistance to stripe rust in barley. Crop Sci 2002; 42:1701–1708.

Castro A. J., Capettini F., Corey A. E., Filichkina T., Hayes P. M., Kleinhofs A., Kudrna D., Richardson K., Sandoval–Islas S., Rossi C., *et al.* Mapping and pyramiding of qualitative and quantitative resistance to stripe rust in barley. Theor Appl Genet 2003a; 107: 922–930.

Castro A. J., Chen X. M., Hayes P. M., Johnston M. Pyramiding quantitative trait locus (QTL) alleles determining resistance to barley stripe rust: Effects on resistance at the seedling stage. Crop Sci 2003b; 43: 651–659.

Cenci A., D'Ovidio R., Tanzarella O.A., Ceoloni C., Porceddu E. Identification of molecular markers linked to *Pm13*, an *Aegilops longissima* gene conferring resistance to powdery mildew in wheat. Theor Appl Genet 1999; 98: 448–454.

Chagué V., Fahima T., Dahan A., Sun G. L., Korol A.B., Ronin Y.I., Grama A., Röder M.S., Nevo E. 1999. Isolation of microsatellite and RAPD markers flanking the *Yr15* gene of wheat using NILs and bulked segregant analysis. Genome 1999; 42: 1050–1056.

Chantret N., Sourdille P., Röder M., Tavaud M., Bernard M., Doussinault G. Location, mapping of the powdery mildew resistance gene *MIRE* and detection of a resistance QTL by bulked segregant analysis (BSA) with microsatellites in wheat. Theor Appl Genet 2000; 100: 1217–1224.

Chantret N., Mingeot D., Sourdille P., Bernard M., Jacquemin J.M., Doussinault G. A major QTL for powdery mildew resistance is stable over time and at two developmental stages in winter wheat. Theor Appl Genet 2001; 103: 962–971.

Chen F. Q., Prehn D., Hayes P. M., Mulrooney D., Corey A., Vivar H. Mapping genes for resistance to barley stripe rust (*Puccinia striiformis* f. sp. *hordei*). Theor Appl Genet 1994; 88: 215–219.

Chen X.M., Moore M., Milus E.A., Long D.L., Line R.F., Marshall D., Jackson L. Wheat stripe rust epidemics and races of *Puccinia striiformis* f.sp. *tritici* in the United States in 2000. Plant Disease 2002; 86:39–46

Cherukuri D.P, Gupta S.K., Charpe A., Koul S., Prabhu K.V., Singh R.B., Haq Q.M.R., Chauhan S.V.S. Identification of a molecular marker linked to an *Agropyron elongatum*–derived gene *Lr19* for leaf rust resistance in wheat. Plant Breed 2003; 122: 204–208.

Collins N., Drake J., Ayliffe M., Sun Q., Ellis J., Hulber S., Pryor T. Molecular characterization of the maize Rp1–D rust resistance haplotype and its mutants. Plant Cell 1999; 11: 1365–1376

Collins N., Park R., Spielmeyer W., Ellis J., Pryor A. J. Resistance gene analogs in barley and their relationship to rust resistance genes. Genome 2001; 44: 375–381.

Collins N.C., Webb C.A., Seah S., Ellis J.G., Hulbert S.H., Pryor A. The isolation and mapping of disease resistance gene analogs in maize. Mol Plant Microbe Interact 1998; 11: 968–978

Collins N.C., Thordal–Christensen H., Lipka V., Bau S., Kombrink E., Qiu J.L., Huckelhoven R., Stein M., Freiadenhoven A., Somerville S.C., Schulze–Lefert P. SNARE–protein–mediated disease resistance at the plant cell wall. Nature 2003; 425: 973–977.

Conry M.J., Dunne B. Influence of number and timing of fungicide applications on the yield and quality of early and later–sown spring malting barley grown in the south–east of Ireland. Jour Agric Sci 2001; 136: 159–167

Cowger C., Hoffer M.E., Mundt C.C. Specific adaptation by *Mycosphaerella graminicola* to a resistant wheat cultivar. Plant Pathol 2000; 49: 445–451.

Danesh D., Young N.D. Partial resistance loci for tomato bacterial wilt show differential race specificity. Rep Tomato Genet Coop 1994; 44: 12–13

Dangl J.L., Jones J.D.G. Plant pathogens and integrated defence responses to infection. Nature 2001; 411: 826–833

Doll H., Jensen H. P. Localization of powdery mildew resistance gene *Ml–ra* on barley chromosome 5. Hereditas 1986; 105: 61–65.

Dyck P. L., Schaller C. Inheritance of resistance in barley to several physiologic races of the scald fungus. Can Jour Genet Cytol 1961; 3: 153–164.

Eriksen L., Shaw M.W., Østergård H. A model of the effect of pseudothecia on genetic recombination and epidemic development in populations of *Mycosphaerella graminicola*. Phytopathology 2001; 91: 240–248 and Erratum 519.

Eriksen L., Munk L. The occurrence of *Mycosphaerella graminicola* and its anamorph *Septoria tritici* in winter wheat during the growing season. Eur Jour Plant Pathol 2003; 109:253–259.

Eriksen L., Borum F., Jahoor A. 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. Theor Appl Genet 2003a; 107:415–527.

Eriksen L., Afshari F., Christiansen M.J., McIntosh R.A., Jahoor A., Wellings C.R. *Yr32* for resistance to stripe (yellow) rust present in the wheat cultivar Carstens V. Theor Appl Genet 2003b; *in press*.

Falak I., Falk D. E., Tinker N. A., Mather D. E. Resistance to powdery mildew in a doubled haploid barley population and its association with marker loci. Euphytica 1999; 107:185–192.

Feuerstein U., Brown A. H. D., Burdon J. J. Linkage of rust resistance genes from wild barley (*Hordeum spontaneum*) with isozyme markers. Plant Breed 1990; 104: 318–324.

Flor H.H. Host–parasite interaction in flax rust– its genetics and other implication. Phytopathology 1955; 45: 680–685

Fluhr R. Sentinels of disease. Plant resistance genes. Plant Physiol 2001; 127:1367–1374

Franckowiak J. D. Revised linkage maps for morphological markers in barley, *Hordeum vulgare*. Barley Genet Newsl 1997; 26: 9–21

Gale M.D., Miller T.E. “The introduction of alien genetic variation into wheat.” In: F.G.H. Lupton (ed.) *Wheat Breeding, Its Scientific Basis*, Chapman and Hall, London, 1987, pp. 173–210

Garvin D.F., Brown A.H.D., Burdon J.J. Inheritance and chromosome locations of scald–resistance genes derived from Iranian and Turkish wild barleys. Theor Appl Genet 1997; 94:1086–1091

Garvin D.F., Brown A.H.D., Raman H., Read B. J. Genetic mapping of the barley *Rrs14* scald resistance gene with RFLP, isozyme and seed storage protein markers. Plant Breed 2000; 119: 193–196

Gervais L., Dedryver F., Morlais J.–Y., Bodusseau V., Negre S., Bilous M., Groos C., Trottet M. Mapping of quantitative trait loci for field resistance to Fusarium head blight in an European winter wheat. Theor Appl Genet 2003; 106: 961–970

Giese H. Powdery mildew resistance genes in the *Ml–a* and *Ml–k* regions on barley chromosome 5. Hereditas 1981; 95: 51–62

Gold J., Harder D., Townley–Smith F., Aung T., Procunier J. Development of a molecular marker for rust resistance genes *Sr39* and *Lr35* in wheat breeding lines. Electr Jour Biotech 1999; 2: 35–40

Görg R., Hollricher K., Schulze–Lefert P. Functional analysis and RFLP–mediated mapping of the *Mlg* resistance locus in barley. Plant Jour 1993; 3: 857–866

Graner A., Jahoor A., Schondelmaier J., Siedler H., Pillen K., Fischbeck G., Wenzel G., Herrmann R. G. Construction of an RFLP map of barley. Theor Appl Genet 1991; 83: 250–256

Graner A., Foroughi–Wehr B., Tekauz A. RFLP mapping of a gene in barley conferring resistance to net blotch (*Pyrenophora teres*). Euphytica 1996; 91: 229–234

Graner A., Streng S., Drescher A, Jin Y., Borovkova I. G., Steffenson B. Molecular mapping of the leaf rust resistance gene *Rph7* in barley. Plant Breed 2000; 119: 389–392

Graybosch R.A., Peterson C.J., Hansen L.E., Worral D., Shelton D.R., Lukaszewski A.J. Comparative flour quality and protein characteristics of 1BL/1RS and 1AL/1RS wheat–rye translocations. Jour Cereal Sci 1993; 17; 95–106

Griffery C.A., Das M.K., Baldwin R.E., Waldenmeier, C.M. Yield loss in winter barley resulting from a new race of *Puccinia hordei* in North America. Plant Disease 1994; 78: 256–260

Guo P.G., Bai G.H., Shaner G.E. AFLP and STS tagging of a major QTL for fusarium head blight resistance in wheat. Theor Appl Genet 2003; 106: 1011–1017.

Habgood R. M., Hayes J. D. The inheritance of resistance to *Rhynchosporium secalis* in barley. Heredity 1971; 27: 25–37

Halterman D., Zhou F., Wei F., Wise R.P., Schulze–Lefert P. The MLA6 coiled–coil, NBS–LRR protein confers *AvrMla6*–dependent resistance specificity to *Blumeria graminis* f.sp. hordei in barley and wheat. Plant Jour 2001; 25: 335–348

Hammond–Kosack K.E., Jones J.D.G. Plant disease resistance genes. Annu Rev Plant Physiol Plant Mol Biol 1997; 48:575–607

Hardwick N.V., Jones D.R., Slough J.E. Factors affecting diseases of winter wheat in England and Wales, 1989–98. Plant Pathol 2001; 50: 453–462

Hartl L., Mohler V., Zeller F.J., Hsam S.L.K., Schweizer G. Identification of AFLP markers closely linked to the powdery mildew resistance genes *Pm1c* and *Pm4a* in common wheat (*Triticum aestivum* L.). Genome 1999; 42: 322–329

Hartl L., Weiss H., Zeller F.J, Jahoor A. Use of RFLP markers for the identification of alleles of the *Pm3* locus conferring powdery mildew resistance in wheat (*Triticum aestivum* L.). Theor Appl Genet 1993; 86: 959–963

Hartl L., Weiss H., Stephan U., Zeller F.J., Jahoor A. Molecular identification of powdery mildew resistance genes in common wheat (*Triticum aestivum* L.). Theor Appl Genet 1995; 90: 601–606

Hayes P., Prehn D., Vivar H., Blake T., Comeau A., Henry I., Johnston M., Jones B., Steffenson B., *et al.*, Multiple disease resistance loci and their relationship to agronomic and quality loci in a spring barley population, Jour Agri Genom;1996 2 <http://www.cabi–publishing.org/gateways/jag/index.html>

Helguera M., Khan I.A., Dubcovsky J. Development of PCR markers for the wheat leaf rust resistance gene *Lr47*. Theor Appl Genet 2000; 100: 1137–1143 and 101: 625–631

Heun M. Mapping quantitative powdery mildew resistance of barley using a restriction fragment–length polymorphism map. Genome 1992; 35: 1019–1025

Hilbers S., Fischbeck G., Jahoor A. Localization of the *Laevigatum* resistance gene *MlLa* against powdery mildew in the barley genome by the use of RFLP markers. Plant Breed 1992; 109:335–338

Hinze K., Thompson R.D., Ritter E., Salamini F., Schulze–Lefert P. Restriction Fragment Length Polymorphism–mediated targeting of the ml–o resistance locus in barley (*Hordeum vulgare*). Proc Natl Acad Sci USA 1991; 88: 3691–3695

Hiura U. Studies on the disease resistance in barley. IV. Genetics of resistance to powdery mildew. Berichte des Ohara Instituts für Landwirtschaft Okayama 1960; 11: 235–300.

Horvath H., Rostoks N., Brueggeman R., Steffenson B., von Wettstein D., Kleinhofs A. Genetically engineered stem rust resistance in barley using the *Rpg1* gene. Proc Natl Acad Sci USA 2003; 100: 364–369.

Hossain M.A., Sparrow D.H.B. Resistance to powdery mildew (*Erysiphe graminis* f.sp. hordei) in the barley cultivar galleon. I. Relationship with known genes for resistance. Euphytica 1991a; 52: 1–9.

Hossain M.A., Sparrow D.H.B. Resistance to powdery mildew (*Erysiphe graminis* f.sp. *hordei*) in the barley cultivar Galleon. II. Chromosomal location and linkage with hordein genes. Euphytica 1991b; 52: 11–17.

Hovmøller M.S. Disease severity and pathotype dynamics of *Puccinia striiformis* f.sp *tritici* in Denmark. Plant Pathol 2001; 50:181–189.

Hovmøller M.S., Caffier V., Jalli M., Andersen O., Besenhofer G., Czembor J.H., Dreiseitl A., Felsenstein F., Fleck A., Heinrics F., *et al.* The European barley powdery mildew virulence survey and disease nursery 1993–1999. Agronomie 2001; 21:193–194

Hsam S.L.K., Zeller F.J. Breeding for powdery mildew resistance in common wheat (*Triticum aestivum* L.). In: R.R. Bélanger, W.R. Bushnell, A.J. Dik, T.L.W. Carver (eds.) *The Powdery Mildews. A Comprehensive Treatise*, American Phytopathological Society, 2002, pp. 219–238

Hu X.Y., Ohm H.W., Dweikat I. Identification of RAPD markers linked to the gene *Pm1* for resistance to powdery mildew in wheat. Theor Appl Genet 1997; 94: 832–840

Huang L., Gill B.S. An RGA – like marker detects all known *Lr21* leaf rust resistance gene family members in *Aegilops tauschii* and wheat. Theor Appl Genet 2001; 103: 1007–1013

Huang X.Q., Hsam S.L.K., Zeller F.J., Wenzel G., Mohler V. Molecular mapping of the wheat powdery mildew resistance gene *Pm24* and marker validation for molecular breeding. Theor Appl Genet 2000; 101: 407–414

Huang X.Q., Wang L.X., Xu M.X., Röder M.S. Microsatellite mapping of the powdery mildew resistance gene *Pm5e* in common wheat (*Triticum aestivum* L.). Theor Appl Genet 2003; 106: 858–865

Ivandic V., Walther U., Graner A. Molecular mapping of a new gene in wild barley conferring complete resistance to leaf rust (*Puccinia hordei* Otth). Theor Appl Genet 1998; 97: 1235–1239

Jahoor A., Fischbeck G. Sources of resistance to powdery mildew in barley lines derived from *Hordeum spontaneum* collected in Israel. Plant Breed 1987a; 99: 274–287

Jahoor A., Fischbeck G. Genetic studies of resistance of powdery mildew in barley lines derived from *Hordeum spontaneum* collected from Israel. Plant Breed 1987b; 99: 265–273

Jahoor A., Fischbeck G. Identification of new genes for mildew resistance of barley at the *Mla* locus in lines derived from *Hordeum spontaneum*. Plant Breed 1993; 110: 116–122

Jahoor A., Jacobi A., Schüller C. M. E., Fischbeck G. Genetic and RFLP studies at the Mla locus conferring powdery mildew resistance in barley Theor Appl Genet 1993; 85: 713–718

Jensen J. Coordinator's report: barley chromosome 1H (5). Barley Genet Newsl 2002; 32:141–143

Jensen J., Jørgensen J. H. The barley chromosome 5 linkage map. II. Extension of the map with four loci. Hereditas 1975; 80: 17–26

Jensen J., Jørgensen J. H. Location of the yellow rust resistance gene *Yr4* on barley chromosome 5. Barley Genet Newsl 1997; 7: 37–40

Jensen J., Backes G., Skinnes H., Giese H. Quantitative trait loci for scald resistance in barley localized by a non–interval mapping procedure. Plant Breed 2002; 121: 124–128

Jin Y., Statler G. D., Franckowiak J. D., Steffenson B. J. Linkage between leaf rust resistance genes and morphological markers in barley. Phytopathology 1993; 83:230–233

Kalavacharla V., Stavely J.R., Myers J.R., McClean P.E. Crg, a gene required for Ur–3–mediated rust resistance in common bean, maps to a resistance gene analog cluster. Mol Plant Microbe Interact 2000; 13: 1237–1242

Keller M., Keller B., Schachermayr G., Winzeler M., Schmid J.E., Stamp P., Messmer M.M. Quantitative trait loci for resistance against powdery mildew in a segregating wheat × spelt population. Theor Appl Genet 1999; 98: 903–912

Kema G.H.J., van Silfhout C.H. Genetic variation for virulence and resistance in the wheat *Mycosphaerella graminicola* pathosystem III. Comparative seedling and adult plant experiments. Phytopathology 1997; 87: 266–272

Kicherer S., Backes G., Walther U., Jahoor A. Localising QTLs for leaf rust resistance and agronomic traits in barley (*Hordeum vulgare* L.). Theor Appl Genet 2000; 100: 881–888

Kjær B. J., Jensen J., Giese H. Quantitative trait loci or heading date and straw characters in barley. Genome 1995; 38:1098–1104

Kleinhofs A., Kilian A., Maroof M. A. S., Biyashev R. M., Hayes P. M., Chen F. Q., Lapitan N., Fenwick A., Blake T. K., Kanazin V., *et al.* A molecular, isozyme and morphological map of the barley (*Hordeum vulgare*) genome. Theor Appl Genet 1993; 86:705–712.

Kloppers F.J., Pretorius Z.A. Effects of consiuations amongst genes Lr13, Lr34 and Lr37 on components of resistance in wheat to leaf rust. Plant Pathol 1997; 46:737–750

Knott D.R. The effect of transfers of alien genes for leaf rust resistance on the agronomic and quality characteristics of wheat. Euphytica 1989; 44: 65–72

Kolb F.L., Bai G–H., Muehlbauer G.J., Anderson J.A., Smith K.P., Fedak G. Host plant resistance genes for fusarium head blight: Mapping and manipulation with molecular markers. Crop Sci 2001; 41: 611–619

Kolmer J.A. Genetics of resistance to wheat leaf rust. Annu Rev Phytopathol 1996; 34: 435–455

Langridge P., Karakousis A., Kretschmer J., Collins N., Manning S. A consensus linkage map of barley. Mol Breed 1995; 1:389

Law C.N., Snape J.W., Worland A.J. Aneuploidy in wheat and its uses in genetic analysis. In: F.G.H. Lupton (ed.) *Wheat Breeding, Its Scientific Basis*, Chapman and Hall, London, 1987, pp. 71–108

Leister D., Kurth J., Laurie D.A., Yano M., Sasaki T., Devos K., Graner A., Schulze–Lefert P. Rapid re–organization of resistance gene homologs in cereal genomes. Proc Natl Acad Sci USA 1998; 95:370–375

Leonard–Schippers C., Gieffers W., Schäfer–Pregl R., Ritter E., Knapp S.J., Salamini F., Gebhardt C. Quantitative resistance to Phytophtora infestans in potato: a case study for QTL mapping in allogamous plant species. Genetics 1994; 137: 67–77

Li Z.–K., Luo L.J., Mei H.W., Paterson A.H., Zhao X.H., Zhong D.B., Wang Y.P., Yu X.Q., Zhu L., Tabien R., *et al.* A "defeated" rice resistance gene acts as a QTL against a virulent strain of *Xanthomonas oryzae* pv *oryzae*. Mol Gen Genet 1999; 261: 58–63

Ling H.–Q., Zhu Y., Keller B. High–resolution mapping of the leaf rust disease resistance gene *Lr1* in wheat and characterisation of BAC clones from the *Lr1* locus. Theor Appl Genet 2003; 106: 875–882

Liu Z.–W., Biyashev R.M., Maroof M.A.S. Development of simple sequence repeat DNA markers and their integration into a barley linkage map. Theor Appl Genet 1996; 93: 869–876

Liu Z., Sun Q., Ni Z., Yang T. Development of SCAR markers linked to the *Pm21* gene conferring resistance to powdery mildew in common wheat. Plant Breed 1999; 118: 215–219

Liu S., Griffey C.A., Maroof M.A.S. Identification of molecular markers associated with adult plant resistance to powdery mildew in common wheat cultivar Massey. Crop Sci 2001; 41: 1268–1275

Liu Z., Sun Q., Ni Z., Nevo E., Yang T. Molecular characterization of a novel powdery mildew resistance gene *Pm30* in wheat originating from wild emmer. Euphytica 2002; 123: 21–29

Liu S., Anderson J.A. Marker assisted evaluation of fusarium head blight resistant wheat germplasm. Crop Sci 2003; 43: 760–766

Ma Z.Q., Sorrells M.E., Tanksley S.D. RFLP markers linked to powdery mildew resistance genes *Pm1*, *Pm2*, *Pm3* and *Pm4* in wheat. Genome 1994; 37: 871–875

Ma J., Zhou R., Dong Y., Wang L., Wang X., Jia J. Molecular mapping and detection of the yellow rust resistance gene *Yr26* in wheat transferred from *Triticum turgidum* L. using microsatellite markers. Euphytica 2001; 120: 219–226

Madsen L.H., Collins N.C., Rakwalska M., Backes G., Sandal N., Krusell L., Jensen J., Waterman E.H., Jahoor A., Ayliffe M., *et al.* Barley disease resistance gene analogs of the NBS–LRR class: Identification and mapping. Mol Gen Genet 2003; 269: 150–161

Mago R., Spielmeyer W., Lawrence G. J., Lagudah E.S., Ellis J.G., Pryor A. Identification and mapping of molecular markers linked to rust resistance genes located on chromosome 1RS of rye using wheat–rye translocation lines. Theor Appl Genet 2002; 104: 1317–1324

Mammadow J. A., Zwonitzer J. C., Biyashev R., Griffey C. A., Jin Y., Steffenson B., Maroof M. A. S. Molecular mapping of leaf rust resistance gene *Rph5* in barley. Crop Sci 2003; 43:388–393

Manninen O., Kalendar R. N., Robinson J., Schulman A. H. Application of *BARE*–1 retrotransposon markers to the mapping of a major resistance gene for net blotch in barley. Mol Gen Genet 2000; 264: 325–334

McCartney C.A., Brûle–Babel A.L., Lamari L., Somers D.J. Chromosomal location of a race–specific gene to *Mycosphaerella graminicola* in the spring wheat ST6. Theor Appl Genet 2003; 107: 1181–1186

McDaniel M. E., Hathcock B. R. Linkage of the Pa_4 and Ml_a loci in barley. Plant Breed 1969; 104:318–324

McIntosh R.A., Hart G.E., Devos K.M., Gale M.D., Rogers W.J. Catalogue of Gene Symbols for Wheat, 1998; http://wheat.pw.usda.gov/ggpages/wgc/98

Mohler V. Methodische Entwicklungen zur Genomanalyse bei Gerste und Weizen (Dissertation). TU München, 1997.

Mohler V., Jahoor A. Allele specific amplification of polymorphic sites for the detection of powdery mildew resistance loci in cereals. Theor Appl Genet 1996; 93: 1078–1082

Mohler V., Hsam S.L.K., Zeller F.J., Wenzel G. An STS marker distinguishing the rye–derived powdery mildew resistance alleles at the *Pm8*/*Pm17* locus of common wheat. Plant Breed 2001; 120: 448–450

Naik S., Gill K.S., Rao V.S.P., Gupta V.S., Tamhankar S.A., Pujar S., Gill B.S., Ranjekar P. K. Identification of a STS marker linked to the *Aegilops speltoides*–derived leaf rust resistance gene *Lr28* in wheat. Theor Appl Genet 1998; 97: 535–540.

Negassa M. Genetics of resistance to powdery mildew in some Ethiopian barleys. Hereditas 1985; 102:123–138

Nelson J. C., Singh R. P., Autrique J. E., Sorrells M. E. Mapping genes conferring and suppressing leaf rust resistance in wheat. Crop Sci 1997; 37: 1928–1935

Nover I., Schwarzbach E. Inheritance studies with a mildew resistant barley mutant. Barley Genet Newsl 1971; 1: 36–37

Otto C.D., Kianian S.F., Elias E.M., Stack R.W., Joppa L.R. Genetic dissection of a major fusarium head blight QTL in tetraploid wheat. Plant Mol. Biol. 2002; 48: 625–632

Park R. F., Karakousis A. Characterization and mapping of gene *Rph19* conferring resistance to *Puccinia hordei* in the cultivar 'Reka 1' and several Australian barleys. Plant Breed 2002; 121: 232–236

Parlevliet J. E. The genetis of seedling resistance to leaf rust, *Puccinia hordei* Otth. in some spring barley cultivars. Euphytica 1976; 25: 249–254.

Parry D.W.P., Jenkinson P., McLeod L. Fusarium ear blight (scab) in small grain cereals–a review. Plant Pathol 1995; 44: 207–238

Patil V. Genetics of *Rhynchosporium secalis* (Oud.) J. J. Davis resistance in barley (*Hordeum vulgare* L.). Agricultural University of Norway, Ås, Norway, 2001

Patil V., Bjørnstad.Å, MacKey J. Molecular mapping of a new gene $Rrs4_{\text{CI 11549}}$ for resistance to barley scald (*Rhynchosporium secalis*). Mol Breed 2003; 12:169–183

Pecchioni N., Faccioli P., Toubia–Rahme H., Vale G., Terzi V. Quantitative resistance to barley leaf stripe (*Pyrenophora graminea*) is dominated by one major locus. Theor Appl Genet 1996; 93: 97–101

Pecchioni N., Vale G., Toubia–Rahme H., Faccioli P., Terzi V., Delogu G. Barley–*Pyrenophora graminea* interaction: QTL analysis and gene mapping. Plant Breed 1999; 118:29–35

Peng J.H., Fahima T., Röder M.S., Huang Q.Y., Dahan A., Li Y.C., Grama A., Nevo E. High–density molecular map of chromosome region harbouring stripe–rust resistance genes *YrH52* and *Yr15* derived from wild emmer wheat, *Triticum dicoccoides*. Genetica 2000; 109: 199–210.

Prins R., Groenewald J.Z., Marais G.F., Snape J.W., Koebner R.M.D. AFLP and STS tagging of *Lr19*, a gene conferring resistance to leaf rust in wheat. Theor Appl Genet 2001; 103: 618–624

Procunier J.D., Townley–Smith T.F., Fox S., Prashar S., Gray M., Kim W.K., Czarnecki E., Dyck P.L. PCR–based RAPD/DGGE markers linked to leaf rust resistance genes *Lr29* and *Lr25* in wheat (*Triticum aestivum* L.). Jour Genet Breed. 1995; 49: 87–92

Qi L., Cao M., Chen P., Li W., Liu D. Identification, mapping, and application of polymorphic DNA associated with resistance gene *Pm21* of wheat. Genome 1996; 39: 191–197

Qi X., Stam P., Lindhout P. Comparison and integration of four barley genetic maps. Genome 1996; 39: 379–394

Qi X., Niks R. E., Stam P., Lindhout P. Identification of QTLs for partial reistance to leaf rust (*Puccinia hordei*) in barley. Theor Appl Genet 1998a; 96: 1205–1215

Qi X., Stam P., Lindhout P. Use of locus–specific AFLP markers to construct a high density molecular map in barley. Theor Appl Genet 1998b; 96: 376–384

Qi X., Jiang G., Chen W., Niks R. E., Stam P., Lindhout P. Isolate–specific QTLs for partial resistance to *Puccinia hordei* in barley. Theor Appl Genet 1999; 99: 877–884

Rajaram S., Pena R.J., Villareal R.L., Mujeeb–Kazi A., Singh R., Gilchrist L. Utilization of wild and cultivated emmer and of diploid wheat relatives in breeding. Israel J Plant Sci 2001; 49: S93–S104 Suppl.

Ramsay L., Macaulay M., Ivanissevich S.D., MacLean K., Cardle L., Fuller J., Edwards K.J., Tuvenson S., Morgante M., Massari A., *et al.* A simple sequence repeat–based linkage map of barley. Genetics 2000; 156: 1997–2005

Raupp W.J., Sukhwinder–Singh, Brown–Guedira G.L., Gill B.S. Cytogenetic and molecular mapping of the leaf rust resistance gene *Lr39* in wheat. Theor Appl Genet 2001; 102: 347–352

Richter K., Schondelmaier J., Jung C. Mapping of quantitative trait loci affecting *Drechslera teres* resistance in barley with molecular markers. Theor Appl Genet 1998; 97:1225–1234.

Roane C. W., Starling T. M. Inheritance of reaction to Puccinia hordei in barley. II. Gene symbols for loci in differential cultivars. Phytopathology 1967; 57:66–68

Robert O., Abelard C., Dedryver F. Identification of molecular markers for the detection of the yellow rust resistance gene *Yr17* in wheat. Mol Breed 1999; 5: 167–175

Röder M.S., Korzum V., Wendehake K., Plaschke J., Tixier M.H., Leory P., Ganal M.W. A microsatellite map of wheat. Genetics 1998; 149: 2007–2023

Rong J.K., Millet E., Manisterski J., Feldman M. A new powdery mildew resistance gene: Introgression from wild emmer into common wheat and RFLP–based mapping. Euphytica 2000; 115: 121–126

Schachermayr G., Feuillet C., Keller B. Molecular markers for the detection of the wheat leaf rust resistance gene *Lr10* in diverse genetic backgrounds. Mol Breed 1997; 3: 65–74

Schönfeld M., Ragni A., Fischbeck G., Jahoor A. RFLP mapping of three new loci for resistance genes to powdery mildew (*Erysiphe graminis* f. sp. *hordei* in barley. Theor Appl Genet 1996; 93: 48–56

Schönfeld R.–M. Identifizierung und molekulare Lokalisierung neuer Resistenzgenloci der Wildgerste (*Hordeum vulgare* ssp. *spontaenum*) gegen Mehltau (*Erysiphe graminis* f.sp. *hordei*). Bericht aus der Agrarwissenschaft (Dissertation TU München), Shaker Verlag, Aachen, 1998

Schüller C., Backes G., Fischbeck G., Jahoor A. RFLP markers to identify the alleles on the *Mla* locus conferring powdery mildew resistance in barley. Theor Appl Genet 1992; 84: 330–338

Schwarz G., Michalek W., Mohler V., Wenzel G., Jahoor A. Chromosome landing at the *Mla* locus in barley (*Hordeum vulgare* L.) by means of high–resolution mapping with AFLP markers. Theor Appl Genet 1999; 98: 521–530

Schweizer G. F., Baumer M., Daniel G., Rugel H., Röder M. S. RFLP markers linked to scald (*Rhynchosporium secalis*) resistance gene *Rh2* in barley. Theor Appl Genet 1995; 90: 920–924

Seah S., Bariana H., Jahier J., Sivasithamparam K., Lagudah E.S. The introgressed segment carrying rust resistance genes *Yr17*, *Lr37* and *Sr38* in wheat can be assayed by a cloned disease resistance gene–like sequence. Theor Appl Genet 2001; 102: 600–605

Seyfarth R., Feuillet C., Schachermayr G., Winzeler M, Keller B. Development of a molecular marker for the adult plant leaf rust resistance gene *Lr35* in wheat. Theor Appl Genet 1999; 99: 554–560

Shaw M.W., Royle D.J. Factors determining the severity of epidemics of *Mycosphaerella graminicola* (*Septoria tritici*) on winter wheat in the UK. Plant Pathol 1993; 42: 882–899

Shen X., Ittu M., Ohm H.W. Quantitative trait loci conditioning resistance to fusarium head blight in wheat line F201R. Crop Sci 2003a; 43: 850–857

Shen X., Zhou M., Lu W., Ohm H. Detection of fusarium head blight resistance QTL in a wheat population using bulked segregant analysis. Theor Appl Genet 2003b; 106: 1041–1047

Shi A.N., Leath S., Murphy J.P. A major gene for powdery mildew resistance transferred to common wheat from wild einkorn wheat. Phytopathology 1998; 88: 144–147

Shi Z.X., Chen X.M., Line R.F., Leung H., Wellings C.R. Development of resistance gene analog polymorphism markers for the *Yr9* gene resistance to wheat stripe rust. Genome 2001; 44: 509–516

Singh N.K., Shepherd K.W., McIntosh R.A. Linkage mapping of genes for resistance to leaf, stem and stripe rusts and *ω*-secalins on the short arm of rye chromosome 1R. Theor Appl Genet 1990; 80: 609–616

Singh R.P., Nelson J.C., Sorrells M.E. Mapping *Yr28* and other genes for resistance to stripe rust in wheat. Crop Sci 2000; 40: 1148–1155

Singrün Ch., Hsam S.L.K., Hartl L., Zeller F.J., Mohler V. Powdery mildew resistance gene *Pm22* in cultivar Virest is a member of the complex *Pm1* locus in common wheat (*Triticum aestivum* L. em Thell.). Theor Appl Genet 2003; 106: 1420–1424

Smith P.H., Koebner R.M.D., Boyd L.A. The development of a STS marker linked to a yellow rust reistance derived from the wheat cultivar Moro. Theor Appl Genet 2002; 104: 1278–1282

Somasco O.A., Qualset C.O., Gilchrist D.G. Single–gene resistance to septoria tritici blotch in the spring wheat cultivar 'Tadinia'. Plant Breed 1996; 115: 261–267

Somers D.J., Fedak G., Savard M. Molecular mapping of novel genes controlling *Fusarium* head blight resistance and deoxynivalenol accumulation in spring wheat. Genome 2003; 46: 555–564

Sourdille P., Robe P., Tixier M.–H., Doussinault G., Pavoine M.–T., Bernard M. Location of *Pm3g*, a powdery mildew resistance allele in wheat, by using a monosomic analysis and by identifying associated molecular markers. Euphytica 1999; 110: 193–198

Spaner D., Shugar L.P., Choo T.M., Falak I., Briggs K.G., Legge W.G., Falk D.E., Ullrich S.E., Tinker N.A., Steffenson B.J., *et al.* Mapping of disease resistance loci in barley on the basis of visual assessment of naturally occurring symptoms. Crop Sci 1998; 38: 843–850

Suenaga K., Singh R.P., Huerta–Espino J., William H.M. Microsatellite markers for genes *Lr34/Yr18* and other quantitative trait loci for leaf rust and stripe rust resistance in bread wheat. Phytopathology 2003; 93: 881–890

Sun G.L., Fahima T., Korol A.B., Turpeinen T., Grama A., Ronin Y.I., Nevo E. Identification of molecular markers linked to the *Yr15* stripe rust resistance gene of wheat originated in wild emmer wheat, *Triticum dicoccoides*. Theor Appl Genet 1997; 95: 622–628

Tacconi G., Cattivelli L., Faccini N., Pecchioni N., Stanca A. M., Valé G. Identification and mapping of a new leaf stripe resistance gene in barley (*Hordeum vulgare* L.). Theor Appl Genet 2001; 102: 1286–1291

Tan B. H. A new gene for resistance to Puccinia hordei in certain Ethiopian barleys. Cereal Rusts Bull 1977; 5: 39–43

Tanksley S.D., Nelson J.C. Advanced backcross QTL analysis: a method for simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding material. Theor Appl Genet 1996; 92: 191–203

Tao W., Liu D., Liu J., Feng Y., Chen P. Genetic mapping of the powdery mildew resistance gene *Pm6* in wheat by RFLP analysis. Theor Appl Genet 2000; 100: 564–568

Thomas W. T. B., Powell W., Waugh R., Chalmers K. J., Barua U. M., Jack P., Lea V., Forster B. P., Swanston J. S., Ellis R. P., *et al.* Detection of quantitative trait loci for agronomic, yield, grain and disease characters in spring barley (*Hordeum vulgare* L.). Theor Appl Genet 1995; 91: 1037–1047.

Thomsen S. B., Jensen H. P., Jensen J., Skou J. P., Jørgensen J. H. Localization of a resistance gene and identification of sources of resistance to barley leaf stripe. Plant Breed 1997; 116:455–459

Toojinda T., Baird E., Booth A., Broers L. H., Hayes P.M., Bos I., Powell W., Thomas W., Vivar H., Young G. Introgression of quantitative trait loci (QTLs) determining stripe rust resistance in barley: an example of marker–assisted line development. Theor Appl Genet 1998; 96:123–131

Toojinda T., Broers L.H., Chen X. M., Hayes P. M., Kleinhofs A., Korte J., Kudrna D., Leung H., Line R.F., Powell W., *et al.* Mapping quantitative and qualitative disease resistance genes in a doubled haploid population of barley (*Hordeum vulgare*). Theor Appl Genet 2000; 101: 580–589

Tuleen N. A., McDaniel M. E. Location of genes *P* and *Pa5*. Barley Newsl 1971; 15: 106–107

van Berloo R., Aalbers H., Werkman A., Niks R. E. Resistance QTL confirmed through development of QTL–NILs for barley leaf rust resistance. Mol Breed 2001; 8: 187–195

van der Plank J.E. *Disease Resistance in Plants*, Academic Press, New York, USA, 1968

van der Voort J.R., Lindeman W., Folkertsma R., Hutten R., Overmars H., van der Vossen E., Jacobsen E., Bakker J. A QTL for broad–spectrum resistance to cyst nematode species (*Globodera* spp.) maps to a resistance gene cluster in potato. Theor Appl Genet 1998; 96: 654–661

van Eeuwijk F.A., Mesterhazi A., Kling C.I., Ruckenbauer P., Sauer L., Bürstmayer H., Lemmens M., Keizer L. C.P., Maurin N., Snijders C.H.A. Assessing non–specificity of resistance in wheat to head blight caused by inoculation with European strains of *Fusarium culmorum, F. graminearum* and *F. nivale* using a multiplicative model for interaction. Theor Appl Genet 1995; 90: 221–228.

Waldron B.L., Moreno–Sevilla B., Anderson J.A., Stack R.W., Frohberg R. C. RFLP mapping of QTL for Fusarium head blight resistance in wheat. Crop Sci 1999; 39: 805–811.

Wang L., Ma J., Zhou R., Wang X., Jia J. 2002. Molecular tagging of the yellow rust resistance gene *Yr10* in common wheat, P.I.178383 (*Triticum aestivum L.*). Euphytica 2002; 124: 71–73.

Wang Z.X., Yano M., Yamanouchi U., Iwamoto M., Monna L., Hayasaka H., Katayose Y., Sasaki T. The *Pib* gene for rice blast resistance belongs to the nucleotide binding and leucine–rich repeat class of plant disease resistance genes. Plant J 1999; 19: 55–64

Wei F.S., Gobelman–Werner K., Morroll S.M., Kurth J., Mao L., Wing R., Leister D., Schulze–Lefert P., Wise R.P. The *Mla* (powdery mildew) resistance cluster is associated with three NBS–LRR gene families and suppressed recombination within a 240–kb DNA interval on chromosome 5S (1HS) of barley. Genetics 1999; 153: 1929–1948

Wellings C.R., McIntosh R.A. *Puccinia striiformis* f.sp. *tritici* in Australasia: pathogenic changes during the first 10 years. Plant Pathol 1990; 39: 316–325

Wiberg A. Genetical studies of spontaneous sources of resistance to powdery mildew. Hereditas 1974; 77:89–148

William M., Singh R.P., Huerta–Espino J., Islas S.O., Hoisington D. Molecular marker mapping of leaf rust resistance gene *Lr46* and its association with stripe rust resistance gene *Yr29* in wheat. Phytopathology 2003; 93: 153–159

Williams K. J., Lichon A., Gianquitto P., Kretschmer J. M., Karakousis A., Manning S., Langridge P., Wallwork H. Identification and mapping of a gene conferring resistance to the spot form of net blotch (*Pyrenophora teres* f *maculata*) in barley. Theor Appl Genet 1999; 99: 323–327

Wilson, R.E. Inheritance of resistance to *Septoria tritici* in wheat. In: A.L. Scharen (ed.) *Septoria of Cereals: Proceedings of the Workshop (August 2–4, 1983)*, Montana State University, 1985, pp. 33–35

Yan G.P., Chen X.M., Line R.F., Wellings C.R. Resistance gene–analog polymorphism markers co–segregating with the *Yr5* gene for resistance to wheat stripe rust. Theor Appl Genet 2003; 106: 636–643

Yoshimura S., Yamanouchi U., Katayose Y., Toki S., Wang Z.X., Kono I., Kurata N., Yano M., Iwata N., Sasaki T. Expression of *Xa1*, a bacterial blight–resistance gene in rice, is induced by bacterial inoculation. Proc Natl Acad Sci USA 1998; 95: 1663–1668

Yu Y.G., Buss G.R., Maroof M.A.S. Isolation of a superfamily of candidate disease–resistance genes in soybean based on a conserved nucleotide–binding site. Proc Natl Acad Sci USA 1996; 93: 11,751–11,756

Zeller F.J., Kong L., Hartl L., Mohler V., Hsam S.L.K. Chromosomal location of genes for resistance to powdery mildew in common wheat (*Triticum aestivum* L. em Thell.) 7. Gene *Pm29* in line Pova. Euphytica 2002; 123: 187–194

Zhong S., Effertz R. J., Jin Y., Franckowiak J. D., Steffenson B. J. Molecular mapping of the leaf rust resistance gene *Rph6* in barley and its linkage relationships with *Rph5* and *Rph7*. Phytopathology 2003; 93:604–609

Zhou F., Kurth J., Wei F., Elliott C., Vale G., Yahiaoui N., Keller B., Somerville S., Wise R., Schulze–Lefert P. Cell–autonomous expression of barley *Mla1* confers race–specific resistance to the powdery mildew fungus via a *Rar1*–independent signaling pathway. Plant Cell 2001; 13: 337–350

Zhou W., Kolb F.L., Bai G., Shaner G., Domier L.L. Genetic analysis of scab resistance QTL in wheat with microsatellite and AFLP markers. Genome 2002; 45: 719–727

Zhou W.–C., Kolb F.L., Bai G.–H., Domier L.L., Boze L.K., Smith N.J. Validation of a major scab resistance with SSR markers and use of marker–assisted selection in wheat. Plant Breed 2003; 122: 40–46

Chapter 9

QTLs AND GENES FOR TOLERANCE TO ABIOTIC STRESS IN CEREALS

Roberto Tuberosa* and Silvio Salvi

Department of Agroenvironmental Sciences and Technology, University of Bologna, Viale Fanin 44, 40127 Bologna, Italy
*Author for correspondence: tuberosa@agrsci.unibo.it

1. INTRODUCTION

World-wide, abiotic stress curtails production and lowers the quality and nutritional value of the grain in cereal crops. Among all abiotic stresses, drought is the most important from an economic standpoint. Furthermore, with current predictions of increased atmospheric CO_2 and concomitant global warming, the proportion of arid areas of the world is likely to increase. Historically, the improvement of tolerance to abiotic stress has been a major goal in many breeding projects; however, there is no unanimous opinion regarding the best strategy to deploy and which traits to target (Blum, 1988). The slow progress obtained through conventional approaches is largely due (i) to poor knowledge of the genetic and physiological basis of the factors imparting tolerance to abiotic stresses and (ii) to the low heritability of yield and its components under unfavourable conditions (Ceccarelli and Grando, 1996).

Molecular approaches targeting the improvement of tolerance to abiotic stress rely on the application of genetic engineering and marker-assisted selection (Nguyen *et al*., 1997; Khush, 1999; Serageldin, 1999; Cushman and Bohnert, 2000; Bruce *et al*., 2002; Gowda *et al.,* 2003). Genetic engineering has been successfully adopted to manipulate the expression of genes affecting the response to abiotic stress (see Chapter 16 by Sreenivasulu *et al*., in this book); this notwithstanding, to date, the impact of genetic engineering on the release of improved cultivars remains marginal, mainly due to the unpredictable effects of ectopic gene expression on yield and other agronomic traits. This

P.K. Gupta and R.K. Varshney (eds.), Cereal Genomics, 253-315.

unpredictability can in part be traced to the complexity of the morpho-physiological mechanisms underlying the adaptive response of plants to abiotic stress. Indeed, a better understanding of such complexity through the recently introduced genomic approaches would allow us to devise more effective strategies for directly manipulating the genome in order to obtain the desired phenotype in a shorter time and at lower cost.

Because the genetic basis of stress tolerance to abiotic stress is largely quantitative (Blum, 1988; Richards, 1996), the discovery of QTLs (Quantitative Trait Loci) governing the genetic variability of traits influencing such tolerance and their manipulation via marker-assisted selection are of pivotal importance for the successful implementation of genomic approaches and their integration within breeding programmes. Herein, we present and discuss the results of a number of studies in cereals describing QTLs that are important in imparting tolerance to adverse environmental conditions. This knowledge will facilitate more targeted approaches via marker-assisted selection and, in a few cases, will lead to the cloning of the genes underlying major QTLs.

2. MORPHO-PHYSIOLOGICAL BASIS OF TOLERANCE TO ABIOTIC STRESS

Although this review does not aim to analyze in detail the morpho-physiological basis of abiotic stress tolerance, it will summarize some general aspects concerning the interpretation of the molecular information generated to investigate the genetic basis of such traits and to evaluate the applicability of such findings to enhance crops' performance. One of the major challenges for exploiting genomic approaches to improve stress tolerance is to ensure that the molecular information provided by QTL analysis and other 'omics' platforms (e.g., microarrays) is produced within a biological context relevant also to field conditions. As an example, the results of experiments that are carried out under artificially induced conditions of stress to identify candidate genes, may have limited relevance to the conditions present in the field. If this is the case, then the value of such results for crop improvement will inevitably be limited. Therefore, the dynamics (i.e. timing and intensity) of the stress treatment under artificially induced conditions should mimic as closely as possible the prevailing conditions encountered by field-grown plants. This is particularly difficult for those stresses (e.g., drought), which commonly unfold over a fairly prolonged period, thus allowing the plant to activate the molecular mechanisms capable of enhancing tolerance to the incumbent stress.

2.1. Root Architecture

The level of tolerance to different abiotic stresses is influenced by a vast number of morpho-physiological traits for which QTLs have been described. Clearly, for those cases where QTLs have been identified under artificial conditions, their value from a breeding standpoint should be assessed under field conditions. Among the morpho-physiological traits relevant to a number of abiotic stresses, root architecture is certainly one that benefits from a QTL approach, also considering the difficulty of selecting for root traits using conventional approaches. Among all the organs of a plant, the root is probably the most plastic in terms of response to environmental factors, particularly water and nutrient availability. One of the best examples is provided by the effects of P and N availability on root size and root hair density in barley (Gahoonia and Nielsen, 1997). To a varying degree, this plasticity is under genetic control, with each genotype being characterized by its own response to the environmental cue.

2.2. Leaf Growth

An interesting example on how QTL analysis and an ecophysiological model (Ben Haj Salah and Tardieu, 1995; 1996) can be integrated to investigate the genetic basis of leaf growth in resposnse to environmental factors has been recently provided by Reymond *et al.* (2003), who have identified QTLs affecting leaf elongation rate in maize as a function of meristem temperature, water vapour pressure difference and soil water status. In this case, the morpho-physiological variable subjected to QTL analysis was the slope of the regression of the leaf elongation rate on the values of the environmental variables. Consistently with the model originally proposed, all the investigated responses were found to be linear and common to different experiments. Of several QTLs identified, most were specific for the response to one variable only. Remarkably, a model based on the combined QTL effects predicted 74% of the variability for leaf elongation rate measured among a random sample of recombinant inbred lines of the mapping population used for QTL detection (Reymond *et al.*, 2003). This example clearly indicates that it is possible to conjugate QTL information and physiological modelling for predicting leaf elongation in maize, irrespectively of fluctuating environmental factors whose effect can be accounted for by the model. Therefore, crop modelling based on QTLs for morpho-physiological traits responding to measurable environmental variables may help us understand and resolve genotype × environment (G × E) interactions under conditions of abiotic stress; in turn, this will offer new opportunities for

collaboration between modellers and geneticists in ideotype breeding for high yield (Yin *et al*., 2003) and may help to dissect yield into characters that might be under simpler genetic control (Asseng *et al.,* 2002; Asseng *et al.,* 2003).

2.3. Abscisic Acid and Osmolytes

At the biochemical level, extensive work has been carried out in cereals to identify QTLs influencing the concentration of abscisic acid (ABA), often referred to as the "stress hormone" in view of its involvement and pivotal role in the adaptive response to drought and other types of abiotic stress (Zeevaart and Creelman, 1988; Quarrie, 1991; Sharp *et al*., 1994). So far, conflicting evidence has been obtained on the association of yield level with genetic variability in the capacity to accumulate ABA under drought-stressed field conditions (Innes *et al.,* 1984; Blum, 1988; Quarrie, 1991, 1996; Landi *et al*., 2001; Tuberosa *et al.*, 2002b). Another biochemical trait for which conflicting evidence has accumulated is the capacity to accumulate osmolytes upon exposure to dehydration and its association with yield (Blum, 1988; Serraj and Sinclair, 2002). Dehydration is commonly experienced by plants exposed not only to drought, but also to salinity, heat or low temperatures. Even though the interpretation of osmotic relations in genetically engineered plants can be cumbersome (Blum *et al*., 1996), transformation experiments have shed light on the mechanisms by which plants may benefit from an altered capacity to accumulate osmolytes (see Chapter 16 by Sreenivasulu *et al*., in this book). It seems plausible that the tradeoff between the metabolic requirements of osmotic adjustment and the potential benefits for the crop varies on a case-by-case basis as a function of the crop as well as the dynamics and severity of the dehydration episode(s).

2.4. Carbon Isotope Discrimination ($\delta^{13}C$)

Another drought-related trait that has sparked considerable interest for an indirect selection to improve cereals' yield is the carbon isotope discrimination ($\delta^{13}C$), which measures the ratio of stable carbon isotopes ($^{13}C/^{12}C$) in the plant dry matter compared to the value of the same ratio in the atmosphere (Farquar *et al.*, 1989). Because of the differences in leaf anatomy and the mechanisms of carbon fixation in species with the C_3 or C_4 pathway, studies on $\delta^{13}C$ have wider implications for C_3 species where, compared to C_4 species, the variation in $\delta^{13}C$ is larger and has a greater impact on crop yield (Farquar *et al.*, 1989). $\delta^{13}C$ is negatively associated with transpiration

efficiency over the period during which the dry mass accumulates (Turner, 1997; Condon *et al*., 2002). Under drought stress, $\delta^{13}C$ is a good predictor of stomatal conductance and water-use efficiency (WUE) in different crops (Turner, 1997; Araus *et al.*, 2002). A number of studies conducted under varying conditions of water availability have shown that in bread wheat the correlation between $\delta^{13}C$ and final grain yield varies from positive, when ample water is available to the crop, to negative in drought conditions (Condon *et al.*, 1993; Condon *et al.*, 2002). These results can be interpreted based on the influence of both stomatal conductance and photosynthetic activity on $\delta^{13}C$, and on the fact that biomass production is limited in wet years by a lower stomatal conductance, which becomes an advantage under drought conditions. Additional complexity is added when other physiological traits, such as leaf temperature, are considered,. In the field, higher canopy temperatures were exhibited by wheat genotypes characterized by high transpiration efficiency consequent to a lower stomatal conductance (Turner, 1997). Therefore, the relationships between $\delta^{13}C$ and grain yield depend on the environmental conditions, the phenology of the crop and the plant organ (e.g., leaf or grain) from which the samples are collected (Royo *et al.,* 2001; Araus *et al.*, 2002). Substantial genetic variation for grain $\delta^{13}C$ has been reported in cereals (Turner, 1997), with high values for broad-sense heritabilities and a low G × E interaction (Richards, 1996; Ellis *et al*., 2002). For these characteristics, $\delta^{13}C$ is an attractive target for improving WUE and yield. However, the high cost required to measure each sample limits its potential applications and makes it an interesting candidate for marker-assisted selection.

Ultimately, our capacity to identify QTLs for morpho-physiological and biochemical traits affecting stress tolerance will largely depend on accurate phenotyping for the target traits. In general, the closer the trait is to the level of "crop organization", the more influence it will have on productivity. The elusive and unpredictable nature of certain abiotic stresses, drought in particular, translates in rather large effects due to G × E interactions, which lead to low across-environment heritability values and low gain from a selection solely based on phenotypic values. Additionally, when different seasons are analyzed, the level of unpredictability of the environmental variables has to be considered.

3. THE QTL APPROACH

Molecular markers allow us to identify and map the QTLs, i.e. the relevant loci responsible for genetic variability of quantitatively inherited traits (Stuber *et al.*, 1987; Stuber *et al.*, 1999; Paterson *et al*., 1991; Tanksley, 1993; Lee,

1995; Prioul *et al*., 1997; Yano and Sasaki, 1997; Pflieger *et al.,* 2001; Asins, 2002; Morgante and Salamini, 2003). The QTL approach also provides information on the simultaneous effects of a chromosome region on a number of traits. The ultimate goals of QTL analysis are to dissect the complex inheritance of quantitative traits into "Mendelian-like" factors amenable to selection through the analysis of the flanking molecular markers and, eventually, to clone the gene(s) underlying major QTLs.

3.1. Methods for QTL Discovery and Characterization

A number of reviews have discussed the statistical approaches and features of QTL analysis (Tanksley, 1993; Churchill and Doerge, 1994; Jansen and Stam, 1994; Zang, 1994; Lee, 1995; Beavis, 1998; Stuber *et al.*, 1999; Hackett, 2002; Broman *et al.*, 2003). The genetic maps required for QTL identification are assembled (Tanksley, 1993; Liu, 1998) using different categories of molecular markers, more frequently RFLPs, AFLPs and SSRs, each having its own advantages and shortcomings. An important point relates to the ratio between the genetic and physical distance, which, particularly in cereals, varies greatly (up to 300-400 fold) along an individual chromosome (Boyko *et al.*, 2002). This will determine the level of genetic resolution achievable in a specific region of a chromosome.

3.1.1. Linkage Analysis using Mapping Populations

When QTL data are available for two or more mapping populations of the same species, it is possible to compare the positions of QTLs by using anchor markers (usually RFLPs and/or SSRs) common to the linkage maps being analyzed and/or through a comparison with a reference map when the number of common markers mapped in more than one mapping populations is limited. Reference maps are available for many crops (for an example in maize, see Davis *et al.*, 1999), which can be used for the comparative analysis of QTLs, ESTs and mutants across different crosses (for an example on QTLs for root traits in maize, see Tuberosa *et al*., 2003) and even across related taxa, provided that syntenic relationships are established (Gale and Devos, 1998; Cattivelli *et al*., 2002a; Cattivelli *et al*., 2002b; Sorrells *et al*., 2003; Tuberosa *et al*., 2003). A number of these maps have been subdivided into sectors (bins) flanked by reliable RFLP and/or SSR markers, i.e. markers whose map information can be accurately transferred from one segregating population to another. A reference map presenting integrated information on

mutants also allows us to compare the map position of a mutant with that of a QTL, thus contributing relevant information for the identification of possible candidate genes for the investigated trait. On this subject, Robertson (1985) suggested that a mutant phenotype at a particular locus may simply be caused by an allele with a much more drastic phenotypic effect in comparison with that of the QTL alleles at the same locus in the mapping population. In maize, Robertson's hypothesis was validated for a QTL for plant height co-localized with the mutant *dwarf3* (Touzet *et al.*, 1995; Winkler and Helentjaris, 1995) but could not be validated when the position of mutants for ABA biosynthesis was compared to that of QTLs for ABA concentration in the leaf of field-grown plants (Tuberosa *et al.,* 1998b).

3.1.2. Linkage Disequilibrium (LD) and Association Studies in Natural Populations/Germplasm

More recently, QTL identification without the analysis of a mapping population has become possible with the development of analytical approaches exploiting the residual linkage disequilibrium (LD) between markers and closely linked, functionally polymorphic QTLs present in a large number of accessions (Buckler and Thornsberry, 2002; Jansen *et al*., 2003). Under this approach, marker-trait association is only expected when a QTL is tightly linked to the marker because the accumulated recombination events occurring during the development of the lines will prevent the detection of any marker/trait association in any situation where the QTL is not tightly linked to a molecular marker. As compared to allogamous species (e.g., maize), autogamous species such as barley and wheat have higher levels of LD and thus provide better opportunities to detect the presence of major QTLs when a similar number of markers are used. In barley, association mapping has already been successfully applied to identify chromosome regions influencing tolerance to salt stress (Pakniyat *et al*., 1997).

3.1.3. Use of Near-Isogenic Lines (NILs)

A widely adopted strategy to obtain more accurate estimates of chromosomal position and effect of a QTL after its primary identification, relies on the development and testing of nearly-isogenic lines (NILs) differing only for the parental allelic constitution at short chromosome segments (usually from ca. 10 to 20 cM) at the QTL region. Phenotyping and molecular genotyping of NILs will allow us to validate (or disprove) the results obtained from standard interval mapping. NILs will be subsequently crossed to produce a new mapping population, which will be referred to as secondary cross. NILs can

be produced from the same population where the primary QTLs were identified by (i) marker-assisted backcross introgression (or substitution) of one QTL allele into one or both parental genetic backgrounds to obtain the so-called backcross-derived lines (BDLs) or (ii) a number of selfing cycles of the progenies which are heterozygous at the QTL region. Alternatively, NILs can be obtained in a more systematic way, regardless of any primary QTL analysis information, by producing an introgression library (IL), namely a set of introgression lines between a donor and a background genome where each line is isogenic to the background parental line with the exception of a single short (ca. 10-20 cM) chromosome segment introgressed from the donor. Within an IL, the donor genome is completely represented among the different IL lines (ILLs). Such resources have been produced in tomato (Eshed and Zamir, 1995) and *Brassica oleracea* (Ramsay *et al*., 1996) and are being produced in maize (Salvi *et al*., 2003). Koumproglou *et al.* (2002) described and tested in Arabidopsis a refined method for producing introgression lines called STAIRS, where introgressed chromosome segments carry only one recombination event per chromosome. Nearly isogenic material suitable for starting positional cloning is also produced by the Advanced Backcross QTL Analysis (ABQA) method, which can be applied to concurrently identify and transfer favourable QTL alleles from unadapted to cultivated germplasm (Tanksley and Nelson, 1996).

3.2. QTLs for Tolerance to Abiotic Stress in Cereals

Table 1 summarizes the main details of a number of studies on QTLs for tolerance to abiotic stresses and/or related traits in cereals. Although the science of QTL discovery is still in its infancy, a great deal of information has already accumulated. For each type of abiotic stress herein considered, a number of case studies have been analyzed in greater detail and commented on accordingly.

3.2.1. Drought

The majority of calories required for feeding mankind are provided by rice, maize, wheat, sorghum, barley and millets. In most of the least developed countries, these crops are grown under rainfed conditions in environments characterized by relatively low and erratic rainfall. Additionally, drought stress also limits the uptake of nitrogen and other nutrients, thus impairing kernel development and negatively affecting its nutritional value. Because yield stability under arid conditions can, in principle, be expedited by the proper use of molecular markers linked to major QTLs (Ribaut *et al*., 2002),

Table 1. Main features of studies investigating QTLs for tolerance to abiotic stress in cereals

Species	Cross	Pop.	Environment	Main traits	Reference
I. Drought					
Rice	Co39 × Moroberekan	RIL	Greenhouse	Root traits, leaf rolling, osmotic adj., grain yield	Champoux *et al*. (1995) Lilley *et al*. (1996)
	IR20 × 63-83	F2	Pots	Leaf size, ABA, tiller number	Quarrie *et al*. (1997)
	IR64 × Azucena	DH	Field	Grain yield, shoot and root traits	Yadav *et al*. (1997) Venuprasad *et al*. (2002)
			Pot with wax layers	Root traits	Zheng *et al*. (2000)
			Field	Leaf roll., leaf size, RWC, growth rate	Courtois *et al*. (2000)
	Bala × Azucena	F2, RIL	Field	Leaf rolling, stomatal conductance, flowering, RWC	Price *et al*. (1997, 2002a)
		RIL	Soil-filled boxes, pots	Root traits, osmotic adjustment	Price and Tomos (1997) Price *et al*. (2000, 2002c)
	CT9993-.× IR62266-.	DH	Greenhouse	Cell-membrane stability	Tripathy *et al*. (2000)
			Field	Root morphology, plant height, grain yield	Babu *et al*. (2003)
	IR58821-. × IR52561-.	RIL	Greenhouse	Root traits	Ali *et al*. (2000)
	Caiapo × *O. rufipogon*	BDL	Field	Plant height, grain yield	Moncada *et al*. (2001)
	IR1552 × Azucena	RIL	Filter paper	Seminal root length	Zhang *et al*. (2001)
	Nipponbare × Kasalath	RIL	Greenhouse	Stomatal density, leaf rolling, osmotic adjustment	Ishimaru *et al*. (2001b)
	IR62266-. × IR60080-.	BDL	Greenhouse	Osmotic adjustment	Robin *et al*. (2003)
	IAC165 × Co39	RIL	Greenhouse	Root traits	Courtois *et al*. (2003)
Sorghum	B35 × TX7078	RIL	Field	Stay-green, grain yield	Tuinstra *et al*. (1997, 1998)
	B35 × TX430	RIL	Field	Stay-green	Crasta *et al*. (1999)

Table 1. Continued

Sorghum	B35 × TX7000	RIL	Field	Stay-green	Subudhi *et al*. (2000), Xu *et al*. (2000), Sanchez *et al*. (2002)
	SC56 × TX7000	RIL	Field	Stay-green	Kebede *et al*. (2001)
	IS9830 × E36-1, N13 × E36-1	RIL	Field	Stay-green	Haussmann *et al*. (2002)
Maize	F2 × Polj17	F_2	Greenhouse	L-ABA, X-ABA, stomatal cond., root traits	Lebreton *et al*. (1995)
	Ac7643S × Ac7729/TZ.	F_3	Field	ASI, grain yield, yield components	Ribaut *et al*. (1996, 1997b,c)
		RIL	Hydroponics	Root traits	Tuberosa *et al*. (2002c)
	Os420 × IABO78	$F_{3:4}$	Field	L-ABA, leaf temperature, RWC, stomatal conductance, grain yield	Tuberosa *et al*. (1998b) Sanguineti *et al*. (1999)
	A632 × B73	F_3	Field	L-ABA, flowering time	Tuberosa *et al*. (1998a)
	B73 × H99	RIL	Field	ASI, plant height, yield components	Frova *et al*. (1999) Sari-Gorla *et al*. (1999)
	F2 × Io	RIL	Greenhouse	Invertase activity, carbohydrates	Pelleschi *et al*. (1999)
		RIL	Field, greenhouse	Leaf elongation rate, X-ABA	Reymond *et al*. (2003)
	Lo964 × Lo1016	F_3	Field, hydroponics	Root traits, ASI, grain yield	Landi *et al*. (2002b) Tuberosa *et al*. (2002c)
Barley	Dicktoo × Morex	RIL	Pots	L-ABA	Sanguineti *et al*. (1994)
	Tadmor × Er/Apm	RIL	Field	CID, growth habit, osmotic adjustment , RWC, grain yield, yield components	Teulat *et al*. (1997, 1998, 2001a,b, 2002, 2003)
	Arta × *H. spontaneum* 41-1	RIL	Field	Plant height, grain yield	Baum *et al*. (2003)
	Barke × *H. spontaneum*	DH-BC	Field	Growth habit, plant height, grain yield, yield components	Talamè *et al*. (2004)
Wheat	SQ1 × Chinese Spring	DH	Field	WUE, NUE, yield	Quarrie *et al*. (1995, 2004)
	W7984 × Opata	RIL	Field	WUE	Zhang and Ping (2002)

Table 1. Continued

Millet	H 77/833–2 × PRLT 2/89–33	F_2, TC	Field	Grain yield, stover yield	Yadav *et al.* (2002)
II. Flooding					
Rice	IR40931-26 × PI543851	F_2, F_3	Submerged pots	Plant survival	Xu and Mackill (1996),
	DX18121 × M-202				Xu *et al.* (2000)
	IR74 × FR13A	RIL	Field	Plant survival	Nandi *et al.* (1997)
	IR74 × Jalmagna	RIL	Greenhouse	Plant elongation	Sripongpangkul *et al.* (2000)
	IR58821 × IR52561	RIL	Pots	Root morphology	Kamoshita *et al.* (2002)
	IR49830 × CT6241	DH	Field	Plant survival, shoot elongation	Toojinda *et al.* (2003)
	FR13A × CT6241	RIL			
	Jao Hom Nin × Khaow Dawk	$F_{2:3}$			
III. Cold					
Rice	Hokuriku × Hyogo-K.	F_2, F_3	Growth chamber	Recovery rate	Misawa *et al.* (2000)
	M-202 × IR50	RIL	Growth chamber	Spikelet sterility	Andaya and Mackill (2003)
Maize	Lo964 × Lo1016	$F_{2:4}$	Greenhouse	Root and seedling traits	Hund *et al.* (2004)
Barley	Dicktoo × Morex	DH	Field	Growth habit, plant survival	Hayes *et al.* (1993),
					Pan *et al.* (1994)
	Arda × Opale	DH	Growth chamber	Plant survival	Tuberosa *et al.* (1997)
Wheat	Chinese Spring × SQ1	DH	Growth chamber	Plant survival, heading date	Dashti *et al.* (2001)
	T. m. mon. × *T. m. aegil.*	RIL	Growth chamber	Regrowth index	Vagujfalvi *et al.* (2003)
IV. Heat					
Maize	T232 × CM37	RIL	*In vitro*	Cellular membrane stability	Ottaviano *et al.* (1991)
			In vitro, pots	HSP expression and pollen germination	Frova and Sari-Gorla (1993, 1994)

Table 1. Continued

V. Salinity					
Rice	IR64 × Azucena	DH	Hydroponics	Seedling traits	Prasad *et al*. (2000)
	IR4630 × IR15324	RIL	Hydroponics	Ions transport	Koyama *et al*. (2001)
Barley	Steptoe × Morex Harrington × TR306	DH	Pots	Germination test and seedling injury test	Mano and Takeda (1997)
	Derkado × B83-12/21/5	DH	Field	Seedling traits, stem weight, tillers	Ellis *et al*. (2002)
VI. Low N					
Rice	Nipponbare × Kasalath	BILs	Field	Rubisco, GS, soluble protein content	Ishimaru *et al*. (2001a) Yamaya *et al*. (2002)
	IR42 × Palawan	F_2	Hydroponics	N uptake, NUE	Fan *et al*. (2001)
Maize	B73 × G79	$F_{2:3}$	Field	NUE	Agrama *et al*. (1999)
	F2 × Io	TC	Field	Grain composition, leaf N, GS, NR, grain yield	Bertin and Gallais (2001), Hirel *et al*. (2001)
VII. High Al					
Rice	IR1552 × Azucena	RIL	Hydroponics	Root length	Wu *et al*. (2000)
	Koshihikari × Kasalath	BIL	Hydroponics	Root length	Ma *et al*. (2002)
	IR64 × *O. rufipogon*	RIL	Filter papers	Root length	Nguyen *et al*. (2003)
Maize	L53 × L1327	$F_{3:4}$	Hydroponics	Seminal root length	Ninamango-Cardenas *et al*. (2003)
Rye	M39A-1–6 × M77A-1	RIL	Petri dishes	Relative tolerance index	Miftahudin *et al*. (2002)
Wheat	Forno × Oberkulmer	RIL	Plastic boxes	Plant survival, seedling growth	Burgos *et al*. (2001)

Table 1. Continued

VIII. Low P					
Rice	IR20 × IR55178-3B-9-3	RIL	Hydroponics	Plant growth and phosphatase activity	Ni *et al*. (1998), Hu *et al*. (2001)
	Nipponbare × Kasalath	NIL	Field, pots	P-uptake, PUE, dry weight	Wissuwa *et al*. (1998, 2002)
	Zhaiyeqing 8 × Jingxi 17	DH	Hydroponics	Root surface area, dry weight	Ming *et al*. (2001)
Maize	B73 × Mo17	RIL	Hydroponics	Shoot and root growth index	Kaeppler *et al*. (2000)
IX. High Fe					
Rice	Nipponbare × Kasalath	BIL	Plastic containers	Leaf bronzing, shoot and root growth index	Wan *et al*. (2003)

Pop.: population, ASI: anthesis silking interval, BIL: backcross inbred lines, CID: carbon isotope discrimination, DH: doubled haploid lines, GS: glutamine synthetase, HSP: heat shock protein, L-ABA: concentration of abscisic acid in the leaf, NR: nitrate reductase, NUE : nitrogen-use efficiency, PUE: phosphorus-use efficiency, RIL: recombinant inbred lines, RWC: relative water content, TC: test-crosses, X-ABA: concentration of abscisic acid in the xylem sap, WUE: water-use efficiency.

discovery of QTLs for WUE and related traits in cereals is particularly important for future productive gains through molecular breeding. Because of its complex genetics, drought tolerance is often considered the most difficult trait to improve. This challenge becomes even greater due to the unpredictability of drought, whose timing and intensity may fluctuate greatly from year to year. For categorizing the mechanisms conferring drought tolerance, this review adopts the nomenclature reported by Ludlow and Muchow (1990), which distinguishes between traits providing escape from drought and those providing resistance to drought, with the latter further categorized in terms of dehydration avoidance and dehydration tolerance. Dehydration avoidance depends on maintenance of turgor through an increase in water uptake and/or reduction in water loss, while dehydration tolerance relies on biochemical mechanisms allowing the cell to tolerate water loss. Greater emphasis is now being laid upon the mechanisms leading to increasing yield *per se*, rather than the characteristics that enhance plant survival under extreme drought, since the latter entails a negative trade-off under less severe circumstances and a lower intrinsic water-use efficiency of the selected materials (Passioura, 2002).

The mechanisms and the traits underlying drought tolerance in cereals and other crops have been discussed in a number of reviews (Boyer, 1982; Ceccarelli, 1984; Blum, 1988 and 1996; Quarrie, 1991; Passioura, 1996, 2002; Richards, 1996 and 2000; Prioul *et al*., 1997; Turner, 1997; Quarrie *et al*., 1999b; Mitra; 2001; Araus *et al*., 2002; Richards *et al*., 2002; Tuberosa *et al*., 2002b; Slafer, 2003). The level of drought tolerance of a particular genotype is determined by a plethora of traits acting and interacting at the biochemical, physiological and morphological levels. While the genes governing the biosynthesis of different compatible solutes have been identified and cloned from a variety of plant species, the genetic basis of most physiological and morphological traits influencing drought tolerance remains elusive. Although these considerations are to a varying degree true for all types of abiotic stresses, they are particularly pertinent for drought.

3.2.1.1. Rice

Because a deep, thick root system positively impacts the yield of upland rice under water stress conditions, great interest has been devoted to the study of QTLs for root architecture and related traits (Lafitte *et al.*, 2002). A rather large set of QTL data for root traits has been obtained from the evaluation of different rice mapping populations (for details, see Table 1). Due to the difficulty of properly scoring root traits, marker-assisted selection could be conveniently deployed for tailoring root architecture (Shen *et al*., 2001). In a pioneering effort, Champoux *et al.* (1995) investigated the overlap of QTLs associated with root morphology and QTLs associated with drought tolerance

in a mapping population derived from a cross between Co39, an *Indica* cultivar of lowland adaptation, and Moroberekan, a traditional *japonica* upland cultivar. In total, 14 chromosomal regions significantly affected field drought avoidance/tolerance and 12 of these QTL regions also influenced root morphology. In a subsequent study, the same mapping population was investigated for the presence of QTLs associated with osmotic adjustment and dehydration tolerance (Lilley *et al.*, 1996). One major QTL for osmotic adjustment and two of the five QTLs influencing dehydration tolerance were found to partially overlap with QTLs for root morphology, thus indicating a genetic association among these traits. Both osmotic adjustment and dehydration tolerance were negatively correlated with root morphological characters associated with drought avoidance.

The rice mapping population that was most extensively investigated for traits influencing WUE has been derived from Bala × Azucena. Bala has a number of shoot-related characteristics for a better adaptation to drought-prone environments, while Azucena has root traits that potentially contribute to drought resistance. As compared to Azucena, Bala has highly sensitive stomata and a greater ability to adjust osmotically; it does not roll its leaves readily, slows growth more rapidly when subjected to drought and has a lower WUE (Price and Tomos, 1997). In a number of studies characterized by a wide range of experimental conditions, Price and co-workers identified and compared several QTLs for root traits, stomatal conductance, leaf relative water content, leaf rolling, leaf drying, etc. (Price and Tomos, 1997; Price *et al.*, 1997; Price *et al.*, 2000; Price *et al.*, 2002a; Price *et al.*, 2002b; Price *et al.*, 2002c). The QTLs for these drought avoidance traits did not overlap with those for root traits as would be expected if all these traits contributed to drought resistance. Therefore, in this population drought avoidance was only partially influenced by QTLs for root-growth; this is attributed to shoot-related mechanisms of drought resistance and to the difficulty of collecting precise data from field trials because of variability in soils and rainfall (Price *et al.,* 2002c). Because QTLs for root morphological traits under low-moisture did not overlap with QTLs for grain yield under well-watered conditions, Price *et al.* (2002c) suggested that marker-assisted selection could combine QTL alleles for higher grain yield with QTL alleles for desirable root morphological traits under low-moisture conditions, thus avoiding yield penalty under favourable conditions.

Results on root QTLs from other rice mapping populations and their utilization to improve drought tolerance have been discussed in other studies (Yadav *et al.*, 1997; Mackill *et al.*, 1999; Courtois *et al.*, 2000; Kamoshita *et al.*, 2002; Lafitte *et al.*, 2002; Courtois *et al.*, 2003), leading to the identification of a number of major QTL regions with a more substantial and

consistent effect in controlling variation in roots and other traits influencing WUE (Ali *et al.*, 2000; Courtois *et al.*, 2000; Tripathy *et al.*, 2000; Zhang *et al.*, 2001a,b; Price 2002b,c; Courtois *et al.*, 2003; Mei *et al*., 2003). In particular, some of the QTLs detected by Courtois *et al.* (2000) for leaf rolling, leaf drying and relative water content overlapped with QTLs influencing root morphology in the same population and QTLs for leaf rolling as reported from other populations. A number of congenic strains (NILs and BDLs) have been derived to test more accurately the effects of single QTLs and for the presence of epistatic interactions (Shen *et al.*, 2001; Courtois *et al.*, 2000; Price *et al.*, 2002c). A recent study on a mapping population from the cross CT9993-5-10-1-M × IR62266-42-6-2 subjected to water stress prior to anthesis identified 47 QTLs for various plant water stress indicators, phenology and production traits (Babu *et al*., 2003). In this case, root traits were positively associated with yield under drought stress, with the region *RG939-RG214* on chromosome 4 showing the strongest effect on root traits and yield. Constitutive root traits, including maximum length, thickness and dry weight of different root layers, were measured by Courtois *et al.* (2003) in a RIL population derived from the cross IAC165 × Co39. For each trait, between one and four main effect QTLs, each accounting for 5 to 25% of the variability, were identified. The most important genomic regions, which carried QTLs for several traits, were found on chromosomes 1, 4, 9, 11 and 12. These QTL locations were in good agreement with previous studies on these traits, confirming the value of these QTL regions in different genetic background.

Osmotic adjustment is one of several characters, which has been traditionally associated with drought tolerance (Blum, 1988; Ludlow and Muchow, 1990; Nuccio *et al*., 1999; Serraj and Sinclair, 2002). In rice, *indica* cultivars have a greater capacity for osmotic adjustment than *japonica* cultivars. An advanced backcross population has been produced using an *Indica* donor, IR62266-42-6-2, to introgress osmotic adjustment into an elite *japonica* cultivar, IR60080-46A. One hundred and fifty BC_3F_3 families evaluated for osmotic adjustment under greenhouse conditions allowed Robin *et al*. (2003) to detect 14 QTLs accounting for 58% of the phenotypic variability. Most, but not all, of the alleles with positive effects came from the donor parent. The QTL locations were in good agreement with previous studies on this trait in rice and in other cereals. Some BC_3F_3 lines carried the favorable alleles at the two markers flanking up to four QTLs. Intercrossing these lines followed by marker-aided selection in their progenies will allow for the recovery of lines with levels of osmotic adjustment equal to that of the donor parent.

3.2.1.2. Sorghum

The vast majority (ca. 80%) of commercial sorghum hybrids in the United

States are grown under rainfed conditions. Although most of them have pre-flowering drought resistance, many are highly susceptible to post-flowering drought, thus making breeding for improving post-flowering drought tolerance an important priority. Drought stress during grain filling causes premature leaf death. Because evaluation of stay-green is difficult and unreliable under field conditions, progress in improving stay-green in sorghum by conventional breeding methods has been quite slow. As a consequence, increasing attention has been devoted to identify QTLs affecting stay-green, a post-flowering drought resistance trait that is also important for maize (Bolaños and Edmeades, 1996; Thomas and Howarth, 2000). The results of QTL studies on drought-related traits in sorghum are also valuable for maize, due to the similarity of the morpho-physiology of these two crops, and due to a high level of synteny between their genomes (Gale and Devos, 1998); additionally, the smaller size of the sorghum genome makes this species particularly attractive for QTL cloning. Significant progress in genome mapping of sorghum has also recently been reported (Menz *et al.*, 2002). The physiological trait associated to drought resistance that has received the main attention is leaf stay-green. The evaluation of 98 RILs obtained from B35 (pre-flowering susceptible but post-flowering tolerant) × TX7078 (pre-flowering tolerant but post-flowering susceptible) allowed for the identification of 13 QTLs affecting one or more post-flowering drought tolerance traits (Tuinstra *et al.*, 1997). Two QTLs characterized by major effects on yield and stay-green under post-flowering drought were shown to affect yield also under fully irrigated conditions. Following these encouraging results, NILs were developed to test the phenotypic effects of three major QTLs affecting agronomic performance in drought and/or well-watered environments (Tuinstra *et al.*, 1998). In most cases, NILs contrasting for a specific QTL allele differed in phenotype as predicted by the previous QTL study. Further analysis indicated that differences in agronomic performance may be associated with a drought tolerance mechanism that also affects heat tolerance. Also, NILs contrasting at QTL marker *tH19/50* differed in yield under both water regimes. The analysis of these NILs indicated that these differences may be influenced by a drought tolerance mechanism that conditions plant water status and the expression of stay-green. NILs contrasted at QTL marker *t329/132* differed in yield and seed weight. In this case, the differences appeared to be caused by two QTLs that are closely linked in repulsion. Stay-green was also investigated in a mapping population developed from B35 × TX7000 and tested in seven environments (Subudhi *et al.*, 2000; Xu *et al.*, 2000). A comparison of the four stay-green QTLs (*Stg1*, *Stg2*, *Stg3* and *Stg4*) discovered in this population with those described in earlier reports (Tuinstra *et al.*, 1997; Crasta *et al.*, 1999) indicated their consistency across genetic backgrounds (Sanchez *et al.*, 2002). Among these four QTLs, *Stg2* is the most

important for the trait stay-green (Subudhi *et al.*, 2000). While *Stg1* and *Stg2* are on linkage group A, *Stg3* and *Stg4* are on linkage groups D and J, respectively. Three of these QTLs influenced also the chlorophyll content under conditions of post-flowering drought stress (Xu *et al.*, 2000). Additionally, the *Stg1* and *Stg2* regions also contain the genes for key photosynthetic enzymes, heat shock proteins, and an ABA-responsive gene, indicating the importance of linkage group A for tolerance to drought and heat-stress as well as yield in sorghum.

For mapping QTLs influencing stay-green, two sorghum RIL populations, each with 226 $F_{3:5}$ lines were developed independently from two crosses IS9830 × E36-1 and N13 × E36-1 (Haussmann *et al.*, 2002). The common parental line, E36-1 of Ethiopian origin, was the source of the enhanced stay-green. Stay-green was evaluated as the green leaf area percentage at 15, 30 and 45 days after flowering. The number of QTLs for the three sets of stay-green data ranged from five to eight, explaining 31 to 42% of the genetic variance. In both RIL populations, stay-green alleles were contributed by both parent lines. Across the three measures of the stay-green trait, three QTLs on linkage groups A, E and G were common to both RIL populations, with the stay-green alleles originating from E36-1. These QTLs were therefore consistent across the tested genetic backgrounds and years.

3.2.1.3. Maize

In maize, as well as in other cereals, the most critical stage in terms of deleterious effects of drought on yield is just prior to and during flowering (Westgate and Boyer, 1985; Jones and Setter, 2000; Saini and Westgate, 2000). Drought at flowering delays the extrusion of silks and increases the interval between pollen shed and silking (i.e. anthesis-silking interval: ASI) as indicated by the negative association of ASI and grain yield under drought (Bolaños and Edmeades, 1993, 1997; Agrama and Moussa, 1996; Chapman and Edmeades, 1999). Because marker-assisted selection has been adopted to select for ASI, the main results obtained so far on the QTL analysis for ASI are reviewed in section 3.4 on MAS. Among the traits that can influence tolerance to drought in maize, extensive work has been carried out to identify QTLs for the concentration of ABA in the leaf (L-ABA) and their associated effects on other drought-related traits and yield (Lebreton *et al.*, 1995; Tuberosa *et al.*, 1998a; Tuberosa *et al.*, 1998b; Sanguineti *et al.*, 1999). Among the 16 QTLs identified in the Os420 × IABO78 background (Tuberosa *et al.*, 1998b), the most important and consistent QTL was mapped near *csu133,* on bin 2.04. The importance of this QTL was validated by Salvi *et al.* (1997) in a parallel study based on the evaluation of F_4 families divergently selected for L-ABA starting from 480 (Os420 × IABO78) F_2 plants (Sanguineti *et al.*, 1996; Landi *et al.*, 2001). For a more targeted

manipulation of the QTLs controlling ABA concentration, it is desirable to dissect the biochemical and physiological bases governing variability for this trait. To this end, Tuberosa *et al.* (1998b) investigated whether mapped mutants affecting ABA biosynthesis might be possible candidates for the QTLs controlling L-ABA. The major rate-determining step for ABA biosynthesis is controlled by *vp14* (Schwartz *et al.*, 1997, Schwartz *et al.*, 2003; Milborrow, 2001), which has been mapped to bin 1.08 (Tan *et al.*, 1997). Based on the RFLP information of markers common to the reference UMC maize map and the Os420 × IABO78 map, it was shown that mutants impaired in ABA biosynthesis did not map within the support intervals of the QTLs influencing L-ABA (Tuberosa *et al.,* 1998b), a result that leaves open the question of what sort of genes may underline these QTLs. Possible candidates include the genes involved in the intensity of the transduction signal associated with turgor loss, a major determinant in the regulation of ABA concentration (Jensen *et al.*, 1996) and/or genes controlling morpho-physiological traits (e.g., root size and architecture, osmotic adjustment, leaf angle, etc.) affecting the water balance of the plant, hence its turgor. Indeed, a fairly extensive overlap among QTLs for L-ABA and QTLs for leaf relative water content was found in Os420 × IABO78: of the 16 QTLs which significantly affected L-ABA, seven QTLs also influenced leaf relative water content (Sanguineti *et al.*, 1999). At 15 of the 16 QTLs, the corresponding relative QTL effects for leaf relative water content and L-ABA were antagonistic (i.e, +/- or -/+), a result, which suggests that in this case L-ABA mainly represented an indicator of the level of drought stress experienced by plants at the time of sampling. This hypothesis was further substantiated by the analysis of the effects shown by QTL regions affecting L-ABA as well as stomatal conductance and/or leaf temperature. Accodingly, when populations divergently selected for L-ABA starting from Os420 × IABO78 and from B88 × Mo17 were evaluated under drought conditions, L-ABA was found to be negatively associated with grain yield (Landi *et al.*, 2001).

Quarrie *et al.* (1999a) reported that recurrent selection for grain yield under drought conditions significantly changed allele frequencies at *csu133* in two populations ("Tuxpeño Sequia" and "Drought Tolerant Population") developed at CIMMYT (Bolaños and Edmeades, 1993; Bolaños *et al.*, 1993). These results further uderline the importance of the QTL region near *csu133* in controlling drought-related traits and yield in maize. Several QTLs were also found to influence ABA concentration in samples of leaf tissue and xylem sap collected from drought-stressed plants in the cross Polj17 × F2 (Lebreton *et al.*, 1995). All chromosomes, with the exception of chromosome 8, harboured QTLs influencing the concentration of ABA. Analogously to what was reported in Os420 × IABO78 (Tuberosa *et al.*, 1998b), also in Polj17 × F2 a major QTL near *csu133* on bin 2.04 affected L-ABA

concentration. The QTLs for xylem ABA concentration showed a poor overlap with those for L-ABA, a result in keeping with the low correlation reported for these two traits in maize (Zhang and Davies, 1990; Tuberosa *et al.*, 1994). In the Polj17 × F2 population investigated for the concentration of ABA in the leaf and in the xylem, Lebreton *et al.* (1995) searched also for QTLs affecting root traits. Four QTLs affected both seminal root number and root number at the base of the stem and seven QTLs influenced root-pulling force (RPF). QTLs for root traits in hydroponics were identified in a mapping population derived from Lo964 × Lo1016 (Tuberosa *et al.*, 2002c). Several QTLs affected primary root length, primary root diameter, primary root weight and/or the weight of the adventitious seminal roots (R2W). The QTL with the most sizeable effects was mapped on chromosome 1 (bin 1.06). In order to verify to what extent some of the QTL regions that influenced root traits in hydroponics also modulated root characteristics in the field, the same mapping population was tested for RPF in three field experiments (Landi *et al.*, 2002b). Among the 19 bins harbouring QTLs for RPF, 11 also harboured a QTL for one or more root traits measured in hydroponics. The most noticeable overlap for QTLs influencing root traits in hydroponics and in the field occurred in bin 1.06. The same bin also harboured QTLs for root traits in Polj17 × F2 (Lebreton *et al.,* 1995), in F288 × F271 (Barriere *et al*., 2001) and in Ac7729 × Ac7643/TZSRW (unpublished results reported in Tuberosa *et al.*, 2003). Additionally, QTLs for ASI and for grain yield under both well-watered and drought-stressed conditions were mapped in the same bin in Lo964 × Lo1016 (Tuberosa *et al.*, 2002c); the same study reported other overlaps between the QTLs for root traits and those affecting grain yield. Among the investigated root traits, QTLs for R2W most frequently and consistently overlapped with QTLs for grain yield. In particular, at four QTL regions (bins 1.06, 1.08, 10.04 and 10.07) a positive association was found between R2W and grain yield. A synopsis of QTLs for root traits, ASI and grain yeld in maize is reported in Tuberosa *et al.* (2002b).

A comparative analysis based on syntenic relationships between rice and maize (Quarrie, 1996) indicated that at least five QTL regions for root traits in the maize population (Polj17 × F2) correspond to QTL regions in rice affecting root characteristics (Champoux *et al*., 1995). The most notable coincidence was between the region near *umc11* on chromosome 1 in maize and the region between *RG104A* and *RG227* on chromosome 3 in rice. This chromosome region of rice has been shown to influence root penetration ability (Price *et al.,* 2000) and root pulling force (Ali *et al.,* 2000).

3.2.1.4. Barley

In barley, carbon discrimination ($\delta^{13}C$) of the shoot tissue has been reported to be more heritable than other seedling traits (Ellis *et al.*, 2002), an observation

similar to what was previously reported in wheat (Ehdaie and Waines, 1994). However, early attempts to develop $\delta^{13}C$ as an indirect assay for yield in barley were unsuccessful (Acevedo, 1993). A possible reason for this is suggested by the results of Ellis *et al.* (2002), who showed that in the spring barley cross Derkado × B83-12/21/5, only one of the three QTLs for yield was associated with a QTL for stem $\delta^{13}C$. A population of 167 RILs developed from the cross Tadmor × Er/Apm has been extensively investigated to identify QTLs for drought-related traits, including $\delta^{13}C$, grain yield and its components (Teulat *et al.*, 1997; Teulat *et al.*, 1998; Teulat *et al.*, 2001a; Teulat *et al.*, 2001b; Teulat *et al.*, 2002; Teulat *et al.*, 2003). Despite the large heterogeneity in water availability among the six Mediterranean environments considered by Teulat *et al.* (2001b), a number of QTLs were consistently detected in several environments, particularly for plant height and kernel weight, thus indicating a prevalently constitutive mode of action of the genes underlying these QTLs. In total, the multiple-environment analysis by Teulat *et al.* (2001b), revealed 24 consistently expressed QTLs, 18 of which were common to previously published work; the remaining six QTLs that were found to be unique either suggested the likely presence of functional polymorphism conferring specific adaptation to Mediterranean conditions or were specific to the Tadmor and Er/Apm genetic background. As to $\delta^{13}C$, ten QTLs were identified in the multiple-environment study (Teulat *et al*., 2002). One QTL was specific to one environment, two showed G × E interaction, six presented main effects across three or two environments and one presented both effects. Heading date did not contribute to the environment and G × E effects acting on $\delta^{13}C$. Seasonal rainfall and the ratio of rainfall to evapo-transpiration made large contributions to the environmental effect, but their influence on G × E was weak. Eight QTLs for $\delta^{13}C$ co-located with QTLs for physiological traits related to plant water status, osmotic adjustment, and/or agronomic traits previously measured on the same population. More recently, the Tadmor × Er/Apm population has been tested in field trials conducted in three Mediterranean countries to identify QTLs for drought-related traits. A total of nine chromosomal regions influenced variability in the relative water content of the leaf. The QTL with the most-consistent effect was mapped on the long arm of chromosome 6H, a region previously reported to control leaf relative water content as well as leaf osmotic potential under water stress and osmotic adjustment in an experiment conducted in growth-chamber conditions with the same population (Teulat *et al.,* 2001a). Therefore, the confirmation of the role of this region in the genetic control of water and turgor status suggests that it can be exploited for breeding purposes in the Mediterranean area.

Wild barley (*Hordeum vulgare* ssp. *spontaneum*) is a valuable source of alleles influencing resistance to abiotic stresses (Ellis *et al.*, 2000; Forster *et*

al., 2000; Ivandic *et al.*, 2000; Robinson *et al.*, 2000; Turpeinen *et al.*, 2001). To this end, 123 BC_1DH lines derived from a *H. vulgare* (cv. Barke) × *H. spontaneum* cross were investigated under rainfed conditions in three Mediterranean countries to identify agronomically favorable QTL alleles contributed by the wild parent (Talamè *et al.*, 2004). Among the 81 putative QTLs found to influence growth habit traits, heading date, plant height, ear length, ear extrusion, grain yield and/or 1000-grain weight, in 43 cases (53%) the wild parental line contributed the alleles with favorable effects. As to grain yield, although the majority (65%) of favourable QTL alleles belonged to *H. vulgare*, at six QTLs the alleles increasing grain yield were contributed by *H. spontaneum*. These results indicate that ABQA provides a useful germplasm enhancement strategy for identifying wild progenitor QTL alleles capable of improving grain yield of the related crop under stressful conditions.

A set of 194 RILs derived from the cross *H. vulgare* (cv. Arta) × *Hordeum spontaneum* 41-1 was evaluated for cold tolerance, days to heading, plant height, total biomass and grain yield at two ICARDA research stations for two years (Baum *et al*., 2003). Under rainfed conditions, in some Mediterranean countries (WANA countries), plant height is a very important trait, since straw can still provide a meaningful source of feed for grazing animals in dry years. QTLs for plant height were detected on chromosomes 2H, 3H and 7H. These QTLs, particularly the one on chromosome 3H, showed pleiotropic effects on days to heading, total biomass and grain yield. QTLs associated with adaptation to the Mediterranean environment such as cold tolerance, days to heading and tiller number were also described.

3.2.1.5. Wheat

In bread wheat, extensive work has been carried out by Quarrie and co-workers to discover QTLs influencing WUE and yield in a mapping population of 95 DH lines derived from the cross Chinese Spring × SQ1 and tested under different water regimes (Quarrie *et al*., 1995; Quarrie *et al*., 2004). A $\delta^{13}C$ analysis carried out on kernels harvested for two consecutive years in rain-fed trials identified QTLs on chromosomes 2A (near *Xpsr375.1*), 2B (near *Xcbd453*) and 4D (near *Xpsr575.1*) (Quarrie *et al.*, 1995). No significant effect on $\delta^{13}C$ was found to be associated with *rht-B1,* which has been reported to affect WUE in Australian wheats (Richards and Condon, 1993). In a subsequent study, the DH lines were field tested in 1999 and 2000 under irrigated and non-irrigated conditions (Quarrie *et al.*, 2004). A number of significant yield QTLs were found under both drought and irrigated conditions. QTLs for grain yield coincident with water treatments were identified on chromosomes 1A and 7A. Other yield QTLs were unique to the rainfed or irrigated treatment. In particular, a major QTL for yield only under

drought conditions was present distal on chromosome arm 1AL. Each of the major yield QTLs detected under both rainfed and irrigated conditions was associated with one of the yield components. The major yield QTL detected in both treatments on 7A was coincident with a QTL for grains per ear. These results confirm the presence of a major QTL for yield on chromosome 7A found previously in other field studies with the same population (Quarrie *et al.*, 1995). The major dwarfing gene *rht-B1* detected in this mapping population on chromosome 4BS had no obvious effect on plant height in 2000, in contrast to the results for 1999, which was a much more favourable year for plant growth. The results for 2000 showed that the QTLs governing variability in grain yield under severe drought stress conditions differed from those acting under irrigated conditions.

3.2.2. QTLs for Tolerance to Salinity

It has been reported that high salinity afflicts approximately 20% of the cultivated surface and 33% of irrigated agricultural lands world-wide (Flowers and Yeo, 1995; Foolad, 2004). Additionally, ca. 1% of arable land is lost annually due to excessive salinization. Tolerance to salinity implies maintaining meristematic growth and sustained photosynthesis, a condition achieved by controlling the rate of salt accumulation and sequestering in leaves so that cytoplasmic concentrations do not reach toxic levels while maintaining the supply of carbon to sinks (Munns, 1993). Although a Na^+ (net) uptake transporter has not yet been identified in higher plants, it appears that Na^+ enters cells through high-affinity K^+ carriers or low-affinity K^+ channels, insufficiently selective for K^+ (Amtmann and Sanders, 1999). In contrast to Na^+ influx, more is known about genes responsible for the compartmentalization of Na^+ to the vacuole. This is accomplished by a tonoplast Na^+/H^+ antiporter, whose ectopic over-expression has been shown to increase salinity tolerance in Arabidopsis (Apse *et al.,* 1999). At the crop level, experimental evidence indicates the quantitative nature of salt tolerance and the influence of the environment (Blum, 1988; Rao and McNeilly, 1999; Ashraf, 2002). Additionally, it has been noted that the degree of tolerance usually increases with plant age (Mano and Takeda, 1997; Foolad, 2004). Tolerance to soil salinity in cereals is associated with Na^+ concentration in the shoot. Reducing Na^+ and Cl^- uptake while maintaining K^+ uptake are characteristics that positively affect growth under saline conditions. Salt tolerance may also be a consequence of high growth rate associated with fast progression of development and early flowering (Munns *et al.,* 2000).

3.2.2.1. Rice

In rice, Koyama *et al.* (2001) distinguished between QTLs for the total

quantity of ions in a shoot and QTLs affecting ion concentration in the shoot. The latter coincided with QTLs for vegetative growth (i.e. vigour). Conversely, it was suggested that the QTLs that were independent of vigour were those controlling quantitative variation in ion uptake. These QTLs were found to govern independently Na^+ uptake, K^+ uptake and Na^+/K^+ selectivity. The major QTLs for Na^+ and K^+ uptake mapped on different chromosomes, a result consistent with the mechanistically different uptake pathways for Na^+ and K^+ in salt-stressed rice (i.e. apoplastic leakage and membrane transport, respectively). Based upon knowledge of the underlying mechanisms of ion uptake in rice, Koyama *et al.* (2001) postulated that QTLs for Na^+ transport are more likely to act through the control of root development, whereas QTLs affecting K^+ uptake are likely to act through the structure or regulation of membrane-sited transport components.

3.2.2.2. Barley

In barley, genetic differences in tolerance to saline conditions have been reported (Richards *et al.,* 1987; Slavich *et al.,* 1990), with a number of genes seemingly affecting Na^+ influx and concentration in cells (Schachtman and Liu, 1999). QTLs affecting salt tolerance at germination and the seedling stage were identified in two barley populations (Steptoe × Morex and Harrington × TR306; Mano and Takeda, 1997). In both cases, the most important QTLs occurred at different loci on chromosome 5H. The QTLs for salt tolerance at the seedling stage were located on chromosomes 2H, 1H, 5H and 6H in the DH lines developed from Steptoe × Morex and on chromosome 5H in the DH lines developed from Harrington × TR306. These positions were different from those of QTLs controlling salt tolerance at germination, suggesting that salt tolerance at germination and at the seedling stage were controlled by different loci. Genetic variability for salt tolerance has also been reported in wild barley by Ellis *et al.* (2002), who suggested that this variability could be utilized for the improvement of existing cultivars. In total, Ellis *et al.* (2002) mapped 12 QTLs for seedling traits in a DH population from the barley cross Derkado × B83-12/21/5 tested under saline solution in hydroponics. Major effects were detected near *ari-eGP* (on 5H) and *sdw1* (on 3H) dwarfing genes, which were also found to influence C and N isotopes discrimination ($\delta^{13}C$ and $\delta^{15}N$).

3.2.3. QTLs for Tolerance to Low Nutrients

Tolerance to low nutrients is especially important in countries (e.g., African nations), where farmers do not have the economic resources to purchase chemical fertilizers. Furthermore, an increased cereal production in the past required a ca. 6.0-fold increase in nitrogen (N) fertilization (Messmer *et al.,*

1984) causing widespread environmental damage to surface water. Thus, increasing tolerance to nutrient-deficiency of cereals represents a more cost-effective and long-term sustainable solution than relying on fertilizer application. Therefore, the release of cultivars more tolerant to low N concentration would reduce production costs while also alleviating the problems of environmental pollution due to an excessive use of N.

3.2.3.1. Nitrogen deficiency

(a) *Rice.* A hydroponic system was deployed to locate QTLs for N-absorbing capacity and N-utilization efficiency (NUE) in F_2 seedlings of a paddy rice population derived from the *indica* cross IR42 × Palawan (Fan *et al*., 2001). Two QTLs linked with NH_4 uptake were detected on chromosomes 2 and 5, with significant additive and partially dominant gene effects. Two QTLs with a dominant gene effect on chromosomes 5 and 6 influenced NO_3 uptake, one QTL with significant additive and dominant gene effects on chromosome 12 influenced NUE and one QTL on chromosome 4 showed a significant additive effect on NH_4 uptake by the roots.

Because the ratio of Rubisco to total leaf N in leaves is the main target in improving photosynthesis in plants, their genetic control in rice leaves was studied with a QTL approach based on a population of backcross inbred lines (BILs) derived from the cross of Nipponbare × Kasalath (Ishimaru *et al*., 2001a). QTLs governing Rubisco content did not overlap with QTLs for total leaf N content, indicating that contents of Rubisco and total leaf N are controlled by different genes. QTLs that controlled the ratio of Rubisco to total leaf N were detected, suggesting the possibility of improving NUE using this trait. A QTL for Rubisco content mapped near a QTL for soluble protein content on chromosome 8 at 5 days after heading and on chromosome 9 at 25 days. In both cases, the QTL profiles overlapped accurately, providing strong circumstantial evidence for pleiotropic effects by the same genes. QTLs other than those detected at 5 or 25 days after heading also influenced soluble protein or Rubisco, indicating that these traits are genetically controlled depending on the growth stages of leaves.

Immunocytological studies have suggested that NADH-dependent glutamate synthase (NADH-GOGAT; EC 1.4.1.14) in developing organs of rice is involved in the utilization of glutamine remobilized from senescing organs through the phloem (Yamaya *et al*., 2002). Because the levels of NADH-GOGAT in the sink organs of *indica* cultivars are lower than those of *japonica* cultivars, over-expression of the NADH-GOGAT gene from *japonica* rice was investigated using Kasalath, an *indica* cultivar. Transgenic Kasalath lines over-producing NADH-GOGAT under the control of a NADH-GOGAT promoter of Sasanishiki, a *japonica* rice, showed an increase in grain

weight (80% as a maximum), demonstrating that NADH-GOGAT is a key factor in N utilization and grain filling in rice. The evaluation of a population of 98 BC_1F_6 developed from a cross between Nipponbare (a *japonica* rice) and Kasalath allowed detection of QTLs for the content of NADH-GOGAT and cytosolic glutamine synthetase (GS1; EC 6.3.1.2) (Yamaya *et al*., 2002). Seven QTLs were detected for the content of GS1 protein, which is likely to be involved in the export of N from senescing organs, and six QTLs were detected for NADH-GOGAT. Some of these QTLs were located in QTL regions for various biochemical and agronomic traits affected by N recycling.

(b) ***Maize.*** Agrama *et al.* (1999) tested a maize population derived from B73 × G79 at two levels of N availability, which led to the identification of six QTLs for grain yield at low N availability, and five QTLs for grain yield at high N availability. Only two regions (chromosome bins 9.6-9.7 and 10.04-10.05) harboured QTLs affecting grain yield at both N levels; in both cases, these chromosome regions have been shown to affect root traits of other populations tested in hydroponics and in the field (Tuberosa *et al.,* 2003). Maize QTLs for NUE have similarly been reported during vegetative growth and grain filling of a mapping population grown with (175 kg N/ha) or without N fertilization (N^+ and N^-, respectively; Bertin and Gallais, 2001; Hirel *et al.,* 2001). QTLs controlling traits of vegetative development, grain yield and its components were found to map in clusters. For all measured traits, the QTLs in N^- were a subset of those in N^+, the only exception being grain protein content, where a higher number of QTLs were detected in N^-. QTLs for N-uptake were mainly detected in N^+, whereas QTLs for NUE were mostly detected in N^-.

3.2.3.2. Phosphorus deficiency

Phosphorus (P) deficiency of soils can be a major yield-limiting factor in cereal production. Several studies involving identification of QTL for tolerance to P-deficiency have been conducted in rice, wheat and maize (discussed in this section). The prospects for integrating conventional techniques with a QTL approach to enhance performance of pearl millet and sorghum on soils low in available P were also recently discussed (Hash *et al.,* 2002).

(a) Rice. In rice, a total of 127 DH lines, derived from an *indica* (Zhaiyeqing 8) × *japonica* (Jingxi 17), were evaluated to locate QTLs for root surface area and dry weight under P-deficient conditions at 10, 20 and 30 days after starting the treatment (DAT; Ming *et al*., 2001). For the relative root surface area, one QTL was detected 10 DAT, but not 30 DAT. QTLs associated with root dry weight were detected 20 and 30 DAT. Because the results indicated an interaction between plant development and QTL effects, it was suggested

that root morphological traits should be evaluated 10 DAT, while sensitivity to low P should be determined 30 DAT.

Eighty-four randomly selected lines from a population of 284 RILs derived from a cross between the *indica* varieties IR20 and IR55178-3B-9-3 were used in a hydroponic culture experiment with different P supplies to identify QTLs for the activity of acid phosphatase (AAP) in roots, particularly for P-deficiency stress induced AAP (Psi-AAP) and relative AAP (RAAP) (Hu *et al*., 2001). One QTL for AAP and three QTLs for Psi-AAP were detected on chromosomes 1, 6 and 12. Two QTLs, one each on chromosomes 6 and 12 influenced both RAAP and Psi-AAP. These results indicated that the genetic control of Psi-AAP differed from that for AAP under normal culture.

Using 98 backcross rice inbred lines derived from a *japonica* (Nipponbare; low P uptake) × *indica* (Kasalath; high P uptake) cross grown on P-deficient soil, four QTLs were identified for P uptake, accounting for 55% of the variability (Wissuwa *et al.,* 1998). Three of the four QTLs also influenced seedling dry weight, a trait correlated with P uptake ($r = 0.96$). For both traits, the QTL linked to marker *C443* on chromosome 12 had a major effect. Two of the three QTLs detected for internal P-use efficiency (PUE), including the major one on chromosome 12, coincided with QTLs for P uptake; however, whereas *indica* alleles increased P uptake, they reduced PUE. It was concluded that this was not due to the tight linkage of two genes in repulsion, but it was rather due to an indirect effect of P uptake on PUE, since most lines with high PUE were characterized by very low P uptake and dry weight and experienced extreme P-deficiency stress. Their higher P-use efficiency resulted from highly suboptimal tissue-P concentrations and did not represent a positive adaptation to low P availability. In addition to the major QTL on chromosome 12 already identified for all other traits, two QTLs, one each on chromosomes 4 and 12 were identified for tiller number. Their positions, however, coincided with QTLs for tiller number reported elsewhere under P-sufficient conditions and therefore these QTLs seem to be unrelated to P-deficiency tolerance. Based on their results, Wissuwa *et al*. (1998) concluded that P-deficiency tolerance was mainly caused by differences in P uptake and not in PUE. Subsequently, Wissuwa *et al.* (2002) developed a NIL for the major QTL linked to marker *C443*. As compared to Nipponbare, under P-deficiency in pots, the NIL carrying an introgressed segment from Kasalath in the proximity of *C443* showed a four-fold higher P uptake whereas both showed similar values under optimum P supply. The QTL near *C443*, therefore, affects the mechanisms specifically increasing P uptake under P deficiency. The high plasticity of the root system of NIL-C443 was shown by a seven-fold increase in the root surface area between days 40 and 150 from the beginning of the experiment, whereas roots of Nipponbare hardly doubled

in size and reached only 15% of their non-stress surface area under control conditions. Therefore, the superior P uptake of NIL-C443 was due to its capacity to maintain root growth under P deficiency. In order to map the major QTL near *C443* with greater accuracy, a secondary mapping population was developed by crossing the NIL containing the QTL from the donor parent to the recurrent parent of low P uptake (Wissuwa *et al*., 2002). Two different mapping strategies were followed. A conventional QTL mapping approach based on the phenotypic evaluation of family means in the F_3 and a substitution-mapping approach. QTL mapping showed that close to 80% of the variation between families was due to a single QTL, named *Pup1* (Phosphorus uptake 1) which was placed in a 3 cM interval flanked by markers *S14025* and *S13126*. Other chromosomal regions and epistatic effects were not significant. The two mapping strategies yielded almost identical results and, in combining the advantages of both, *Pup1* could be mapped with a high level of confidence.

(b) Wheat. In hexaploid wheat, a population (ITMI pop) of 114 RILs derived from the cross W7984 × Opata85, were evaluated in a hydroponic experiment under P-sufficient and P-deficient conditions (Weidong *et al*., 2001). PUE for shoot (SPUE), or whole plant (WPUE) was defined as shoot or whole plant biomass production per mg of absorbed P by the whole plant. Under high P conditions, two and three QTLs were identified for SPUE and WPUE, respectively, while under low-P conditions, three and two QTLs influenced SPUE and WPUE, respectively. One digenic epistatic interaction influenced SPUE and WPUE under high-P environment; one digenic interaction was also found for WPUE at P-deficient level. The QTL on chromosome 7A influenced WPUE at both P levels and was involved in the epistasis influencing WPUE under P-sufficient conditions, underlining the importance of this region for PUE in wheat. Additionally, the results indicated that chromosome 5A was more important for PUE under high P-conditions, while chromosome 2D was important at low P-levels.

(c) Maize. Plant growth at low soil P concentration is affected both by physiological factors inherent to the crop as well as by its interactions with the soil biota. Among the different species colonizing the soil, very little is known concerning the role of mycorrhiza in nutrient uptake of crops. In maize, QTLs for growth at low P and response to mycorrhizal fungi were identified in a B73 × Mo17 RIL population (Kaeppler *et al*., 2000). Three QTLs were found to influence growth at low P in the absence of mycorrhizas based on shoot weight and one QTL, which controlled mycorrhizal responsiveness. The authors suggested that this variation could be harnessed to develop cultivars better adapted for regions of the world with P deficiency and for reduced input production systems.

3.2.4. QTLs for Aluminium Tolerance

Aluminium (Al) toxicity is one of the major constraints for plant development in acid soils in many countries (e.g., Brazil). Furthermore, acid soils with high levels of Al retard root growth, causing increased crop sensitivity to drought and impairing nutrient acquisition. Cultivars genetically adapted to cope with low-pH soils may offer a more environmentally compatible solution under such conditions.

3.2.4.1. Rice

Among small-grain cereal crops, rice shows the highest tolerance to Al. The response of rice to Al stress was preliminarily investigated by Ma *et al.* (2002) using a RIL population derived from a cross between the *japonica* cv. Koshihikari and the *indica* cv. Kasalath. As compared to Kasalath, Koshihikari showed higher tolerance to various Al concentrations, and exhibited a lower Al content in root apexes, suggesting that exclusion mechanisms rather than internal detoxification are more active in Koshihikari. Although Al-induced secretion of citrate was observed in both Koshihikari and Kasalath, it is not the mechanism conferring Al tolerance because the two cvs. did not differ in the amount of citrate secreted. The evaluation of a population of 183 backcross inbred lines (BILs) derived from a Koshihikari × Kasalath cross identified three putative QTLs on chromosomes 1, 2 and 6, accounting for 27% of the phenotypic variability for Al tolerance. The existence of QTLs for Al tolerance was confirmed through the evaluation of substitution lines for the corresponding chromosomal segments. In rice, QTLs for Al tolerance were also mapped using a population derived from the cross of an Al-susceptible rice line with an *Oryza rufipogon*, Al-tolerant line evaluated using a nutrient solution with and without 40 ppm of active Al^{3+} (Nguyen *et al.,* 2003). Nine QTLs were identified including one for root length (RL) under non-stress (control) conditions (CRL), three for root length under Al stress (SRL) and five for relative root length (RRL). *O. rufipogon* contributed favourable alleles for each of the five QTLs for RRL, a trait of primary importance for Al tolerance. Comparative genetic analysis showed that QTLs for RRL, which were mapped on chromosomes 1 and 9, appear to be consistent among different rice populations. Interestingly, comparative mapping showed that a major QTL for RRL on rice chromosome 3, which accounted for a large portion (ca. 25%) of the phenotypic variatbility, is also conserved across cereal species.

3.2.4.2. Maize

In maize, five QTLs were mapped on chromosomes 2, 6 and 8, accounting for 60% of the phenotypic variability (Ninamango-Cardenas *et al.,* 2003). The QTL peak near the RFLP marker *umc043* is close to loci encoding for

enzymes involved in the synthesis of organic acids, a widely proposed mechanism for Al tolerance in plants. The QTL on chromosome 2 was mapped in the same region as *Alm2*, a locus, which has also been associated with Al tolerance in maize. Also in this case, the comparison of the map positions of QTLs with those of loci known to influence the target trait has provided a means to identify possible candidate genes for the QTLs.

3.2.4.3. Wheat

A hexaploid wheat population of 91 RILs segregating for Al tolerance was investigated to provide a more detailed genetic linkage map of the chromosome arm 4DL, which previous studies had shown to harbour genes conferring resistance to Al (Milla and Gustafson, 2001). The *Alt(BH)* gene was confined to a 5.9-cM interval between markers *Xgdm125* and *Xpsr914*. In addition, utilizing a set of wheat deletion lines for chromosome arm 4DL, the *Alt(BH)* gene was physically mapped to the distal region of the chromosome where the ratio between the genetic/physical distances appears feasible for the map-based cloning of the gene.

3.2.4.4. Rye

Among the *Triticeae*, rye is considered to be the most Al-tolerant crop. An AFLP screening of a RIL population derived from the cross between an Al-tolerant and an Al-sensitive line showed that a single locus controlled aluminium tolerance (Miftahudin *et al.,* 2002).

3.2.5. QTLs for Tolerance to Flooding

In regions subjected to heavy rainfall, particularly in the tropics, soil waterlogging is a major environmental constraint limiting crops' productivity.

3.2.5.1. Rice

Extensive work has been carried out on rice, since a large portion of the surface cultivated with lowland rice is affected by flooding. Counterintuitively to the widespread belief that rice is highly tolerant to submergence, rice plants are seriously damaged by several days of total submergence. Both submergence tolerance and plant elongation are the major mechanisms that contribute to the adaptation of deepwater rice to flooding. The first major QTL influencing submergence tolerance in rice was reported by Nandi *et al*. (1997) on chromosome 9, while other QTLs were detected on chromosomes 6, 7, 11 and 12. Sripongpangkul *et al.* (2000) identified several genes/QTLs that control plant elongation and submergence tolerance. The most important QTL (*QIne1*), which mapped near *sd-1* on chromosome 1, in a region contributed by the parent adapted to deepwater (cv. Jalmagna),

stimulated internal elongation and contributed significantly to submergence tolerance under flooding. The second locus was a major gene, *sub1* on chromosome 9, which contributed to submergence tolerance only. At the third QTL (*QIne4*) on chromosome 4, the allele from IR74 (non-elongating parent, susceptible to flooding) had a sizeable effect on internal elongation. An additional locus that interacted strongly with both *QIne1* and *QIne4* appeared near *RG403* on chromosome 5, suggesting a complex epistatic relationship among these three loci.

In a recent review, Jackson and Ram (2003) assessed the characteristics of the underwater environment that may damage rice plants and summarized the progress achieved so far to identify QTL alleles from the submergence-tolerant cv. FR13A for improving tolerance to flooding. The available physiological evidence indicates the harmful effects of an imbalance between production and consumption of assimilates. This imbalance is exacerbated by an accelerated leaf extension and senescence, both of which appear to be ethylene-mediated and largely absent from FR13A and derived cvs. More especially, DNA markers flanking a major submergence tolerance QTL were shown to be potentially useful in marker-assisted selection programmes attempting to improve submergence tolerance. Three rice populations, each derived from a single cross between two cultivars differing in their response to submergence, were evaluated under different field conditions, locations and years to identify QTLs governing plant survival and other associated traits (Toojinda *et al*., 2003). A major QTL region on chromosome 9 consistently affected (across years and mapping populations) plant survival, plant height, stimulation of shoot elongation, visual tolerance score and leaf senescence. A number of secondary QTLs influencing flooding tolerance were localized on chromosomes 1, 2, 5, 7, 10 and 11. These secondary QTLs were specific to particular traits, genetic backgrounds or environments.

3.2.5.2. Wheat

In some marginal and more northern agroclimatic zones, wheat often suffers from flooding during the early stages of growth. The effect of waterlogging is more severe for winter wheat (*Triticum aestivum*) as compared to winter spelt (*Triticum spelta*). The response of a wheat × spelt RIL population (Forno x Oberkulmer) to flooding stress in the early phase of germination was investigated by Burgos *et al.* (2001). Lines characterized by greater tolerance to a 48h flooding treatment just after imbibition showed less electrolyte leakage (r = -0.79), indicating greater membrane integrity and better survival. Five QTLs accounting for 41% of the phenotypic variance for survival to flooding mapped on chromosomes 2B, 3B, 5A and 7D. Plants with a fast coleoptile growth during flooding were less susceptible to flooding, as indicated by the fact that flooding tolerance was best correlated with the mean

germination time (r = 0.80). Ten QTLs were found for seedling growth index after flooding, accounting for 36% of the phenotypic variance. These QTLs were localized on chromosomes 2A, 2B, 2D, 3A, 4B, 5A, 5B, 6A and 7D.

3.2.6. QTLs for Tolerance to Low Temperature

Among all crops, cereals are probably exposed to the widest range in temperature. As an example, barley and wheat cultivation spans from the hot and arid lowlands of Mediterranean countries to the cold regions at high elevation or in plains at high latitude in the Northern hemisphere. However, low temperatures can damage cereal production both in the temperate regions and at high elevations in the tropics.

3.2.6.1. Rice

Rice is particularly susceptible to low temperatures at the booting stage, i.e. approximately 10 days before heading. *Japonica* cultivars are known to be more tolerant than *indica*s. A bioassay was deployed by Misawa *et al.* (2000) to evaluate a rice mapping population for the response of seedlings to chilling temperature conditions under which the *indica* parent Hokuriku 142 did not survive, while the *japonica* parent Hyogo-Kitanishiki showed complete survival. QTLs affecting the response to low temperature were detected on chromosomes 1, 3, 9 and 11. The QTLs on chromosomes 3, 9 and 11 showed partial synteny with chromosomes 5 of wheat and 5H of barley carrying genes for freezing tolerance. Andaya and Mackill (2003) tested 191 RILs derived from a cross between a tropical *indica*, IR50 (cold susceptible) and a temperate *japonica*, M-202 (cold tolerant), and measured spikelet sterility of plants grown at 12 °C for 5 days under controlled conditions at the booting stage. QTLs affecting cold tolerance were mapped on chromosomes 1, 2, 3, 5, 6, 7, 9 and 12. At the two QTLs with the largest effect, each one accounting for ca. 17% of the phenotypic variance, the tolerant parent conributed the allele increasing cold tolerance. At two of the eight QTLs affecting cold tolerance, the favourable allele was contributed by IR50.

3.2.6.2. Maize

Improving early vigour of maize is crucial for its adaptation to the climatic conditions of central Europe and the northern Mediterranean, where early sowing is an important strategy for avoiding the negative effects of summer drought. A set of 168 $F_{2:4}$ families of the Lo964 × Lo1016 cross was grown in a sand-vermiculite substrate at 15/13 °C until the one-leaf stage, when they were evaluated for four shoot and two seed traits (Hund *et al*., 2004). In total, 20 QTLs influenced the four shoot and two seed traits that were examined. The speed of germination was shown to be positively associated with the

length of the lateral primary roots. The analysis of the total root weight and of the length and diameter of primary and seminal roots led to the identification of 38 QTLs. Seven of the QTLs reported by Tuberosa *et al*. (2002c) were confirmed to have the same genomic location and a similar additive effect.

3.2.6.3. Barley

In barley, although early work of Hayes *et al.* (1993) identified only one major multilocus cluster of linked QTLs on the long arm of chromosome 5H in the Dicktoo × Morex population, nine QTLs for freezing tolerance were later reported on chromosomes 2H, 3H, 5H and 6H in the Arda x Opale background (Tuberosa *et al.,* 1997). The presence of a locus (*Dhn1*) encoding for a dehydrin within the QTL interval on chromosome 5H, led to the initial suggestion that this locus might influence frost tolerance (Hayes *et al.,* 1993, Pan *et al.* 1994). However, further work did not support this hypothesis. Dehydrins (DHNs) are one of the typical families of plant proteins that accumulate in response to dehydration, low temperature, osmotic stress or ABA treatment, or during seed maturation (Close, 1997). Three genes (*Dhn1*, *Dhn2* and *Dhn9*) encoding low-molecular weight DHNs map within a 15-cM region of barley chromosome 5H that overlaps a QTL for winter-hardiness, while other *Dhn* genes encoding low- and high-molecular weight DHNs are located on chromosomes 3H, 4H and 6H (Choi *et al*., 2000). Zhu *et al*. (2000) examined the expression of specific *Dhn* genes under conditions associated with expression of the winter-hardiness phenotype. Their observations were consistent with the hypothesis that the major chilling-induced DHNs help to prime plant cells for acclimatization to more intense cold, which then entails adaptation to dehydration during the freeze-thaw cycling. A role for chromosome 5H-encoded DHNs in acclimatization to more intense cold seems plausible, even though it is not the basis of the major heritable variability in winter-hardiness that has been observed within the Dicktoo x Morex population.

3.2.6.4. Wheat

In hexaploid wheat, the studies of Sutka and co-workers using chromosome substitution lines pointed out the quantitative nature of frost tolerance in wheat (Sutka and Snape, 1989; Veisz and Sutka, 1993; Sutka, 1994) and provided a useful framework for more informative studies based on QTL analysis. QTL mapping has shown that frost tolerance and vernalization requirements are governed by two tightly linked loci on chromosome 5A, namely *Vrn1* and *Fr1* (Galiba *et al*., 1995; Sutka *et al*., 1999). An additional frost tolerance locus, *Fr2*, has been identified on chromosome 5D in the homeologous region corresponding to that harbouring *Fr1* on chromosome 5A, thus providing evidence for the orthology of these two QTLs (Snape *et al*., 1997). More recently, a new locus for frost tolerance designated *Fr-A2*

has been reported (Vagujfalvi *et al*., 2003). This locus was mapped on the long arm of chromosome 5A of diploid wheat (*Triticum monococcum*), 30 cM proximal to the major frost tolerance locus *Fr-A1*. Remarkably, plants of the frost-tolerant and frost-susceptible parental lines grown at 15 °C differed in the transcription level of the cold-induced gene *Cor14b*. When transcript levels of this gene were determined in each line of the mapping population and mapped as a QTL, its peak precisely overlapped with the QTL peak for frost survival at the *Fr-A2* locus, thus providing strong circumstantial evidence that frost tolerance was mediated by differential regulation of the expression of the *Cor14b* gene. Previous work in hexaploid wheat showed that *Cor14b* is regulated by two loci on chromosome 5A, one in the same chromosome region as the *T. monococcum Fr-A2* locus and the other closely linked to *Fr-A1*. Since CBF transcriptional activators in Arabidopsis regulate *Cor* genes and are involved in frost tolerance, Vagujfalvi *et al.* (2003) mapped the cold-regulated CBF-like barley gene *Cbf3* on the *T. monococcum* map. *Cbf3* was mapped on the peak of the *Fr-A2* QTL for frost tolerance, suggesting that the observed differential regulation of *Cor14b* at the *Fr-A2* locus is due to allelic variation at the *XCbf3* locus, and that this transcriptional activator gene might be a candidate gene for the *Fr-A2* frost tolerance locus on wheat chromosome 5A.

3.2.6.5. Synteny between wheat and barley QTLs

A number of reviews have recently analyzed in greater detail the syntenic relationships of QTLs for freezing tolerance in wheat and barley (Snape *et al.* 2001; Cattivelli *et al.*, 2002a,b), and strategies for their isolation using rice genomic tools have been discussed (Snape *et al*., 2001). A survey of major loci and QTLs influencing tolerance to abiotic stresses in the *Triticeae* pointed out the importance of chromosome group 5 for frost and salt tolerance and chromosome group 7 for drought tolerance (Cattivelli *et al.*, 2002b). Additionally, despite the involvement of a common set of genes in the molecular response to cold and drought, different QTLs underlined the response to these two types of stresses.

3.2.7. QTLs for Tolerance to High Temperature

Cereals are often exposed to episodes of heat stress during the grain filling phase. High temperatures disrupt the deposition of reserve products, thus lowering final yield and the quality of the grain. In maize, a pioneering study by Ottaviano *et al.* (1991) investigated cellular membrane stability (CMSt) as a physiological index to evaluate thermostability in a set of RILs derived from a cross between a thermotolerant (T232) and a thermosensitive (CM37) line. In total, six QTLs were identified accounting for 53% of the genetic

variability for CMSt, whose heritability was 0.73. The same RIL was tested by Frova and Sari-Gorla (1993) to search for QTLs governing variability in heat shock protein (HSP) synthesis. In higher plants, within-species qualitative polymorphism for HSPs is extremely rare, even between genotypes showing different heritable levels of thermotolerance. Significant differences were detected in the level of expression of five HSPs with a varying number of QTLs (from three to eight) accounting from 35 to 60% of the genetic variability observed for these bands. The analysis of the correlation between the variability of HSPs and that of CMSt did not reveal any significant association of the two traits (Frova and Sari-Gorla, 1993).

Beacuse pollen thermotolerance is an important component of the adaptability of crops to high temperature, the degree of injury caused by high temperature to pollen germinability (IPGG) and pollen tube growth (IPTG) was investigated in 45 RILs in maize derived from T232 × CM37 (Frova and Sari-Gorla, 1994). Both traits revealed quantitative variability and high heritability. The chromosomal localization of five QTLs for IPGG and six QTLs for IPTG, indicated that different sets of genes controlled these traits. Additionally, IPGG and IPTG were shown to be basically independent of pollen germination ability and pollen tube growth rate under non-stress conditions.

In hexaploid wheat (*T. aestivum*), two heat-stress related QTLs influencing grain filling duration (GFD) were identified in a mapping population of F_2 plants derived from the cross Ventnor (heat-tolerant) × Karl 92 (heat-susceptible). The two QTLs linked to microsatellites *Xgwm11* and *Xgwm293* accounted for 11 and 12%, respectively, of the total phenotypic variability in GFD (Yang *et al*., 2002).

3.3. Cloning QTLs for Abiotic Stress Tolerance in Cereals

Although QTL cloning is well advanced in mammals (Korstanje and Paigen, 2002), only eight QTLs have so far been formally cloned in plants: three in tomato controlling fruit characteristics (*fw2.2*, *Ovate* and *Brix9-2-5*), three in rice for flowering time (*Hd1*, *Hd6* and *Hd3a*) and one in Arabidopsis for flowering time (*CRY2*). All these QTLs have been cloned through positional cloning by refining their map position to a small physical region and then deploying different approaches to specifically validate the involvement of a specific gene at the target region. In another approach, the maize QTL *Tb1* (*Teosinte-branch1*, controlling inflorescence architecture) was first cloned by using a complementation test between the QTL and a known mutant earlier recognized as one of the possible candidate genes; subsequently, the mutant

allele was physically cloned by transposon tagging. A few more QTLs have been either fine mapped (Salvi *et al*., 2002, 2003) or functionally linked to specific candidate genes (Pelleschi *et al.,* 1999; Hirel *et al.,* 2001; Vagujfalvi *et al.,* 2003). Among the above-mentioned QTLs, those controlling flowering time are relevant for improving tolerance to abiotic stress.

3.3.1. Positional Cloning

Figure 1 outlines the main steps necessary for the positional cloning of a QTL. Further details and comments on QTL cloning are provided by Stein and Graner in Chapter 11 of this volume. After a major QTL has been carefully chosen based on a number of accurate studies, the first step for its positional cloning requires the production of a large secondary cross in a nearly isogenic background where only the target QTL segregates. The large number of plants (ca. 1,500 or more) in the segregating population allows for the recovery of a sufficiently high number of recombination events in the target region, an essential prerequisite to achieve the desired level of map resolution. Another important prerequisite for the successful completion of the high resolution mapping is the availability of markers in the target region. While in the past new markers were usually added through bulk segregant analysis (Salvi *et al.,* 2002), the vast database on synteny between rice and cereals provides a rich source of additional markers. Following the fine mapping, the markers more tightly linked to the QTL are used to anchor the genetic region to a physical chromosome region. Genes or genomic sequences carrying alternative alleles that are found to be completely linked with the QTL are then functionally tested with a number of different approaches (e.g., genetic engineering, identification of knockouts, complementation of mutants and/or expression profiles). Usually, in order to support the validation of a particular candidate gene, circumstantial evidence such as the presence of candidate genes at homologous chromosome positions in highly syntenic species and/or pattern of gene expression from other genotypes is also sought (Glazier *et al.,* 2002).

The fine mapping step based on the cross between congenic strains aims to narrow the QTL position to a genetic interval corresponding, in physical terms, to a DNA sequence including very few genes, and ideally only one, as outlined under the general paradigm known as “chromosome landing” (Tanksley *et al*., 1995). This would reduce or completely avoid the need of chromosome walking. Thus, the step needs a careful design in order to recover an adequately large number of crossover events around the QTL position, to identify new molecular markers closer to the QTL and to unambiguously define the QTL genotype of the scored individuals.

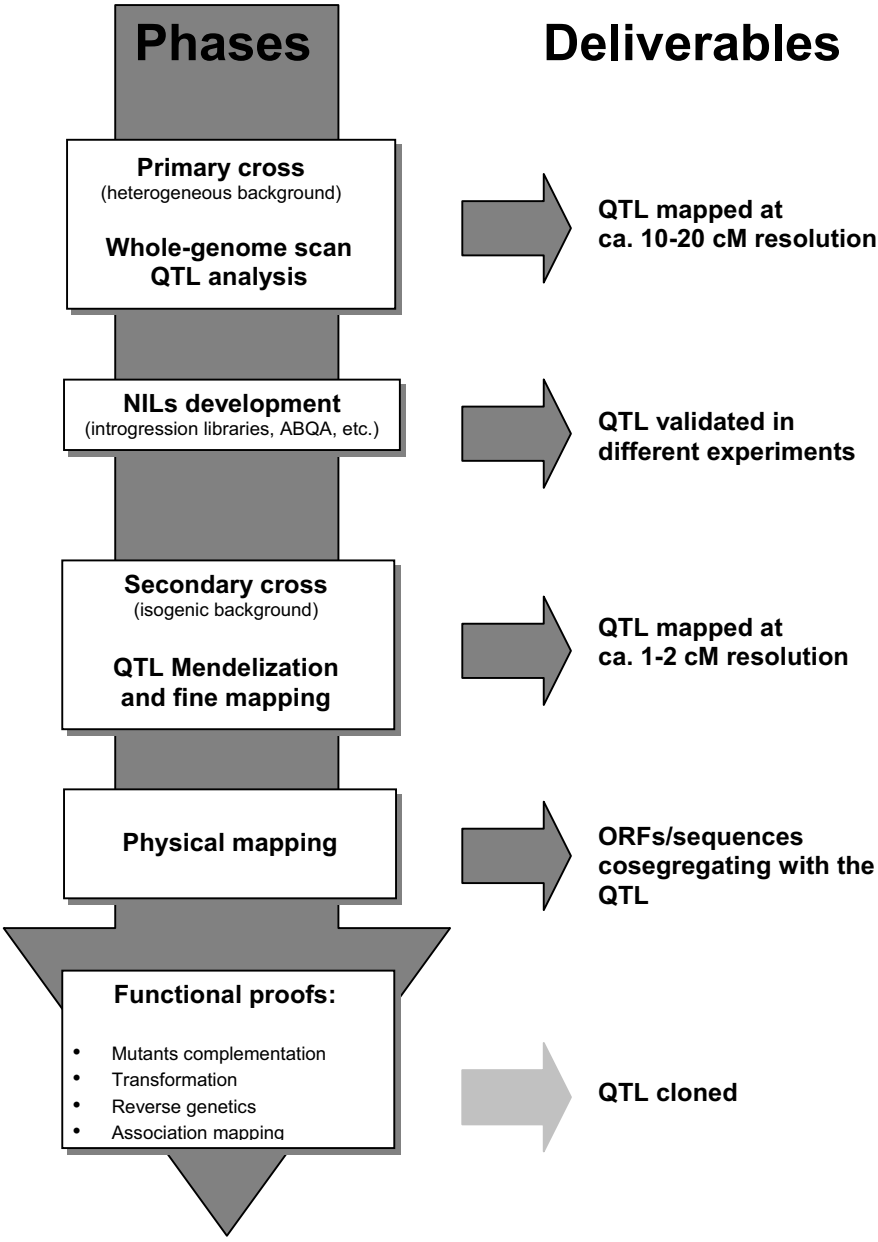


Figure 1. Main steps required for the positional cloning of a QTL.

3.3.2. The 'Omics' Approach: Linking Candidate Genes and QTLs for Tolerance to Abiotic Stress

The recent progress in the mass-scale profiling of the transcriptome, proteome and metabolome (see also Chapter 16 by Sreenivasulu *et al*., in this book) offers the possibility of investigating the concerted response of thousands of genes to abiotic stress. This approach receives support from the entire genome sequence and annotated genes that are available in rice (Goff *et al*., 2002). A number of studies have applied mass-scale profiling of mRNAs to investigate the changes in gene expression in response to environmental stress (Deyholos and Galbraith, 2001; Kawasaki *et al*., 2001; Chen *et al*., 2002; Ozturk *et al*., 2002; Zinselmeier *et al*., 2002; Maathuis *et al*., 2003; Oono *et al.*, 2003; Yu and Setter, 2003). An interesting application of transcriptome analysis is the identification of the so-called eQTLs, namely QTLs that influence the level of expression (hence the "e") of a particular gene. In this case, the analysis of the level of gene expression carried out on each progeny of a mapping population will identify eQTLs influencing the observed variability among progenies in mRNA level of the profiled genes. Circumstantial evidence as to the importance of each open reading frame (ORF) in governing variability for yield under conditions of abiotic stress can be obtained by comparing the map position of QTLs for yield with the map position of the ORF itself and the corresponding eQTLs. In plants, eQTLs have been reported for the first time in maize (Schadt *et al*., 2003). Clearly, the costs for profiling a large number of RNA samples required to identify eQTLs within a mapping population are still prohibitive to conceive the routine application of this approach. Instead, transcriptome profiling is better suited for studies involving a limited number of samples differing at key genomic regions, such as in the case for congenic strains (e.g., NILs, BDLs, ILLs, etc.) as well as samples derived from bulk segregant analysis. Besides providing clues for the identification of candidate genes, transcriptome profiling of congenic QTL strains offers the opportunity of enriching the region with additional markers, provided that the different levels of expression of the investigated ORFs are caused by a sequence polymorphism (e.g., SNP or insertion/deletion) that can be utilized for producing a useful marker. SNP markers are particularly interesting because their profiling can be automated.

Deciphering gene function can also be facilitated by information gathered through the profiling of the proteome and metabolome, which, as compared to the transcriptome, are functionally "closer" to the observed phenotypic traits used to select for tolerance to abiotic stress. Recent work in yeast and mammalian cells has shown that abiotic stress can profoundly alter the translational machinery (Ideker *et al*., 2001; Patel *et al.,* 2002; Uesono and

Toh, 2002). The rapid improvement and automation in the techniques required to quantify proteins allows for the quantification of up to ca. 2,000-2,500 proteins in a single sample (Yates, 1998). In rice, more than 2,000 proteins were detected reproducibly in drought-stressed and well-watered leaves (Salekdeh *et al*., 2002). Among the 1,000 proteins that were reliably quantified, 42 changed significantly in abundance and/or position. The three most marked changes were identified for the actin depolymerizing factor, the chloroplastic glutathione-dependent dehydroascorbate reductase and a homologue of the S-like ribonucleases. Profiling the proteome of a mapping population offers the opportunity to identify QTLs influencing protein quantity (PQLs, Protein Quantity Loci; Damerval *et al*., 1994; Touzet *et al*., 1995; de Vienne *et al*., 1999; Thiellement *et al*., 1999; Zivy and de Vienne, 2000; Pflieger *et al.,* 2001; Consoli *et al*., 2002). Co-localization of a PQL with its protein-coding locus would indicate that allelic differences at that locus influence the expression of the protein, whereas co-localization between a PQL and a QTL for a different trait would suggest an association between the candidate protein and trait variation. Interesting examples of the application of proteomics in maize in order to identify suitable candidate genes for QTLs governing resistance to drought have been provided by the work of de Vienne *et al*. (1999) and Pelleschi *et al*. (1999). In maize grown under water-limiting conditions it has been suggested that the observed reduction in acid invertase activity could impair sink strength because photosynthates cannot be converted rapidly to starch (Zinselmeier *et al*., 1995; Edmeades *et al*., 2000).

Metabolome profiling is aimed at the identification and quantification of all metabolites in a given biological sample (Fell, 2001; Fiehn *et al.,* 2001). Metabolic databases provide a valuable framework to predict biochemical pathways and products given a certain genotype and to reveal the phenotype of specific mutations (Fell, 2001), thus contributing to a more comprehensive view of the functional characteristics under investigation. With the present technology, ca. 2,000 different metabolites can be profiled in a single sample (Fiehn, 2002). When applied to a mapping population, metabolome profiling can be used to identify QTLs regulating the level of a particular metabolite and verify its coincidence with QTLs for yield and/or genes involved in metabolic pathways.

In maize, the changes occurring during a drought episode in the level of a limited number of key metabolites such as sugars and starch in the reproductive organs and in the growing kernel have been investigated (Zinselmeier *et al.,* 1995; Zinselmeier *et al*., 1999). QTLs for invertase activity have been described in a maize population subjected to drought stress (Pelleschi *et al*., 1999). The number of QTLs for invertase activity detected

under drought (nine in total) was more than twice the number detected under well-watered conditions (four in total), an indirect indication of the important role of this enzyme under drought conditions. One QTL common to both treatments was located near *Ivr2*, an invertase-encoding gene on bin 5.03. A number of QTLs influencing invertase activity were found to map in close proximity to carbohydrate QTLs with the two main clusters mapping on bins 1.03 and 5.03. Drought produced an early stimulation of acid-soluble invertase activity in adult leaves, whereas the activity of the cell wall invertase was found to be unaffected. This response was closely related to the mRNA level for only one (*Ivr2*) of the invertase genes.

Since a cause-effect relationship often exists between the activity of enzymes involved in N/C metabolism and N accumulation in plants, QTL studies provide an opportunity to test the validity of the candidate gene approach to identify genes underlining QTLs for NUE. In maize, a number of genes that encode N- and C-metabolic enzymes have been mapped near the peak of QTLs for vegetative development and for grain yield and its components (Hirel *et al.,* 2001). Examples of such genes were glutamine synthetase (glutamate-ammonia ligase), ADPGppase, sucrose phosphate synthase, sucrose synthase and invertase (β-fructofuranosidase). These genes were suggested as candidates for a pleiotropic action on some of the reported QTLs, particularly when all available information is consistent with the expected physiological effects of such genes. The most striking coincidence was for a gene (*gln4)* encoding glutamine synthetase, which maps in bin 5.07. It is also worth mentioning that one of the QTLs evident at both N levels can be assigned to bin 1.06, a region influencing grain yield under drought (Tuberosa *et al.,* 2002c), a condition likely to limit nutrient assimilation as well. More recently, an example of the validity of the candidate gene approach has been provided in a wheat study (Vagujfalvi *et al*., 2003), whose details have been reviewed earlier in the section 3.2.6.4. More compelling evidence on the role of a particular ORF can then be obtained by altering (down- or up-regulating) its expression by genetic engineering, or by the analysis of mutants (e.g., knockout mutants, TILLING, etc.) and/or through linkage-disequilibrium (LD) analysis of the target region carried out in a sufficiently large sample of unrelated genotypes (Buckler and Thornsberry, 2002).

Although only a very limited number of candidate genes for QTLs have been identified through the deployment of the different 'omics' platforms, their application will streamline this process, especially when it is possible to establish a plausible cause-effect relationship of the target trait with the expression level and/or the product of the putative candidate gene under investigation.

3.4. Marker-Assisted Selection for Enhancing Tolerance to Abiotic Stress

Due to the complex genetic basis of yield, improvement and stabilizing crop performance under conditions of abiotic stress deploying conventional breeding methodologies remains a slow, laborious and largely empirical process. Progress in this direction through a direct selection has been hampered by the low heritability of yield, particularly under drought, and by its large G × E interaction (Blum, 1988; Ceccarelli and Grando, 1996). As an alternative to a direct selection for yield under drought conditions, morpho-physiological traits genetically correlated with yield have been targeted (Blum, 1988; Bolaños and Edmeades, 1996; Turner, 1997). The successful application of this strategy requires traits, which are easy and cheap to score, ideally at a rather early growth stage and preferably before flowering, and characterized by high heritability as well as a high genetic correlation with yield. Because the genetic basis of morpho-physiological traits influencing abiotic stress tolerance is prevalently quantitative, hence not highly heritable, and their measurement is often laborious, to date only a handful of successful examples are available for an indirect phenotypic selection leading to increased tolerance to stress (Richards, 1996; Bolaños and Edmeades, 1996). When the observation of the phenotype alone does not guarantee the identification of the desired genotype, marker-aided selection (MAS) offers the opportunity for improving the efficiency of selection. Streamlining MAS procedures requires quick DNA extraction protocols coupled with high-throughput genotyping, preferably based on the scoring of markers, which do not require the use of gels (Salvi *et al*., 2001). In principle, once QTLs have been identified, introgression of the favourable alleles and their pyramiding into elite germplasm (e.g., parental lines, populations, etc.) becomes possible through MAS (Ribaut and Hoisington, 1998; Ribaut and Betran, 2000).

Although MAS has been extensively deployed for improving tolerance to biotic stresses (Young, 1999; Witcombe and Hash, 2000) and/or to accelerate the recovery of the recurrent parent in backcrossing procedures (Ribaut *et al.,* 1997a; Ribaut *et al.,* 2002), to date only few applications of MAS have been described for the improvement of quantitative traits, including abiotic stress

tolerance (see 3.4 in this article). Two reasons for this discrepancy are the difficulty in identifying major QTLs with a consistently sizeable effect on final yield and the high cost of MAS, a primary concern for its succesful implementation (Stuber *et al*., 1999).

Efforts based on the application of MAS to improve drought resistance are in progress at CIMMYT for maize (Ribaut *et al*., 1997a; Ribaut *et al*., 1997b; Ribaut *et al*., 1997c; Ribaut *et al*., 2002), at ICRISAT for pearl millet and sorghum (Crouch and Serraj, 2002; Yadav *et al*., 2002; Hash *et al.*, 2003) and at IRRI for rice (Atlin and Lafitte, 2002). In maize, the availability of molecular markers linked to a number of QTLs influencing ASI offers the opportunity to select for ASI, hence drought tolerance, also when drought does not occur at flowering (Ribaut *et al*., 1997c). At CIMMYT, a backcross marker-assisted selection (BC-MAS) project has targeted five QTLs influencing ASI (Ribaut *et al*., 1996; Ribaut *et al*., 1997c; Ribaut *et al*., 2002). The line CML247 was used as the recurrent parent and the line Ac7643 was used as the drought-tolerant donor characterized by a short ASI. CML247, an elite line with high yield *per se* under well-watered conditions, is drought susceptible and shows long ASI under drought. The QTL regions with alleles for short ASI were transferred through MAS from Ac7643 into CML247. Testcrosses between two testers and ca. 70 lines derived through BC-MAS were evaluated for three years under different water regimes. Under severe stress conditions reducing yield by at least 80%, the mean of the selected lines outyielded the unselected control. However, this advantage decreased at a lower stress intensity and disappeared when drought reduced yield less than 40%. Across the water-limited trials, a few genotypes consistently outperformed the controls. It was encouraging to observe that under well-watered conditions, the selected lines were as productive as the control lines. Notwithstanding the encouraging results of this BC-MAS approach, Ribaut *et al.* (2002) cautioned that MAS is unable to predict the phenotype of any particular genotype based on its allelic composition, a constraint particularly relevant when the target trait is influenced by epistatic interactions. Another major limitation to a more widespread applicatiion of MAS pertains to the high costs associated to QTL discovery and validation, as well as the release of superior lines. To overcome these limitations, novel strategies have been implemented at CIMMYT that are aimed at improving both the cost-effectiveness of MAS and delivering new germplasm instead of improved versions of existing lines which, as compared to the former, are of more limited value (Ribaut and Betran, 2000). These strategies rely on (i) the construction of a consensus map that combines information related to QTL characterization and gene expression and on (ii) the identification, through functional genomics approaches of a set of key genes/pathways involved in

maize drought response that will be used as selection tools in breeding programmes (Ribaut *et al.,* 2002).

Owing to the key role played by roots in determining rice yield under rainfed conditions, MAS for root depth is being deployed at IRRI to more specifically tailor root morphology to the range of environments present in the different rice-growing regions (Mackill *et al*., 1999). A deep, thick root system has been shown to favourably affect yield of upland rice under water-limited conditions (Nguyen *et al*., 1997). Different mapping populations have been analyzed in an attempt to identify QTLs governing the variability observed for root traits (Yadav *et al*., 1997; Mackill *et al*., 1999; Courtois *et al*., 2000; Price *et al*., 2000; Kamoshita *et al*., 2002; Lafitte *et al*., 2002; Courtois *et al*.,2003). This extensive body of data has allowed a detailed comparison of the QTL position for root traits across populations, leading to the identification of a number of QTL regions more consistently affecting variability in root morphology (Courtois *et al*., 2000; Price *et al*., 2000; Zhang *et al*., 2000; Courtois *et al*., 2003;). In rice, MAS has also been successfully applied to improve submergence tolerance of KDML105, a valuable jasmine rice cultivar widely grown in rain-fed lowland areas of Thailand (Siangliw *et al.,* 2003). To this end, three submergence-tolerant cultivars (FR13A, IR49830-7-1-2-2 and IR67819F2-CA-61) were crossed with KDML105. The transfer to the recipient KDML105 of a major QTL allele improving submergence tolerance was achieved by four backcrosses. Plants of a BC_4F_3 line that retained a critical region on chromosome 9 transferred from the tolerant lines were also tolerant to complete submergence, while preserving all the agronomically desirable characteristics and aromatic properties of KDML105. This study clearly indicates the feasibility of deploying MAS in breeding programmes for improving the level of submergence tolerance in lowland rice.

From an experimental standpoint, an interesting application of MAS is the derivation of congenic stocks (e.g., NILs, BDLs, ILLs, etc.) which differ only for the alleles present at the target QTL region, whose size ranges from ca. 5 to 30 cM. This approach allows us to "Mendelize" QTLs, an essential prerequisite to more accurately test the effects of the targeted QTLs and to undertake their positional cloning (Salvi *et al*., 2002; Salvi *et al.,* 2003). A number of NILs have already been obtained for QTLs of traits influencing drought tolerance in rice (Courtois *et al*., 2000; Shen *et al.,* 2001; Price *et al*., 2002a), sorghum (Tuinstra *et al.,* 1998) and maize (Landi *et al*., 2002a).

To what extent MAS may become a more routine practice will largely depend on the associated costs. In a recent study, Morris *et al*. (2003) have critically analysed and compared the cost-effectiveness of conventional and MAS

breeding for the introgression of an elite allele at a single dominant gene into an elite maize line. Neither method showed clear superiority: while conventional breeding schemes were found to be less expensive, MAS-based schemes were indeed faster. A more detailed analysis of the benefits and shortcomings deriving from the use of MAS is provided by Koebner in Chapter 10 of this book.

4. SUMMARY AND OUTLOOK

The improvements in crops' yield achieved during the past decades are now limited by water availability (Passioura, 2002), which is considered to be the most challenging environmental stress. The release of cultivars characterized by increased tolerance to drought and other abiotic stresses will largely depend on our ability to tailor crops' genomes accordingly. Even though some of its aspects may remain unfamiliar, the genomic approach, when appropriately intersected with other relevant disciplines, will positively impact our understanding of plant metabolism and, eventually, our breeding activities (Miflin, 2000; Tuberosa *et al.,* 2002a; Morandini and Salamini, 2003).

Genomics interfaced with crop modelling approaches provides a means to help dissecting G × E interactions and to resolve the hierarchical complexity underlying yield into traits that might be under simpler genetic control (Asseng *et al.,* 2002; Asseng *et al*., 2003; Reymond *et al*., 2003; Tardieu, 2003). Essentially, models are constructed as decision-making tools for management but may be of use in detecting prospective traits for selection within a breeding programme. Using sound information on crop physiology and empirical relationships, these models can simulate crops' performance, including G × E interactions (Slafer, 2003). Specific models integrating different level of complexity to predict biomass and yield as a function of environmental variables are available (Asseng *et al*., 2002; Asseng *et al*., 2003). Such models can integrate information on the genetic basis of yield, which is included in the form of genetic coefficients. In future, it may be possible to incorporate in these models information concerning the action of single QTLs (Reymond *et al*., 2003; Tardieu, 2003) in order to optimize MAS programmes. Further opportunities for collaboration between modellers and geneticists in ideotype breeding for high crop yield have recently been illustrated (Yin *et al*., 2003).

As to the traits to be dissected through QTL analysis and other molecular approaches in programmes aimed at improving drought tolerance,

photosynthetic efficiency (for an example in maize, see Jeanneau *et al*., 2003) and root architecture are likely to receive greater attention. These traits cannot be easily manipulated through conventional approaches. Other traits may also provide meaningful contributions, should it be possible to devise appropriate screening techniques and if sources of genetic variability suitable for QTL discovery and cloning are available. One example is offered by mycorrhizal colonization, a complex trait whose manipulation may influence water use and nutrient uptake of crops. Although QTLs for mycorrhizal responsiveness have been reported in maize (Kaeppler *et al.,* 2000), limited information is available on the genetic control of the interaction between mycorrhiza and crops' roots (Barker *et al*., 2002).

On the molecular side, extensive EST databases and unigene sets derived from cDNA libraries of different tissues and organs, involved in stress tolerance, are excellent sources for building functional maps that in some cases could help in the identification of plausible candidates for QTLs (Bohnert and Cushman, 2000; Ishimaru *et al*., 2001c). The release of the entire genome sequences of Arabidopsis (Bevan *et al*., 2001) and rice (Goff *et al.,* 2002) has greatly increased the level of resolution of comparative mapping studies exploiting syntenic relationships. The application of this approach to rice and related cereals has revealed a more complex picture (Goff *et al.,* 2002) than that previously reported through the coarser comparative analysis based on the map position of common RFLP markers (Gale and Devos, 1998). However, accumulating evidence for the abundance of chromosomal rearrangements and the recent demonstration on the lack of microcollinearity among cereals (Sorrells *et al.,* 2003) may discourage the map-based cloning approach for isolation of loci controlling important quantitative traits. From a strictly organizational standpoint, the degree of genome rearrangements that have occurred in different species after diverging from a common ancestor proportionally reduce the possibility for the successful exploitation of synteny among genomes of distantly related species, as is the case with Arabidopsis and cereals (Paterson *et al.,* 1996; Gale and Devos, 1998; Van Buuren *et al.,* 2002; Bowers *et al.,* 2003). Thus, in practical terms, relying extensively on a phylogenetically distant species as a model (e.g., Arabidopsis as a model for cereals) may have, on a case-by-case basis, its own shortcomings both from a genomic and physiological standpoint. The future availability of sequence information for whole genomes will greatly enhance the accuracy of gene/QTL discovery based on LD approaches. When sequencing information is unavailable, haplotype analysis of a set of populations can also provide useful information for QTL discovery (Jansen *et al*., 2003). Physical maps of contiged genomic clones (e.g., BACs) will allow for the mapping of any sequence, even if no polymorphism is available, thus streamlining the process of building highly

saturated maps, an essential prerequisite to positional cloning of QTLs. More importantly, polymorphisms at ORFs will allow us to identify sequence variation with functional relevance at the phenotypic level and to validate candidate genes for tolerance to abiotic stress. Assigning a function to genes and validating their roles will be facilitated by a number of experimental approaches, such as RNA-mediated interference (Cogoni and Macino, 2000), forward and reverse mutagenesis (Maggio *et al.,* 2001), including homologous recombination (Hanin and Paszkowski, 2003) and TILLING (Targeting Induced Local Lesions IN Genomes; Till *et al*., 2003).

Model species other than rice may also play an increasingly important role for the discovery of genes/QTLs that may help in improving tolerance of cereals to abiotic stresses. This may be particularly true with model species like Arabidopsis*,* which, as compared to cereals, is comparatively easy to phenotype in large numbers. The small size and very short life cycle, the small genome, the availability of the annotated genome sequence, and a vast number of mutants coupled with a very efficient transformation system, all contribute to make Arabidopsis a particularly attractive species for the identification of QTLs and genes regulating the response to abiotic stress. The work carried out in Arabidopsis will be of particular value in elucidating the molecular mechanisms involved in the perception of abiotic stress as well as in the signal transduction pathway leading to the activation of the suite of genes involved in the adaptive response to stress. Surprisingly, so far only limited work has been carried out in Arabidopsis to identify QTLs for tolerance to abiotic stresses. Recently, the discovery of a number of genes (e.g., *DREB*, Jaglo-Ottosen *et al.*, 1998; Kasuga *et al.*, 1999) quickly triggered by abiotic stresses and encoding for transcription factors controlling the expression of drought-responsive genes has sparked great interest for investigating their role in crops. *DREB* homologues have been identified in maize (van Buuren *et al.*, 2002) and wheat (Vagujfalvi *et al*., 2003), where a *DREB* gene has been shown to be a strong candidate for a QTL conferring freezing tolerance.

Future efforts to improve tolerance to abiotic stress should emphasize allele mining in a germplasm context broader than that presently explored, which has been mainly limited to elite materials. The exploitation of a wider genetic basis will provide additional opportunities to identify rare alleles capable of improving crop adaptation to harsh environments and increasing their sustainability. In this connection, the utilization of the advanced-backcross QTL analysis (ABQA) approach will be instrumental for identifying valuable QTL alleles in wild accessions and for introgressing such alleles in a reasonably short time into elite germplam Additionally, progenies obtained from subsequent intercrossings of multiparental mating schemes including highly contrasted genotypes will improve our capacity to identify and resolve

QTLs. As compared to the biparental crosses that are traditionally deployed for QTL discovery, multiparental mating schemes explore a wider allelic diversity and a larger number of unique recombination events (Li *et al.,* 2001; Korstanje and Paigen; Mott and Flint, 2002), thus providing increased opportunities to unveil at a much finer genetic resolution the presence of a QTL. Although setting up and implementing such multiparental crossing schemes is a costly undertaking, the benefits expected from this approach certainly make this a worthy investment, particularly in a long-term perspective.

5. CONCLUDING REMARKS

Although it is often mentioned that we have entered the "post-genomic" era, a daunting amount of work remains to be completed before we can achieve a thorough comprehension of the complex interactions governing crops' adaptation to abiotic stresses. The next years will witness increased experimental activities aimed at assigning functions to genes and to devise strategies on how best to deploy this information for tailoring crops' genotypes that are able to better withstand environmental constraints. Although this challenging task will require several decades to achieve reasonable success, the benefits to be harnessed will largely outweigh the efforts. During the past 20 years the rate of population growth outpaced the rate of increase in food-grain production, thus indicating that unless this trend is offset by the release of cultivars with a higher yield potential and stability, food shortages will occur in the present century (Khush, 1999). The remarkable increase in cereals' yield during the past century has been equally attributed to breeding and better agronomic practices. A higher resistance to environmental stress, coupled with an improved ability to maximize yield under low-stress conditions have both contributed to such spectacular gains and to stabilize yield across fluctuating environmental conditions (Duvick *et al*., 1999; Miflin, 2000). To maintain the previous linear gains in cereals' yield and to counteract the negative effects due to the increased unpredictability of weather patterns, depletion of irrigation water, soil salinization and increased cost of fertilizers will require an unprecedented effort among different categories of researchers. The real challenge faced by plant scientists is how to fruitfully exploit and effectively integrate into existing breeding programmes the materials and the information generated through QTL analysis and other genomic platforms. This integration will further our knowledge of the physiologic and genetic basis of tolerance to abiotic stresses and will expedite the release of improved varieties better suited for a more sustainable farming and for adequately feeding the burgeoning population of our planet.

6. REFERENCES

Acevedo E. Potential of carbon isotope discrimination as a selection criterion in barley breeding. In: J.R. Ehleringer, A.E. Hall, G.D. Farquhar (eds.) *Stable isotopes and plant carbon-water relations*, Academic Press, New York, 1993, pp. 399-416

Agrama H.A.S., Moussa M.E. Mapping QTLs in breeding for drought tolerance in maize (*Zea mays* L.). Euphytica 1996; 91: 89-97

Agrama H.A.S., Zakaria A.G., Said F.B., Tuinstra M. Identification of quantitative trait loci for nitrogen use efficiency in maize. Mol Breed 1999; 5: 187-195

Ali M.L., Pathan M.S., Zhang J., Bai G., Sarkarung S., Nguyen H.T. Mapping QTLs for root traits in a recombinant inbred population from two *indica* ecotypes in rice. Theor Appl Genet 2000; 101: 756-766

Amtmann A., Sanders D. Mechanisms of Na^+ uptake by plant cells. Adv Bot Res 1999; 29: 75-112

Andaya V.C., Mackill D.J. QTLs conferring cold tolerance at the booting stage of rice using recombinant inbred lines from a *japonica* x *indica* cross. Theor Appl Genet 2003; 106: 1084-1090

Apse M.P., Aharon G.S., Snedden W.A., Blumwald E. Salt tolerance conferred by overexpression of a vacuolar Na^+/H^+ antiport in Arabidopsis. Science 1999; 285: 1256-1258

Araus J.L., Slafer G.A., Reynolds M.P., Royo C. Plant breeding and drought in C_3 cereals: what should we breed for? Ann Bot 2002; 89: 925-940

Ashraf M. Salt tolerance of cotton: some new advances. Crit Rev Plant Sci 2002; 21: 1-30

Asins M.J. Present and future of quantitative trait locus analysis in plant breeding. Plant Breed 2002; 121: 281-291

Asseng S., Turner N.C., Ray J.D., Keating B.A. A simulation analysis that predicts the influence of physiological traits on the potential yield of wheat. Eur Jour Agron 2002; 17: 123-141

Asseng S., Turner N.C., Botwright Y., Condon A.G. Evaluating the impact of a trait for increased specific leaf area on wheat yields using a crop simulation model. Agron Jour 2003; 95: 10-19

Atlin G.N., Lafitte H.R. Marker-assisted breeding versus direct selection for drought tolerance in rice. In: N.P. Saxena, J.C. O'Toole (eds.) *Field screening for drought tolerance in crop plants with emphasis on rice. Proceedings of an International Workshop on Field Screening for Drought Tolerance in Rice*, 2002, pp. 71-81

Babu R.C., Nguyen B.D., Chamarerk V., Shanmugasundaram P., Chezhian P., Jeyaprakash P., Ganesh S., Palchamy A., Sadasivam S., Sarkarung S.*, et al.* Genetic analysis of drought resistance in rice by molecular markers: Association between secondary traits and field performance. Crop Sci 2003; 43: 1457-1469

Barker S.J., Duplessis S., Tagu D. The application of genetic approaches for investigations of mycorrhizal symbioses. Plant and Soil 2002; 244: 85-95

Barrière Y., Gibelin C., Argillier O., Méchin V. Genetic analysis in recombinant inbred lines of early dent forage maize. I - QTL mapping for yield, earliness, starch and crude protein contents from per se value and top cross experiments. Maydica 2001; 46: 253-266

Baum M., Grando S., Backes G., Jahoor A., Sabbagh A., Ceccarelli S. QTLs for agronomic traits in the Mediterranean environment identified in recombinant inbred lines of the cross 'Arta' x *H. spontaneum* 41-1. Theor Appl Genet 2003; 107: 1215-1225

Beavis W.D. QTL analysis: power, precision, and accuracy. In: A. Paterson (ed.) *Molecular dissection of complex traits*, CRC Press, Boca Raton, 1998, pp. 145-162

Ben-Haj-Salah H., Tardieu F. Temperature affects expansion rate of maize leaves without change in spatial distribution of cell length. Analysis of the coordination between cell division and cell expansion. Plant Physiol 1995; 109: 861-870

Ben-Haj-Salah H., Tardieu F. Quantitative analysis of the combined effects of temperature, evaporative demand and light on leaf elongation rate in well-watered field and laboratory-grown maize plants. Jour Exp Bot 1996; 47: 1689-1698

Bertin P., Gallais A. Genetic variation for nitrogen use efficiency in a set of recombinant inbred lines II - QTL detection and coincidences. Maydica 2001; 46: 53-68

Bevan M., Mayer K., White O., Eisen J.A., Preuss D., Bureau T., Salzberg S.L., Mewes H.W. Sequence and analysis of the Arabidopsis genome. Curr Opin Plant Biol 2001; 4: 105-110

Blum A. *Breeding for Stress Environments*, CRC Press, Boca Raton, 1988

Blum A. Crop responses to drought and the interpretation of adaptation. Plant Growth Regul 1996; 20: 135-148

Blum A., Munns R., Passioura J.B., Turner N.C. Genetically engineered plants resistant to soil drying and salt stress: How to interpret osmotic relations? Plant Physiol 1996; 110: 1051-1053

Bohnert H.J., Cushman J.C. The ice plant cometh: Lessons in abiotic stress tolerance. Plant Growth Regul 2000; 19: 334-346

Bolaños J., Edmeades G.O. Eight cycles of selection for drought tolerance in lowland tropical maize. I. Responses in grain yield, biomass, and radiation utilization. Fields Crop Res 1993; 31: 233-252

Bolaños J., Edmeades G.O. The importance of the anthesis-silking interval in breeding for drought tolerance in tropical maize. Fields Crop Res 1996; 48: 65-80

Bolaños J., Edmeades G.O. The importance of the anthesis-silking interval in breeding for drought tolerance in tropical maize. In: G.O. Edmeades, M. Bänziger, H.R. Mickelson, C.B. Peña Valdivia (eds.) *Developing Drought- and Low N- Tolerant Maize, Proceedings of a Symposium*, CIMMYT, 1997, pp. 355-368

Bolaños J., Edmeades G.O., Martinez L. Eight cycles of selection for drought tolerance in lowland tropical maize. III. Responses in drought adaptive physiological and morphological traits. Fields Crop Res 1993; 31: 269-286

Bowers J.E., Chapman B.A., Rong J.K., Paterson A.H. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature 2003; 422: 433-438

Boyer J.S. Plant productivity and the environment. Science 1982; 218: 443-448

Boyko E., Kalendar R., Korzun V., Fellers J., Korol A., Schulman A.H., Gill B.S. A high-density cytogenetic map of the *Aegilops tauschii* genome incorporating retrotransposons and defense-related genes: insights into cereal chromosome structure and function. Plant Mol Biol 2002; 48: 767-790

Broman K.W., Wu H., Sen S., Churchill G.A. QTL mapping in experimental crosses. Bioinformatics 2003; 19: 889-890

Bruce W.B., Edmeades G.O., Barker T.C. Molecular and physiological approaches to maize improvement for drought tolerance. Jour Exp Bot 2002; 53: 13-25

Buckler E.S.I., Thornsberry J.M. Plant molecular diversity and applications to genomics. Curr Opin Plant Biol 2002; 5: 107-111

Burgos M.S., Messmer M.M., Stamp P., Schmid J.E. Flooding tolerance of spelt (*Triticum spelta* L.) compared to wheat (*Triticum aestivum* L.) - a physiological and genetic approach. Euphytica 2001; 122: 287-295

Cattivelli L., Baldi P., Crosatti C., Di Fonzo N., Faccioli P., Grossi M., Mastrangelo A.M., Pecchioni N., Stanca A.M. Chromosome regions and stress-related sequences involved in resistance to abiotic stress in *Triticeae*. Plant Mol Biol 2002a; 48: 649-665

Cattivelli L., Baldi P., Crosatti C., Grossi M., Valè G., Stanca A. Genetic bases of barley physiological response to stressful conditions. In: G. Slafer, J.L. Molina-Cano, R. Savin, J.L. Araus, I. Romagosa (eds.) *Barley Science-Recent advances from molecular biology to agronomy of yield and quality*, Haworth Press Inc., New

York, 2002b, pp. 307-360
Ceccarelli S. Plant responses to water stress: a review. Genetica Agraria 1984; 38: 43-73
Ceccarelli S., Grando S. Drought as a challenge for the plant breeder. Plant Growth Regul 1996; 20: 149-155
Champoux M.C., Wang G., Sarkarung S., Mackill D.J., O'Toole J.C., Huang N., McCouch S.R. Locating genes associated with root morphology and drought avoidance in rice via linkage to molecular markers. Theor Appl Genet 1995; 90: 969-981
Chapman S.C., Edmeades G.O. Selection improves drought tolerance in tropical maize populations: II Direct and correlated responses among secondary traits. Crop Sci 1999; 39: 1315-1324
Chen W., Provart N.J., Glazebrook J., Katagiri F., Chang H.S., Eulgem T., Mauch F., Luan S., Zou G., Whitham S.A., *et al.* Expression profile matrix of Arabidopsis transcription factor genes suggests their putative functions in response to environmental stresses. Plant Cell 2002; 14: 559-574
Choi D.W., Koag M.C., Close T.J. Map locations of barley *Dhn* genes determined by gene-specific PCR. Theor Appl Genet 2000; 101: 350-354
Churchill G.A., Doerge R.W. Empirical threshold values for quantitative trait mapping. Genetics 1994; 138: 963-971
Close T.J. Dehydrins: a commonality in the response of plants to dehydration and low temperature. Physiol Plant 1997; 100: 291-296
Cogoni C., Macino G. Post-transcriptional gene silencing across kingdoms. Curr Opin Genet Dev 2000; 10: 638-643
Condon A.G., Richards R.A., Farquhar, G.D. Relationships between carbon isotope discrimination, water use efficiency and transpiration efficiency for dryland wheat. Aust Jour Agric Res 1993; 44: 1693-1711
Condon A., Richards R., Rebetzke G., Farquhar G. Improving intrinsic water-use efficiency and crop yield. Crop Sci 2002; 42: 122-131
Consoli L., Lefévre C.L., Zivy M., de Vienne D., Damerval C. QTL analysis of proteome and transcriptome variations for dissecting the genetic architecture of complex traits in maize. Plant Mol Biol 2002; 48: 575-581
Courtois B., McLaren G., Sinha P.K., Prasad K., Yadav R., Shen L. Mapping QTLs associated with drought avoidance in upland rice. Mol Breed 2000; 6: 55-66
Courtois B., Shen L., Petalcorin W., Carandang S., Mauleon R., Li Z. Locating QTLs controlling constitutive root traits in the rice population IAC 165 x Co39. Euphytica 2003; 134: 335-345
Crasta O.R., Xu W.W., Rosenow D.T., Mullet J., Nguyen H.T. Mapping of post flowering drought resistance traits in grain sorghum: association between QTLs influencing premature senescence and maturity. Mol Gen Genet 1999; 262: 579-588
Crouch J.H., Serraj R. DNA marker technology as a tool for genetic enhancement of drought tolerance at ICRISAT. In: N.P. Saxena, J.C. O'Toole (eds.) *Field Screening for Drought Tolerance in Crop Plants with Emphasis on Rice. Proceed Intern Workshop on Field Screening for Drought Tolerance in Rice*, 2002, pp. 155-170
Cushman J.C., Bohnert H.J. Genomic approaches to plant stress tolerance. Curr Opin Plant Biol 2000; 3: 117-124
Damerval C., Maurice A., Josse J.M., de Vienne D. Quantitative trait loci underlying gene product variation: a novel perspective for analyzing regulation of genome expression. Genetics 1994; 137: 289-301
Dashti H., Yazdi-Samadi B., Abd-Mishani C., Ghannadha M.R. QTL analysis on cold resistance and heading date in wheat, using doubled haploid lines. Iran Jour Agric Sci 2001; 32: 157-177
Davis G.L., McMullen M.D., Baysdorfer C., Musket T., Grant D., Staebell M., Xu G., Polacco M., Koster L., Melia-Hancock S., *et al.* A maize map standard with

sequenced core markers, grass genome reference points and 932 expressed sequence tagged sites (ESTs) in a 1736 locus map. Genetics 1999; 152: 1137-1172

de Vienne D., Leonardi A., Damerval C., Zivy M. Genetics of proteome variation for QTL characterization: application to drought-stress responses in maize. Jour Exp Bot 1999; 50: 303-309

Deyholos M., Galbraith D. High-density microarrays for gene expression analysis. Cytometry 2001; 43: 229-238

Duvick D.N., Coors J.G., Pandey S. Heterosis: feeding people and protecting natural resources. In: J.G. Coors (ed.) *The genetics and exploitation of heterosis in crops. Proceedings of an international symposium*, CIMMYT, 1999, pp. 19-29

Edmeades G.O., Bolanõs J., Elings A., Ribaut J.M., Bänziger M., Westgate M.E. The role and regulation of the anthesis-silking interval in maize. In: M.E. Westgate, K.J. Boote (eds.) *CSSA Special Publication No 29*, Madison, 2000, pp. 43

Ehdaie B., Waines J.G. Genetic analysis of carbon isotope discrimination and agronomic characters in a bread wheat cross. Theor Appl Genet 1994; 88: 1023-1028

Ellis R.P., Forster B.P., Robinson D., Handley L.L., Gordon D.C., Russell J.R., Powell W. Wild barley: a source of genes for crop improvement in the 21st century? Jour Exp Bot 2000; 51: 9-17

Ellis R.P., Forster B.P., Gordon D.C., Handley L.L., Keith R.P., Lawrence P., Meyer R., Powell W., Robinson D., Scrimgeour C.M., Young G.R., Thomas W.T.B. Phenotype/genotype associations for yield and salt tolerance in a mapping population segregating for two dwarfing genes. Jour Exp Bot 2002; 53: 1-14

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

Fan P., Tao Q.N., Wu P. Quantitative trait loci (QTL) underlying N-absorbing capacity and N-utilization efficiency in paddy rice and their genetic effects. Plant Nutr Fert Sci 2001; 7: 159-165

Farquhar G.D., Ehleringer J.R., Hubick K.T. Carbon isotope discrimination and photosynthesis. Annu Rev Plant Physiol Plant Mol Biol 1989; 4: 503-537

Fell D.A. Beyond genomics. Trends Genet 2001; 17: 680-682

Fiehn O. Metabolomics - the link between genotypes and phenotypes. Plant Mol Biol 2002; 48: 155-171

Fiehn O., Kloska S., Altmann T. Integrated studies on plant biology using multiparallel techniques. Curr Opin Biotech 2001; 12: 82-86

Flowers T.J., Yeo A.R. Breeding for salinity tolerance in crop plants: where next? Aust Jour of Plant Physiol 1995; 22: 875-884

Foolad M.R. Recent advances in genetics of salt tolerance in tomato. Plant, Cell, Tissue, Organ Cult 2004; 72: 101-119

Forster B.P., Ellis R.P., Thomas W.T.B., Newton A.C., Tuberosa R., This D., El Enein R.A., Bahri M.H., Ben Salem M. The development and application of molecular markers for abiotic stress tolerance in barley. Jour Exp Bot 2000; 51: 19-27

Frova C., Sari-Gorla M. Quantitative expression of maize HSPs: genetic dissection and association with thermotolerance. Theor Appl Genet 1993; 86: 213-220

Frova C., Sari-Gorla M. Quantitative trait loci (QTLs) for pollen thermotolerance detected in maize. Mol Gen Genet 1994; 245: 424-430

Frova C., Krajewski P., Di Fonzo N., Villa M., Sari Gorla M. Genetic analysis of drought tolerance in maize by molecular markers. I. Yield components. Theor Appl Genet 1999; 99: 280-288

Gahoonia T.S., Nielsen N.E. Variation in root hairs of barley cultivars doubled soil phosphorus uptake. Euphytica 1997; 98: 177-182

Gale M.D., Devos K.M. Plant comparative genetics after 10 years. Science 1998; 282: 656-659

Galiba G., Quarrie S.A., Sutka J., Morgounov A., Snape J.W. RFLP mapping of the vernalization (*Vrn1*) and frost resistance (*Fr1*) genes on chromosome 5A of wheat. Theor Appl Genet 1995; 90: 1174-1179

Glazier A.M., Nadeau J.H., Aitman T.J. Finding genes that underlie complex traits. Science 2002; 298: 2345-2349

Goff S.A.*et al.* A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science 2002; 296: 92-100

Gowda M., Venu R.C., Roopalakshmi K., Sreerekha M.V., Kulkarni R.S. Advances in rice breeding, genetics and genomics. Mol Breed 2003; 11: 337-352

Hackett C.A. Statistical methods for QTL mapping in cereals. Plant Mol Biol 2002; 48: 585-599

Hanin M., Paszkowski J. Plant genome modification by homologous recombination. Curr Opin Plant Biol 2003; 6: 157-162

Hash C.T., Schaffert R.E., Peacock J.M. Prospects for using conventional techniques and molecular biological tools to enhance performance of 'orphan' crop plants on soils low in available phosphorus. Plant and Soil 2002; 245: 135-146

Hash C.T., Folkertsma R.T., Ramu P., Reddy B.V.S., Mahalakshmi V., Sharma H.C., Rattunde H.F.W., Weltzien E.R., Haussmann B.I.G., Ferguson M.E., *et al.* Marker-assisted breeding across ICRISAT for terminal drought tolerance and resistance to shoot fly and striga in sorghum. In: *"In the Wake of the Double Helix: From the Green Revolution to the Gene Revolution"*, 2003, pp. 81 (available at www.doublehelix.too.it)

Haussmann B.I., Mahalakshmi V., Reddy B.V., Seetharama N., Hash C.T., Geiger H.H. QTL mapping of stay-green in two sorghum recombinant inbred populations. Theor Appl Genet 2002; 106: 133-142

Hayes P.M., Blake T., Chen T.H.H., Tragoonrung S., Chen F., Pan A., Liu B. Quantitative trait loci on barley (*Hordeum vulgare* L.) chromosome 7 associated with components of winterhardiness. Genome 1993; 36: 66-71

Hirel B., Bertin P., Quilleré I., Bourdoncle W., Attagnant C., Dellay C., Gouy A., Cadiou S., Retailliau C., Falque M., Gallais A. Towards a better understanding of the genetic and physiological basis for nitrogen use efficiency in maize. Plant Physiol 2001; 125: 1258-1270

Hu B., Wu P., Liao C.Y., Zhang W.P., Ni J.J. QTLs and epistasis underlying activity of acid phosphatase under phosphorus sufficient and deficient condition in rice (*Oryza sativa* L.). Plant and Soil 2001; 230: 99-105

Hund A., Fracheboud Y., Soldati A., Frascaroli E., Salvi S., Stamp P. Mapping QTL regulating morpho-physiological root and shoot traits in maize seedlings grown at low temperature. Theor Appl Genet 2004, submitted.

Ideker T., Thorsson V., Ranish J.A., Christmas R., Buhler J., Eng J.K., Bumgarner R., Goodlett D.R., Aebersold R., Hood L. Integrated genomic and proteomic analyzes of a systematically perturbed metabolic network. Science 2001; 292: 929-934

Innes P., Blackwell R.D., Quarrie S.A. Some effects of genetic variation in drought-induced abscisic acid accumulation on the yield and water use of spring wheat. Jour of Agri Sci 1984; 102: 341-351

Ishimaru K., Kobayashi N., Ono K., Yano M. Are contents of Rubisco, soluble protein and nitrogen in flag leaves of rice controlled by the same genetics? Jour Exp Bot 2001a; 52: 1827-1833

Ishimaru K., Shirota K., Higa M., Kawamitsu Y. Identification of quantitative trait loci for adaxial and abaxial stomatal frequencies in *Oryza sativa*. Plant Physiol Biochem 2001b; 39: 173-177

Ishimaru K., Yano M., Aoki N., Ono K., Hirose T., Lin S.Y., Monna L., Sasaki T., Ohsui R. Toward the mapping of physiological and agronomic characters on a rice function map: QTL analysis and comparison between QTLs and expressed sequence tags. Theor Appl Genet 2001c; 102: 793-800

Ito O., O' Toole J., Hardy B. Genetic improvement of rice for water-limited environments. In: *Workshop on Genetic Improvement of Rice for Water-Limited Environment*, 1999, pp.

Ivandic V., Hackett C.A., Zhang Z.J., Staub J.E., Nevo E., Thomas W.T.B., Forster B.P. Phenotypic responses of wild barley to experimentally imposed water stress. Jour Exp Bot 2000; 51: 2021-2029

Jackson M.B., Ram P.C. Physiological and molecular basis of susceptibility and tolerance of rice plants to complete submergence. Ann Botany 2003; 91: 227-241

Jaglo Ottosen K.R., Gilmour S.J., Zarka D.G., Schabenberger O., Thomashow M.F. Arabidopsis CBF1 overexpression induces *COR* genes and enhances freezing tolerance. Science 1998; 280: 104-106

Jansen R.C., Stam P. High resolution of quantitative traits into multiple loci via interval mapping. Genetics 1994; 136: 1447-1455

Jansen R.C., Jannick J.L., Beavis W.D. Mapping quantitative trait loci in plant breeding populations: use of parental haplotype sharing. Crop Sci 2003; 43: 829-834

Jeanneau M., Gerentesb D., Foueillassarc X., Zivy, M., Vidala J., Toppand A., Perez P. Improvement of drought tolerance in maize: towards the functional validation of the *Zm-Asr1* gene and increase of water-use efficiency by over-expressing C_4PEPC. Biochimie 2002; 84: 1127-1135

Jensen A.B., Busk P.K., Figueras M., Albà M.M., Peracchia G., Messeguer R., Goday A., Pagès M. Drought signal transduction in plants. Plant Growth Regul 1996; 20: 105-110

Jones R.J., Setter T.L. Hormonal regulation of early kernel development. In: M.E. Westgate, K.J. Boote (eds.) *Physiology and modelling kernel set in maize*, Crop Science Society of America, Madison, 2000, pp. 25-42

Kaeppler S.M., Parke J.L., Mueller S.M., Senior L., Stuber C., Tracy W.F. Variation among maize inbred lines and detection of quantitative trait loci for growth at low phosphorus and responsiveness to Arbuscular mycorrhizal fungi. Crop Sci 2000; 40: 358-364

Kamoshita A., Wade L.J., Ali M.L., Pathan M.S., Zhang J., Sarkarung S., Nguyen H.T. Mapping QTLs for root morphology of a rice population adapted to rainfed lowland conditions. Theor Appl Genet 2002; 104: 880-893

Kasuga M., Liu Q., Miura S., Yamaguchi-Shinozaki K., Shinozaki K. Improving plant drought, salt, and freezing tolerance by gene transfer of a single stress-inducible transcription factor. Nature Biotech 1999; 17: 287-291

Kawasaki S., Deyholos M., Borchert C., Brazille S., Kawai K., Galbraith D.W., Bohnert H.J. Temporal succession of salt stress responses in rice by microarray analysis. Plant Cell 2001; 12: 889-906

Kebede H., Subudhi P.K., Rosenow D.T., Nguyen H.T. Quantitative trait loci influencing drought tolerance in grain sorghum (*Sorghum bicolor* L. Moench). Theor Appl Genet 2001; 103: 266-276

Khush G.S. Green revolution: preparing for the 21st century. Genome 1999; 42: 646-655

Korstanje R., Paigen B. From QTL to gene: the harvest begins. Nature Genet 2002; 31: 235-236

Koumproglou R., Wilkes T.M., Townson P., Wang X.Y., Beynon J., Pooni H.S., Newbury H.J., Kearsey M.J. STAIRS: a new genetic resource for functional genomic studies of Arabidopsis. Plant Jour 2002; 31: 355-364

Korstanje R, Paigen B. From QTL to gene: the harvest begins. Nature Genetics 2002; 31: 235-236

Koyama M.L., Levesley A., Koebner R.M.D., TJ F., Yeo A.R. Quantitative trait loci for component physiological traits determining salt tolerance in rice. Plant Physiol 2001; 125: 406-422

Lafitte H.R., Courtois B., Arraudeau M. Genetic improvement of rice in aerobic systems: progress from yield to genes. Field Crop Res 2002; 75: 171-190

Landi P., Sanguineti M.C., Conti S., Tuberosa R. Direct and correlated responses to divergent selection for leaf abscisic acid concentration in two maize populations. Crop Sci 2001;

41: 335-344
Landi P., Salvi S., Sanguineti M.C., Stefanelli S., Tuberosa R. Development and preliminary evaluation of near-isogenic lines differing for a QTL which affects leaf ABA concentration. Maize Genet Coop Newslett 2002a; 76: 7-8
Landi P., Sanguineti M.C., Darrah L.L., Giuliani M.M., Salvi S., Tuberosa R. Detection of QTLs for vertical root pulling resistance in maize and overlaps with QTLs for root traits in hydroponics and for grain yield at different water regimes. Maydica 2002b; 47: 233-243
Lebreton C., Lazic-Jancic V., Steed A., Pekic S., Quarrie S.A. Identification of QTL for drought responses in maize and their use in testing causal relationships between traits. Jour Exp Bot 1995; 46: 853-865
Lee M. DNA markers and plant breeding programmes. Advances in Agronomy 1995; 55: 265-344
Li Z.K., Khush G.S., Brar D.S., Hardy B. QTL mapping in rice: a few critical considerations. In: G. Khush, D. Brar (eds.) *Fourth International Rice Genetics Symposium*, 2001, pp. 153-171
Lilley J.M., Ludlow M.M., McCouch S.R., O'Toole J.C. Locating QTL for osmotic adjustment and dehydration tolerance in rice. Jour Exp Bot 1996; 47: 1427-1436
Liu B.H. *Statistical Genomics: Linkage, Mapping and QTL Analysis*, CRC Press, Boca Raton, 1998
Ludlow M.M., Muchow R.C. A critical evaluation of traits for improving crop yields in water-limited environments. Adv Agron 1990; 43: 107-153
Ma J.F., Shen R.F., Zhao Z.Q., Wissuwa M., Takeuchi Y., Ebitani T., Yano M. Response of rice to Al stress and identification of quantitative trait loci for Al tolerance. Plant and Cell Physiol 2002; 43: 652-659
Maathuis F.J.M., Filatov V., Herzyk P., Krijger G.C., Axelsen K.B., Chen S.X., Green B.J., Li Y., Madagan K.L., Sanchez Fernandez R*., et al.* Transcriptome analysis of root transporters reveals participation of multiple gene families in the response to cation stress. Plant Jour 2003; 35: 675-692
Mackill D.J., Nguyen H.T., Zhang J.X. Use of molecular markers in plant improvement programs for rainfed lowland rice. Fields Crop Res 1999; 64: 177-185
Maggio A., Hasegawa P.M., Bressan R.A., Consiglio M.F., Joly R.J. Unravelling the functional relationship between root anatomy and stress tolerance. Aust Jour Plant Physiol 2001; 28: 999-1004
Mano Y., Takeda K. Mapping quantitative trait loci for salt tolerance at germination and at the seedling stage of barley (*Hordeum vulgare* L.). Euphytica 1997; 94: 263-272
Mei H.W., Luo L.J., Ying C.S., Wang Y.P., Yu X.Q., Guo L.B., Paterson A.H., Li Z.K. Gene actions of QTLs affecting several agronomic traits resolved in a recombinant inbred rice population and two testcross populations. Theor Appl Genet 2003; 107: 89-101
Menz M.A., Klein R.R., Mullet J.E., Obert J.A., Unruh N.C., Klein P.E. A high density genetic map of *Sorghum bicolor* (L.) Moench based on 2926 AFLP®, RFLP and SSR markers. Plant Mol Biol 2002; 48: 483-499
Messmer M.J., Lambert R.J., Hageman R.H. Classification of certain N traits as criteria for the identification of production maize genotypes. Crop Sci 1984; 24: 605-610
Miflin B. Crop improvement in the 21st century. Jour Exp Bot 2000; 51: 1-8
Miftahudin R., Scoles G.J., Gustafson J.P. AFLP markers tightly linked to the aluminum-tolerance gene *Alt3* in rye (*Secale cereale* L.). Theor Appl Genet 2002; 104: 626-631
Milborrow B.W. The pathway of biosynthesis of abscisic acid in vascular plants: a review of the present state of knowledge of ABA biosynthesis. Jour Exp Bot 2001; 52: 1145-1164
Milla M.A., Gustafson J.P. Genetic and physical characterization of chromosome 4DL in wheat. Genome 2001; 44: 883-892
Ming F., Zheng X.W., Mi G.H., Zhu L.H., Zhang F.S. Detection and verification of quantitative trait loci affecting tolerance to low phosphorus in rice. Jour Plant Nutr

2001; 24: 1399-1408
Misawa S., Mori N., Takumi S., Yoshida S., Nakamura C. Mapping of QTLs for low temperature response in seedlings of rice (*Oryza sativa* L.). Cereal Res Commun 2000; 28: 33-40
Mitra J. Genetics and genetic improvement of drought resistance in crop plants. Curr Sci 2001; 80: 758-763
Moncada P., Martinez C., Borrero J.M.C., Gauch H.J., Guimaraes E., Tohme J., McCouch S. Quantitative trait loci for yield and yield components in an *Oryza sativa* × *Oryza rufipogon* BC_2F_2 population evaluated in an upland environment. Theor Appl Genet 2001; 102: 41-52
Morandini P., Salamini F. Plant biotechnology and breeding: allied for years to come. Trends Plant Sci 2003; 8: 70-75
Morgante M., Salamini F. From plant genomics to breeding practice. Curr Opin Biotech 2003; 14: 214-219
Morris M., Dreher K., Ribaut J.M., Khairallah M. Money matters (II): costs of maize inbred line conversion schemes at CIMMYT using conventional and marker-assisted selection. Mol Breed 2003; 11: 235-247
Mott R., Flint J. Simultaneous detection and fine mapping of quantitative trait loci in mice using heterogeneous stocks. Genetics 2002; 160: 1609-1618
Munns R. Physiological processes limiting plant-growth in saline soils: some dogmas and hypotheses. Plant Cell Environ 1993; 16: 15-24
Munns R., Passioura J.B., Guo J., Chazen O., Cramer G.R. Water elations and leaf expansion: importance of timing. Jour Exp Bot 2000; 51: 1495-1504
Nandi S., Subudhi P.K., Senadhira D., Manigbas N.L., Sen-Mandi S., Huang N. Mapping QTLs for submergence tolerance in rice by AFLP analysis and selective genotyping. Mol Gen Genet 1997; 255: 1-8
Ni J.J., Wu P., Senadhira D., Huang N. Mapping QTLs for phosphorus deficiency tolerance in rice (*Oryza sativa* L.). Theor Appl Genet 1998; 97: 1361-1369
Nguyen H.T., Babu R.C., Blum A. Breeding for drought resistance in rice: physiology and molecular genetics considerations. Crop Sci 1997; 37: 1426-1434
Nguyen B.D., Brar D.S., Bui B.C., Nguyen T.V., Pham L.N., Nguyen H.T. Identification and mapping of the QTL for aluminum tolerance introgressed from the new source, *Oryza rufipogon* Griff., into *indica* rice (*Oryza sativa* L.). Theor Appl Genet 2003; 106: 583-593
Ninamango-Cardenas F.E., Guimaraes C.T., Martins P.R., Parentoni S.N., Carneiro N.P., Lopes M.A., Moro J.R., Paiva E. Mapping QTLs for aluminum tolerance in maize. Euphytica 2003; 130: 223-232
Nuccio M.L., Rhodes D., McNeil S.D., Hanson A.D. Metabolic engineering of plants for osmotic stress resistance. Curr Opin Plant Biol 1999; 2: 128-134
Oono Y., Seki M., Nanjo T., Narusaka M., Fujita M., Satoh R., Satou M., Sakurai T., Ishida J., Akiyama K*., et al.* Monitoring expression profiles of Arabidopsis gene expression during rehydration process after dehydration using ca. 7000 full-length cDNA microarray. Plant Jour 2003; 34: 868-887
Ottaviano E., Sari-Gorla M., Pe E., Frova C. Molecular markers (RFLPs and HSPs) for genetic dissection of thermotolerance in maize. Theor Appl Genet 1991; 81: 713-719
Ozturk Z.N., Talamé V., Deyholos M., Michalowski C.B., Galbraith D.W., Gozukirmizi N., Tuberosa R., Bohnert H.J. Monitoring large-scale changes in transcript abundance in drought- and salt-stressed barley. Plant Mol Biol 2002; 48: 551-573
Pakniyat H., Powell W., Baird E., Handley L.L., Robinson D., Scrimgeour C.M., Nevo E., Hackett C.A., Caligari P.D.S., Forster B.P. AFLP variation in wild barley (*Hordeum spontaneum* C. Koch) with reference to salt tolerance and associated ecogeography. Genome 1997; 40: 332-341
Pan A., Hayes P.M., Chen F., Chen T.H.H., Blake T., Wright S., Karasi I., Bedo Z. Genetic

analysis of the components of winterhardiness in barley (*Hordeum vulgare* L.). Theor Appl Genet 1994; 89: 900-910

Passioura J.B. Drought and drought tolerance. Plant Growth Regul 1996; 20: 79-83

Passioura J.B. Environmental biology and crop improvement. Funct Plant Biol 2002; 29: 537-546

Patel J., McLeod L.E., Vries R.G.J., Flynn A., Wang X.M., Proud C.G. Cellular stresses profoundly inhibit protein synthesis and modulate the states of phosphorylation of multiple translation factors. Eur Jour Biotech 2002; 269: 3076-3085

Paterson A.H., Damon S., Hewitt J.K., Zamir D., Rabinowitch H.D., Lincoln S.E., Lander E.S., Tanksley S.D. Mendelian factors underlying quantitative traits in tomato: comparison across species, generations, and environments. Genetics 1991; 127: 181-197

Paterson A.H., Lan T.-H., Reischmann K.P., Chang C., Lin Y.-R., Liu S.-C., Burow M.D., Kowalski S.P., Katsar C.S., Del Monte T.A. Toward a unified genetic map of higher plants, transcending the monocot-dicot divergence. Nature Genet 1996; 14: 380-382

Pelleschi S., Guy S., Kim J.Y., Pointe C., Mahe A., Barthes L., Leonardi A., Prioul J.L. *Ivr2*, a candidate gene for a QTL of vacuolar invertase activity in maize leaves. Gene-specific expression under water stress. Plant Mol Biol 1999; 39: 373-380

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

Prasad S.R., Bagali P.G., Hittalmani S., Shashidhar H.E. Molecular mapping of quantitative trait loci associated with seedling tolerance to salt stress in rice (*Oryza sativa* L.). Curr Sci 2000; 78: 162-164

Price A.H., Tomos A.D. Genetic dissection of root growth in rice (*Oryza sativa* L.). II: Mapping quantitative trait loci using molecular markers. Theor Appl Genet 1997; 95: 143-152

Price A., Young E., Tomos A. Quantitative trait loci associated with stomatal conductance, leaf rolling and heading date mapped in upland rice (*Oryza sativa*). New Phytologist 1997; 137: 83-91

Price A., Steele K., Moore B., Barraclough P., Clark L. A combined RFLP and AFLP linkage map of upland rice (*Oryza sativa* L.) used to identify QTLs for root-penetration ability. Theor Appl Genet 2000; 100: 49-56

Price A.H., Cairns J.E., Horton P., Jones H.G., Griffiths H. Linking drought-resistance mechanisms to drought avoidance in upland rice using a QTL approach: progress and new opportunities to integrate stomatal and mesophyll responses. Jour Exp Bot 2002a; 53: 989-1004

Price A., Steele K., Gorham J., Bridges J., Moore B., Evans J., Richardson P., Jones R. Upland rice grown in soil-filled chambers and exposed to contrasting water-deficit regimes I. Root distribution, water use and plant water status. Field Crop Res 2002b; 76: 11-24

Price A., Townend J., Jones M., Audebert A., Courtois B. Mapping QTLs associated with drought avoidance in upland rice approach grown in the Philippines and West Africa. Plant Mol Biol 2002c; 48: 683-695

Prioul J.L., Quarrie S., Causse M., de Vienne D. Dissecting complex physiological functions through the use of molecular quantitative genetics. Jour Exp Bot 1997; 48: 1151-1163

Quarrie S.A. Implications of genetic differences in ABA accumulation for crop production. In: W.J. Davies, H.G. Jones (eds.) *Abscisic Acid: Physiology and Biochemistry*, Bios Scientific Publishers, Oxford, UK, 1991, pp. 227-243

Quarrie S.A. New molecular tools to improve the efficiency of breeding for increased drought resistance. Plant Growth Regul 1996; 20: 167-178

Quarrie S.A., Laurie D.A., Zhu J.H., Lebreton C., Semikhodskii A., Steed A., Witsenboer H., Calestani C. QTL analysis to study the association between leaf size and abscisic acid accumulation in droughted rice leaves and comparisons across cereals. Plant Mol Biol 1997; 35: 155-165

Quarrie S.A., Steed A., Semikhdoski A., Lebreton C., Calestani C., Clarkson D.A., Tuberosa

R., Sanguineti M.C., Melchiorre R., Prioul J. Identification of quantitative trait loci regulating water and nitrogen-use efficiency in wheat. In: *2nd STRESSNET Conference*, 1995, pp. 175-180

Quarrie S.A., Lazic Jancic V., Kovacevic D., Steed A., Pekic S. Bulk segregant analysis with molecular markers and its use for improving drought resistance in maize. Jour Exp Bot 1999a; 50: 1299-1306

Quarrie S.A., Stojanovic J., Pekic S. Improving drought resistance in small grained cereals: A case study, progress and prospects. Plant Growth Regul 1999b; 29: 1-21

Quarrie S.A., Steed A., Calestani C., Semikhodskii A., Lebreton C., Chinoy C., Steele N., Pljevljakusic D., Waterman E., Weyen J., *et al.* A genetic map of hexaploid wheat (*Triticum aestivum* L.) from the cross Chinese Spring × SQ1 and its use to compare QTLs for grain yield across a range of environments. Theor Appl Genet 2004; submitted

Ramsay L.D., Jennings D.E., Bohuon E.J.R., Arthur A.E., Lydiate D.J., Kearsey M.J., Marshall D.F. The construction of a substitution library of recombinant backcross lines in *Brassica oleracea* for the precision mapping of quantitative trait loci. Genome 1996; 39: 558-567

Rao S.A., McNeilly T. Genetic basis of variation for salt tolerance in maize (*Zea mays* L). Euphytica 1999; 108: 145-150

Reymond M., Muller B., Leonardi A., Charcosset A., Tardieu F. Combining quantitative trait loci analysis and an ecophysiological model to analyze the genetic variability of the responses of maize leaf growth to temperature and water deficit. Plant Physiol 2003; 131: 664-675

Ribaut J.M., Betran J. Single large-scale marker-assisted selection (SLS-MAS). Mol Breed 2000; 5: 531-541

Ribaut J.M., Hoisington D. Marker-assisted selection: new tools and strategies. Trends Plant Sci 1998; 3: 236-239

Ribaut J.M., Hoisington D.A., Deutsh J.A., Jiang C., Gonzalez-de-Leon D. Identification of quantitative trait loci under drought conditions in tropical maize. 1. Flowering parameters and the anthesis-silking interval. Theor Appl Genet 1996; 92: 905-914

Ribaut J.M., Hu X., Hoisington D., Gonzales-de-Leon D. Use of STSs and SSRs as rapid and reliable preselection tools in a marker-assisted selection backcross scheme. Plant Mol Biol Rep 1997a; 15: 156-164

Ribaut J.M., Jiang C., Gonzalez-de-Leon D., Edmeades G., Hoisington D.A. Identification of quantitative trait loci under drought conditions in tropical maize. 2. Yield components and marker-assisted selection strategies. Theor Appl Genet 1997b; 94: 887-896

Ribaut J.M., Gonzalez de Leon D., Jiang C., Edmeades G.O., Hoisington D. Identification and transfer of ASI quantitative trait loci (QTL): a strategy to improve drought tolerance in maize lines and populations. In: G.O. Edmeades, M. Bänziger, H.R. Mickelson, C.B. Peña Valdivia (eds.) *Proc Symp Developing Drought- and Low N-tolerant Maize*, CIMMYT, 1997c, pp. 396-400

Ribaut J.M., Banziger M., Betran J., Jiang C., Edmeades G.O., Dreher K., Hoisington D. Use of molecular markers in plant breeding: drought tolerance improvement in tropical maize. In: M.S. Kang (ed.) *Quantitative Genetics, Genomics, and Plant Breeding*, CABI Publishing, Wallingford, 2002, pp. 85-99

Richards R.A. Defining selection criteria to improve yield under drought. Plant Growth Regul 1996; 20: 157-166

Richards R.A. Selectable traits to increase crop photosynthesis and yield of grain crops. Jour Exp Bot 2000; 51: 447-458

Richards R.A., Condon A.G. Challenges ahead in using carbon isotope discrimination in plant-breeding programs. In: J.R. Ehleringer, A.E. Hall, G.D. Farquhar (eds.) *Stable isotopes and plant carbon-water relations*, Academic Press, San Diego, CA, USA, 1993, pp. 451-462

Richards R., Dennet C., Qualset C., Epstein E., Norlyn J., Winslow M. Variation in yield of grain and biomass in wheat, barley, and triticale in a salt-affected field. Fields Crop Res 1987; 15: 277-287

Richards R., Rebetzke G., Condon A., van Herwaarden A. Breeding opportunities for increasing the efficiency of water use and crop yield in temperate cereals. Crop Sci 2002; 42: 111-121

Robertson D.S. A possible technique for isolating genic DNA for quantitative traits in plant. Jour Theor Biol 1985; 117: 1-10

Robin S., Pathan MS., Courtois B., Lafitte R., Carandang S., Lanceras S., Amante M., Nguyen HT., Li-Z. Mapping osmotic adjustment in an advanced back-cross inbred population of rice. Theor Appl Genet 2003, 107, 1288-1296

Robinson D., Handley L.L., Scrimgeour C.M., Gordon D.C., Forster B.P., Ellis R.P. Using stable isotope natural abundances ($\delta^{15}N$ and $\delta^{13}C$) to integrate the stress responses of wild barley (*Hordeum spontaneum* C. Koch) genotypes. Jour Exp Bot 2000; 51: 41-50

Royo R., Villegas D., García del Moral L.F., El Hani S., Aparicio N., Rharrabti Y., Araus, J.L. Comparative performance of carbon isotope discrimination and canopy temperature depression as predictors of genotype differences in durum wheat yield in Spain. Aust Jour Agric Res 2001; 53: 1-9

Saini H.S., Westgate M.E. Reproductive development in grain crops during drought. Adv Agron 2000; 68: 59-96

Salekdeh G.H., Siopongco J., Wade L.J., Ghareyazie B., Bennett J. A proteomic approach to analyzing drought- and salt-responsiveness in rice. Fields Crop Res 2002; 76: 199-219

Salvi S., Tuberosa R., Sanguineti M.C., Landi P., Conti S. Molecular marker analysis of maize populations divergently selected for abscisic acid concentration in the leaf. Maize Genet Coop Newslett 1997; 71: 15-16.

Salvi S., Tuberosa R., Phillips R.L. Development of PCR-based assays for allelic discrimination in maize by using the 5-nuclease procedure. Mol Breed 2001; 8: 169-176

Salvi S., Tuberosa R., Chiapparino E., Maccarerri M., Veillet S., Van Beuningen L., Isaac P., Edwards K., Phillips R.L. Toward positional cloning of *Vgt1*, a QTL controlling the transition from the vegetative to the reproductive phase in maize. Plant Mol Biol 2002; 48: 601-613

Salvi S., Morgante M., Fengler K., Meeley B., Ananiev E., Svitashev S., Bruggemann E., Niu X., Li B., Tingey S.C., Tomes D., Miao G.H., Phillips R.L., Tuberosa R. Progress in the positional cloning of *Vgt1*, a QTL controlling flowering time in maize. In: *57th Annual Corn and Sorghum Research Conference*, 2003, pp. 1-18

Sanchez A.C., Subudhi P.K., Rosenow D.T., Nguyen H.T. Mapping QTLs associated with drought resistance in sorghum (*Sorghum bicolor* L. Moench). Plant Mol Biol 2002; 48: 713-726

Sanguineti M.C., Tuberosa R., Stefanelli S., Noli E., Blake T.K., Hayes P.M. Utilization of a recombinant inbred population to localize QTLs for abscisic acid content in leaves of drought-stressed barley (*Hordeum vulgare* L.). Russ J Plant Physiol 1994; 41: 572-576

Sanguineti M.C., Conti S., Landi P., Tuberosa R. Abscisic acid concentration in maize leaves: genetic control and response to divergent selection in two populations. Maydica 1996; 41: 193-203

Sanguineti M.C, Tuberosa R., Landi P., Salvi S., Maccaferri M., Casarini E., Conti S. QTL analysis of drought-related traits and grain yield in relation to genetic variation for leaf abscisic acid concentration in field-grown maize. Jour Exp Bot 1999; 50: 1289-1297

Sari Gorla M., Krajewski P., Di Fonzo N., Villa M., Frova C. Genetic analysis of drought tolerance in maize by molecular markers. II. Plant height and flowering. Theor Appl Genet 1999; 99: 289-295

Schachtman D., Liu W. Molecular pieces to the puzzle of the interaction between potassium and sodium uptake in plants. Trends Plant Sci 1999; 4: 281-287

Schadt E., Monks S., Drake T., Lusis A., Che N., Colinayo V., Ruff T., Milligan S., Lamb J., Cavet G., Linsley P., Mao M., Stoughton R., Friend S. Genetics of gene expression surveyed in maize, mouse and man. Nature 2003; 422: 297-302

Schwartz S.H., Tan B.C., Gage D.A., Zeevaart J.A.D., McCarty D.R. Specific oxidative cleavage of carotenoids by VP14 of maize. Science 1997; 276: 1872-1874

Schwartz S.H., Qin X.Q., Zeevaart J.A.D. Elucidation of the indirect pathway of abscisic acid biosynthesis by mutants, genes, and enzymes. Plant Physiol 2003; 131: 1591-1601

Serageldin I. Biotechnology and food security in the 21st century. Science 1999; 285: 387-389

Serraj R., Sinclair T.R. Osmolyte accumulation: can it really help increase crop yield under drought conditions? Plant, Cell Environ 2002; 25: 333-341

Sharp R., Wu Y., Voetberg G., Saab I., LeNoble M. Confirmation that abscisic acid accumulation is required for maize primary root elongation at low water potentials. Jour Exp Bot 1994; 45: 1743-1751

Shen L., Courtois B., McNally K.L., Robin S., Li Z. Evaluation of near-isogenic lines of rice introgressed with QTLs for root depth through marker-aided selection. Theor Appl Genet 2001; 103: 75-83

Siangliw M., Toojinda T., Tragoonrung S., Vanavichit A. Thai jasmine rice carrying QTLch9 (SubQTL) is submergence tolerant. Ann Bot 2003; 91: 255-261

Slafer G.A. Genetic basis of yield as viewed from a crop physiologist's perspective. Ann Appl Biol 2003; 142: 117-128

Slavich P.G., Read B.J., Cullis B.R. Yield response of barley germplasm to field variation in salinity quantified using the EM-38. Aust Jour Exp Agri 1990; 30: 551-556

Snape J.W., Semikhodskii A., Fish L., Sarma R.N., Quarrie S.A., Galiba G., Sutka J. Mapping frost tolerance loci in wheat and comparative mapping with other cereals. Acta Agron Hung 1997; 45: 268-270

Snape J.W., Sarma R., Quarrie S.A., Fish L., Galiba G., Sutka J. Mapping genes for flowering time and frost tolerance in cereals using precise genetic stocks. Euphytica 2001; 120: 309-315

Sorrells M.E., La Rota M., Bermudez-Kandianis C.E., Greene R.A., Kantety R., Munkvold J.D., Miftahudin, Mahmoud A., Ma X., Gustafson P.J., *et al.* Comparative DNA sequence analysis of wheat and rice genomes. Genome Research 2003; 13: 1818-1827

Sripongpangkul K., Posa G.B.T., Senadhira D.W., Brar D., Huang N., Khush G.S., Li Z.K. Genes/QTLs affecting flood tolerance in rice. Theor Appl Genet 2000; 101: 1074-1081

Stuber C.W., Edwards M.D., Wendel J.F. Molecular marker facilitated investigations of quantitative trait loci in maize. II. Factors influencing yield and its component traits. Crop Sci 1987; 27: 639-648

Stuber C.W., Polacco M., Senior M.L. Synergy of empirical breeding, marker-assisted selection, and genomics to increase crop yield potential. Crop Sci 1999; 39: 1571-1583

Subudhi P.K., Rosenow D.T., Nguyen H.T. Quantitative trait loci for the stay green trait in sorghum (*Sorghum bicolor* L. Moench): consistency across genetic backgrounds and environments. Theor Appl Genet 2000; 101: 733-741

Sutka J. Genetic control of frost tolerance in wheat (*Triticum aestivum* L.). Euphytica 1994; 77: 277-282

Sutka J., Snape J.W. Location of a gene for frost resistance on chromosome 5A of wheat. Euphytica 1989; 42: 41-44

Sutka J., Galiba G., Vagujfalvi A., Gill B.S., Snape J.W. Physical mapping of the *Vrn-A1* and *Fr1* genes on chromosome 5A of wheat using deletion lines. Theor Appl Genet 1999; 99: 199-202

Talamè V., Sanguineti M.C., Chiapparino E., Bahri H., Ben Salem M., Ellis R., Forster B.P., Rhouma S., Zoumarou W., Tuberosa R. Identification of agronomically valuable QTL alleles in wild barley (*Hordeum spontaneum*). Ann Appl Biol 2004; submitted

Tan B.C., Schwartz S.H., Zeevaart J.A.D., McCarty D.R. Genetic control of abscisic acid biosynthesis in maize. Proc Natl Acad Sci USA 1997; 94: 12235-12240

Tanksley S.D. Mapping polygenes. Ann Rev Genet 1993; 27: 205-233

Tanksley S., Nelson J. Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. Theor Appl Genet 1996; 92: 191-203

Tanksley S.D., Ganal M.W., Martin G.B. Chromosome landing: a paradigm for map-based gene cloning in plants with large genomes. Trends Genet 1995; 11: 63-68

Tardieu F. Virtual plants: modelling as a tool for the genomics of tolerance to water deficit. Trends Plant Sci 2003; 8: 9-14

Teulat B., Monneveux P., Wery J., Borries C., Souyris I., Charrier A., This D. Relationships between relative water content and growth parameters under water stress in barley: a QTL study. New Phytol 1997; 137: 99-107

Teulat B., This D., Khairallah M., Borries C., Ragot C., Sourdille P., Leroy P., Monneveux P., Charrier A. Several QTLs involved in osmotic-adjustment trait variation in barley (*Hordeum vulgare* L.). Theor Appl Genet 1998; 96: 688-698

Teulat B., Borries C., This D. New QTLs identified for plant water status, water-soluble two water regimes. Theor Appl Genet 2001a; 103: 161-170

Teulat B., Merah O., Souyris I., This D. QTLs for agronomic traits from a Mediterranean barley progeny grown in several environments. Theor Appl Genet 2001b; 103: 774-787

Teulat B., Merah O., Sirault X., Borries C., Waugh R., This D. QTLs for grain carbon isotope discrimination in field-grown barley. Theor Appl Genet 2002; 106: 118-126

Teulat B., Zoumarou-Wallis N., Rotter B., Ben-Salem M., Bahri H., This D. QTL for relative water content in field-grown barley and their stability across Mediterranean environments. Theor Appl Genet 2003; 108: 181-188

Thiellement H., Bahrman N., Damerval C., Plomion C., Rossignol M., Santoni V., de Vienne D., Zivy M. Proteomics for genetic and physiological studies in plants. Electrophoresis 1999; 20: 2013-2026

Thomas H., Howarth C.J. Five ways to stay green. Jour Exp Bot 2000; 51: 329-337

Till B.J., Reynolds S.H., Greene E.A., Codomo C.A., Enns L.C., Johnson J.E., Burtner C., Odden A.R., Young K., Taylor N.E., Henikoff J.G., Comai L., Henikoff S. Large-scale discovery of induced point mutations with high-throughput TILLING. Genome Res 2003; 13: 524-530

Toojinda T., Siangliw M., Tragoonrung S., Vanavichit A. Molecular genetics of submergence tolerance in rice: QTL analysis of key traits. Ann Bot 2003; 91: 243-253

Touzet P., Winkler R.G., Helentjaris T. Combined genetic and physiological analysis of a locus contributing to quantitative variation. Theor Appl Genet 1995; 91: 200-205

Tripathy J.N., Zhang J., Robin S., Nguyen T.T., Nguyen H.T. QTLs for cell-membrane stability mapped in rice (*Oryza sativa* L.) under drought stress. Theor Appl Genet 2000; 100: 1197-1202

Tuberosa R., Sanguineti M.C., Landi P. Abscisic acid concentration in the leaf and xylem sap, leaf water potential, and stomatal conductance in drought-stressed maize. Crop Sci 1994; 34: 1557-1563

Tuberosa R., Galiba G., Sanguineti M.C., Noli E., Sukta J. Identification of QTL influencing freezing tolerance in barley. Acta Agron Hung 1997; 45: 413-417

Tuberosa R., Parentoni S., Kim T.S., Sanguineti M.C., Phillips R.L. Mapping QTLs for ABA concentration in leaves of a maize cross segregating for anthesis date. Maize Genet Coop Newslett 1998a; 72: 72-73

Tuberosa R., Sanguineti M.C., Landi P., Salvi S., Casarini E., Conti S. RFLP mapping of quantitative trait loci controlling abscisic acid concentration in leaves of drought-stressed maize (*Zea mays* L.). Theor Appl Genet 1998b; 97: 744-755

Tuberosa R., Gill B.S., Quarrie S. Cereal genomics: ushering in a brave new world. Plant Mol Biol 2002a; 48: 445-449

Tuberosa R., Salvi S., Sanguineti M.C., Landi P., Maccaferri M., Conti S. Mapping QTLs regulating morpho-physiological traits and yield: case studies, shortcomings and perspectives in drought-stressed maize. Ann Bot 2002b; 89: 941-963

Tuberosa R., Sanguineti M.C., Landi P., Giuliani M.M., Salvi S., Conti S. Identification of QTLs for root characteristics in maize grown in hydroponics and analysis of their overlap with QTLs for grain yield in the field at two water regimes. Plant Mol Biol 2002c; 48: 697-712

Tuberosa R., Salvi S., Sanguineti M.C., Maccaferri M., Giuliani S., Landi P. Searching for QTLs controlling root traits in maize: a critical appraisal. Plant and Soil 2003; 255: 35-54

Tuinstra M.R., Grote E.M., Goldsbrough P.B., Ejeta G. Genetic analysis of post-flowering drought tolerance and components of grain development in *Sorghum bicolor* (L.) Moench. Mol Breed 1997; 3: 439-448

Tuinstra M.R., Ejeta G., Goldbrough P. Evaluation of near-isogenic sorghum lines contrasting for QTL markers associated with drought tolerance. Crop Sci 1998; 38: 835-842

Turner N.C. Further progress in crop water relations. Adv Agron 1997; 528: 293-338

Turpeinen T., Tenhola T., Manninen O., Nevo E., Nissila E. Microsatellite diversity associated with ecological factors in *Hordeum spontaneum* populations in Israel. Mol Ecol 2001; 10: 1577-1591

Uesono Y., Toh E.A. Transient inhibition of translation initiation by osmotic stress. Jour Biol Chem 2002; 277: 13848-13855

Vagujfalvi A., Galiba G., Cattivelli L., Dubcovsky J. The cold-regulated transcriptional activator *Cbf3* is linked to the frost-tolerance locus *Fr-A2* on wheat chromosome 5A. Mol Gen Genet 2003; 269: 60-67

Van Buuren M., Salvi S., Morgante M., Serhani B., Tuberosa R. Comparative genomic mapping between a 754 kb region flanking DREB1A in *Arabidopsis thaliana* and maize. Plant Mol Biol 2002; 48: 741-750

Veisz O., Sutka J. Ditelosomic analysis of frost resistance in wheat (cv Chinese Spring). Cereal Res Commun 1993; 21: 263-267

Venuprasad R., Shashidhar H.E., Hittalmani S., Hemamalini G.S. Tagging quantitative trait loci associated with grain yield and root morphological traits in rice (*Oryza sativa* L.) under contrasting moisture regimes. Euphytica 2002; 128: 293-300

Wan J.L., Zhai H.Q., Wan J.M., Ikehashi H. Detection and analysis of QTLs for ferrous iron toxicity tolerance in rice, *Oryza sativa* L. Euphytica 2003; 131: 201-206

Weidong C., J. Jizeng, J. Jiyun. Identification and interaction analysis of QTL for phosphorus use efficiency in wheat seedlings. In: *14th Intern Plant Nutrition Colloquium, Plant Nutrition: Food Security and Sustainability of Agro ecosystems through Basic and Applied Research*, Hannover, Germany, 2001, pp. 76-77

Westgate M.E., Boyer J.S. Osmotic adjustment and the inhibition of leaf, root, stem and silk growth at low water potentials in maize. Planta 1985; 164: 540-549

Winkler R., Helentjaris T. The maize *dwarf3* gene encodes a cytochrome P450 mediated early step in gibberellin biosynthesis. Plant Cell Environ 1995; 7: 1307-1317

Wissuwa M., Yano M., Ae N. Mapping of QTLs for phosphorus-deficiency tolerance in rice (*Oryza sativa* L.). Theor Appl Genet 1998; 5-6: 777-783

Wissuwa M., Wegner J., Ae N., Yano M. Substitution mapping of *Pup1*: a major QTL increasing phosphorus uptake of rice from a phosphorus-deficient soil. Theor Appl Genet 2002; 105: 890-897

Witcombe J.R., Hash C.T. Resistance gene deployment strategies in cereal hybrids using marker-assisted selection: gene pyramiding, three-way hybrids and synthetic parent populations. Euphytica 2000; 112: 175-186

Xu K.N., Mackill D.J. A major locus for submergence tolerance mapped on rice

chromosome 9. Mol Breed 1996; 2: 219-224
Xu K.N., Xu X., Ronald P.C., Mackill D.J. A high-resolution linkage map of the vicinity of rice submergence tolerance locus Sub1. Mol Gen Genet 2000; 263: 681-689
Xu W.W., Subudhi P.K., Crasta O.R., Rosenow D.T., Mullet J.E., Nguyen H.T. Molecular mapping of QTLs conferring stay-green in grain sorghum (*Sorghum bicolor* L. Moench). Genome 2000; 43: 461-469
Yadav R., Courtois B., Huang N., McLaren G. Mapping genes controlling root morphology and root distribution in a doubled-haploid population of rice. Theor Appl Genet 1997; 94: 619-632
Yadav R.S., Hash C.T., Bidinger F.R., Cavan G.P., Howarth C.J. Quantitative trait loci associated with traits determining grain and stover yield in pearl millet under terminal drought-stress conditions. Theor Appl Genet 2002; 104: 67-83
Yamaya T., Obara M., Nakajima H, Sasaki S, Hayakawa T; Sato T. Genetic manipulation and quantitative-trait loci mapping for nitrogen recycling in rice Jour Exp Bot 2002; 53: 917-925
Yang J., Sears R.G., Gill B.S., Paulsen G.M. Quantitative and molecular characterization of heat tolerance in hexaploid wheat. Euphytica. 2002; 126: 275-282
Yano M., Sasaki T. Genetic and molecular dissection of quantitative traits in rice. Plant Mol Biol 1997; 35: 145-153
Yates J.R. Mass spectrometry and the age of the proteome. Jour Mass Spect 1998; 33: 1-19
Yin X.Y., Stam P., Kropff M.J., Schapendonk A.H.C.M. Crop modeling, QTL mapping, and their complementary role in plant breeding. Agron Jour 2003; 95: 90-98
Young N.D. A cautiously optimistic vision for marker-assisted breeding. Mol Breed 1999; 5: 505-510
Yu L.X., Setter T.L. Comparative transcriptional profiling of placenta and endosperm in developing maize kernels in response to water deficit. Plant Physiol 2003; 131: 568-582
Zeevaart J.A.D., Creelman R.A. Metabolism and physiology of abscisic acid. Annu Rev Plant Physiol Plant Mol Biol 1988; 39: 439-473
Zeng Z.B. Precision mapping of quantitative trait loci. Genetics 1994; 136: 1457-1468
Zhang Z.B., Xu P. Reviews of wheat genome. Hereditas-Beijing 2002; 24: 389-394
Zhang J., Davies W.J. Does ABA in the xylem control the rate of leaf growth in soil dried maize and sunflower plants? Jour Exp Bot 1990; 41: 1125-1132
Zhang W.P., Shen X.Y., Wu P., Hu B., Liao C.Y. QTLs and epistasis for seminal root length under a different water supply in rice (*Oryza sativa* L.). Theor Appl Genet 2001a; 103: 118-123
Zhang J., Zheng H.G., Aarti A., Pantuwan G., Nguyen T.T., Tripathy J.N., Sarial A.K., Robin S., Babu R.C., Nguyen B.D*., et al.* Locating genomic regions associated with components of drought resistance in rice: comparative mapping within and across species. Theor Appl Genet 2001b; 103: 19-29
Zheng H.G., Babu R.C., Pathan M.S., Ali L., Huang N., Courtois B., Nguyen H.T. Quantitative trait loci for root-penetration ability and root thickness in rice: comparison of genetic backgrounds. Genome 2000; 43: 53-61
Zhu B., Choi D.W., Fenton R., Close T.J. Expression of the barley dehydrin multigene family and the development of freezing tolerance. Mol Gen Genet 2000; 264: 145-153
Zinselmeier C., Westgate M.E., Schussler J.R., Jones R.J. Low water potential disrupts carbohydrate metabolism in maize (*Zea mays* L.) ovaries. Plant Physiol 1995; 107: 385-391
Zinselmeier C., Jeong B., Boyer J. Starch and the control of kernel number in maize at low water potentials. Plant Physiol*;* 1999; 121: 25-36

Zinselmeier C., Sun Y., Helentjaris T., Beatty M., Yang S., Smith H., Habben J. The use of gene expression profiling to dissect the stress sensitivity of reproductive development in maize. Fields Crop Res 2002; 75: 111-121

Zivy M., de Vienne D. Proteomics: a link between genomics, genetics and physiology. Plant Mol Biol 2000; 44: 575-580

Chapter 10

MARKER ASSISTED SELECTION IN THE CEREALS: THE DREAM AND THE REALITY

Robert M.D. Koebner

Crop Genetics Department, John Innes Centre, Colney Lane, Norwich NR4 7UH, UK
Email*:* robert.koebner@bbsrc.ac.uk

1. INTRODUCTION

A variety of DNA-based markers has become available over the last two decades (see Chapter 2 by Somers in this book). While the development of some of these markers has been driven directly by the requirement for better tools for genetic analysis, others have been a happy by-product of genomics research. In particular, the growing feasibility of applying high throughput platforms to molecular marker technology has been primarily facilitated by developments in genomics research. Thus in 1996, the mapping of 5000 microsatellite loci in the human genome merited a major publication in Nature (Dib *et al.*, 1996), but by 2003, the number of known human single nucleotide polymorphisms (SNPs) had already exceeded five million (http://www.ncbi.nlm.nih.gov/SNP/snp_summary.cgi). Therefore the expectation is that the number of SNP assays in the cereals is set to increase dramatically. With the coming together of a large number of markers and the expected development of realistic high throughput platforms, it is now timely to explore present and future prospects of using marker technology in real cereal breeding.

Markers provide a tool for indirect selection, but the definition of marker-trait associations is the second element necessary for the deployment of marker-assisted selection (MAS). Elucidating the inheritance of quantitatively inherited traits by searching for linkage between the loci controlling them and factors which segregate in a simple Mendelian

P.K. Gupta and R.K. Varshney (eds.), Cereal Genomics, 317-329.

fashion is a concept that has its roots in the early days of genetics. Thus, already in 1918, Payne showed that the genetics of bristle number in *Drosophila* could be simplified by the demonstration of its linkage to loci mapped to the X chromosome. Similarly, easily scorable colour phenotypes were used to predict continuously variable traits such as seed weight in *Phaseolus* (Sax, 1923), and fruit size in tomato (Lindstrom, 1924). But because of the lack of a sufficient supply of simply inherited morphological variants, little progress in marker assisted plant breeding was made until 1960s, when biochemical markers, based mainly on the histological staining of specific enzyme activities, were developed. At that time, the discovery of isozyme-determining loci linked to genes determining plant traits enjoyed a brief period of prominence. For instance, in maize, linkage was established between a number of such marker loci and factors determining yield (Stuber *et al.*, 1980) and some of these markers were used to select for yield improvement (Stuber *et al.*, 1982). However, their widespread use in breeding was limited, mainly because of difficulties in scaling up the assays to the levels needed in breeding programmes. In rice, isozyme analysis was most prominently used to define a number of broad geographical types (Glaszmann, 1987), a grouping, which has remained largely intact, even after the advent of molecular markers. In wheat, two biochemical assays developed in 1980s are still in use across a number of diverse breeding programmes (see below). A general feature of biochemical markers is that although their unit cost is relatively low, their potential for high throughput genotyping is limited, because neither sample preparation nor processing can be subjected to automation. Thus their application is largely limited to either the genotype building for providing suitable parents for crossing, or to genotype improvement via directed back-crossing, where the number of individuals to be screened is relatively modest. Many of the difficulties encountered with the deployment for MAS using morphological or biochemical markers have now been largely overcome by current DNA-based ones.

2. BENEFITS OF MAS

MAS can, in principle, increase the precision of breeding in at least four ways: first, the breeder can select on a single plant basis for a trait (or trait combination), where this may be neither appropriate nor possible by conventional phenotypic selection (CPS) - either because of poor heritability, or because the trait per se is difficult to score or cost-ineffective to analyse. Both these scenarios are commonplace in breeding programmes. Second, MAS facilitates the maintenance, and ultimately the fixation of a number of individual genetic components, which acting together, define the overall expression of the trait. This pertains even when, as is the case for many

quantitative traits, each individual component may make only a relatively modest contribution to the overall determination of the phenotype. With the exception of race-specific disease resistance and a small number of quality related traits (such as waxy endosperm), most breeders' traits are under such polygenic control. Third, in segregating generations, it allows the selection of not only recessive genes, but also of those not readily amenable to CPS, without any need for validation at each generation via a progeny test. This is of particular importance in backcross programmes, where the aim is to correct an established genotype (the recurrent parent) for a single weakness (for example, susceptibility to disease) by introducing the minimum amount of genetic material from the donor of the desirable trait. Finally, markers can help in the choice of parents in crossing programmes: in some situations, this can be directed to maximise diversity, which is desirable for the exploitation of heterosis; in other situations, the aim may be to minimise it, in order to preserve intact adaptive gene complexes painstakingly built up in elite inbred germplasm.

3. PREREQUISITES AND CONSTRAINTS FOR MAS

A critical requirement for MAS is, of course, the existence of sufficient marker polymorphism. It has long been known that the level of DNA marker polymorphism is high in maize (Evola *et al.*, 1986), but despite this, the large-scale deployment of MAS did not gather any significant momentum until relatively recently, some 15 years after the first RFLP-based genetic map of maize was developed (Helentjaris *et al.* 1986). Even in the less polymorphic cereals, prominently wheat, the level of polymorphism is not likely to be the major constraint on MAS uptake, although it has been argued in the past to be so. This is because, as noted above, SNP technology is likely to remove any effective limitation on marker discovery, in an even more spectacular way than the development of microsatellite assays has already begun to do.

As noted by Lee (1995), MAS has every chance of being effective in situations where CPS can achieve genetic gain; therefore, whether or not it can be deployed will be largely determined by an analysis of the costs and benefits to the breeding effort. This translates into issues of efficiency, which include, along with the major factor of cost per assay, also considerations of throughput per unit time, and timeliness. The calculation of the unit cost of MAS is complex, and depends on what is included and what is not. Dreher *et al.* (2003) recently presented a detailed analysis of CIMMYT's costings for SSR genotyping and CPS in maize, and pointed out how case-dependent these are. Throughput capacity is a critical consideration, since no matter how

cheap an assay is, to be applicable to a breeding project, it must be scalable to the population sizes generally used, which is itself a function of the number of discrete genes/QTL which are likely to be segregating in the material. Finally, the timeliness factor relates to those traits, which are either phenotypable only late in the growing season (so that MAS can save resources by allowing selection early in the season), or are first amenable to CPS only in advanced generations, so that early generation MAS will result in a considerable saving of resources.

4. STATUS OF MAS IN THE CEREALS

Although the potential benefits of MAS are substantial, its actual uptake in practical breeding programmes, among the cereals and elsewhere, has been patchy. Only relatively recently has it begun to make more than a marginal impact on breeding methodology. The extremes in terms of large scale MAS deployment in the cereals are represented by maize, where uptake is substantial, and wheat, where its extent is less spectacular. Barley is similar to wheat in terms of breeding system, but has enjoyed more progress, possibly as a result of its simpler (non-polyploid) genetics; while rice is particularly relevant because of its global importance both as a crop species in its own right, and as a model species for the cereals in general. The remaining small-grain cereals (oats, rye, triticale, sorghum, the millets and tef) are largely too minor to have enjoyed any significant investment in marker discovery and commercial deployment.

4.1. Maize

Maize breeding in developed countries is carried out largely by large private sector entities (which do not freely, for understandable competition reasons, publicise their procedures). This contrasts with the situation in other major cereal species, where breeding is more commonly, although not universally, carried out by public sector organisations. Globally, maize production is dominated by F_1 hybrids, and this has far-reaching consequences on the financial returns of its breeding. First, seed sales of the same genotype can be made each season (there is no loss of revenue as a result of the use of farm-saved seed), and second, the inbred components of a successful hybrid are not available to competitors to use as parental material for their own programmes of variety improvement. In contrast, cultivars of wheat (similarly barley) are unprotected, except through Plant Breeders' Rights, both from farm-saved seed and from inclusion into a competitor's breeding programme. However,

for rice, recent improvements in the management of male sterility have resulted in a significant expansion of F_1 hybrid breeding, especially in China and in the USA. The expectation is therefore that MAS will eventually grow in importance in rice, mirroring the trend of maize.

Following the development of the maize RFLP genetic map, the late 1980s saw the exploration of opportunities for the commercial application of MAS in maize breeding by companies such as Dekalb Genetics, which formed a joint research venture with a marker company (Native Plants, Inc.) for this purpose (Johnson, 2003). Efforts were also made to assess the selection efficiency of MAS, particularly where the MAS targets involved one or more QTL. A typical conclusion was arrived at by Eathington *et al.* (1997), who showed that markers were able to significantly increase the precision with which superior advanced selfed-generation lines could be identified. It is intuitive that MAS will work in any situation where CPS has been able to achieve progress, provided that the prerequisite marker-trait associations are known. Thus, where QTLs are themselves unreliable, either because of epistasis (leading to unpredictability of expression in genetic backgrounds other than in the one in which they have been detected) or to genotype x environment interactions (so that the effect is environment-dependent), MAS directed at QTL is also unreliable; but where such interactions are insignificant, genetic progress is predictable and MAS, if economically justifiable, will be advantageous.

An important example of the successful deployment of MAS in maize is the use of three SSR markers, namely *umc1066*, *phi057* and *phi112* (developed at CIMMYT) for development of quality protein maize (QPM) through marker aided transfer of *opaque2* gene in backcross programmes (Dreher et al., 2000). In this programme, the genotype of the recurrent parent could also be recovered using MAS within three generations, instead of six generations generally needed in conventional backcross breeding.

Major investment in MAS infrastructure, however, is being made by the large private sector players in maize breeding (in the USA primarily Pioneer Hi-Bred, Syngenta and Monsanto; in Europe, in addition to these, KWS and Limagrain). In Monsanto's US operation, the past five years have seen 'the development of thousands of new marker assays, a 17-fold increase in the acquisition rate of marker data, and a decrease in unit data point cost of 75%' (M. Edwards, Monsanto Company, St. Louis USA, personal communication). Much of this scale-up is dependent upon the automation of the whole process of marker genotyping, and is increasingly reliant on SNPs as the marker technology. Yield-directed MAS typically targets around 20 QTL, and is achieved by incorporating three generations of purely marker-based selection

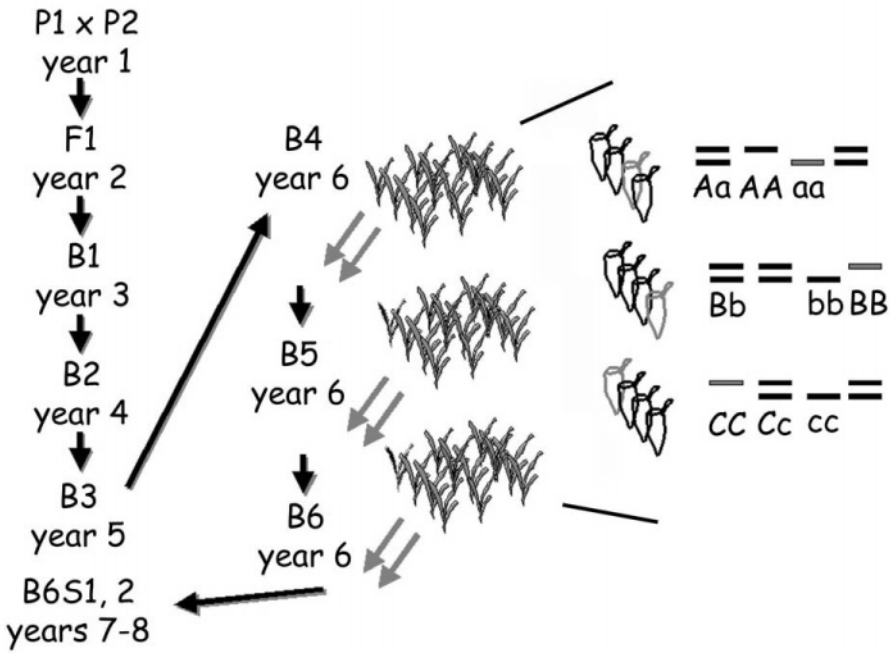


Figure 1. Three rounds of MAS incorporated into a breeding scheme for maize inbred selection (adapted from M. Edwards, pers. comm.).

within one year, thereby adding one year to the conventional five year cycle of inbred line production (Fig. 1).

This protocol is claimed to generate a doubling in breeding gain over CPS, and to produce a saving of one to two years in the time required to backcross a specific trait into an elite inbred. Such a level of commitment to MAS technology is likely to be self-reinforcing, so that as experience and infrastructure related to the technology and its deployment grow, its use will become more pervasive.

4.2. Wheat

In contrast to maize, wheat is a naturally inbreeding species. Although a level of heterosis has been demonstrated, difficulties in enforcing cross-pollination in a reliable and cost-effective way have hindered the development of F_1 wheat hybrids for commercial cultivation. Most varietal development programmes are based on versions of the long-established pedigree breeding system, where large F_2 populations are generated, and selection is carried out in early generations for highly heritable, qualitative

traits (such as disease resistance) and in later generations for the quantitative traits (primarily yield and quality). Since most varieties are bred and grown as pure breeding lines, it is much more difficult for the wheat breeder than it is for the maize breeder to protect the product. Thus, although the volume of seed sale may be comparable, the value of the wheat seed market is much smaller, and thus the economic margin for the breeder is far lower. This has far-reaching implications on the economics of MAS deployment in wheat.

The use of MAS in wheat has a history of about 20 years, and until very recently, involved the exploitation of two main non-DNA based assays. The first, which has been retained with only slight modifications ever since its inception, exploits a correlation between bread-making quality and allelic status at the *Glu-1* (endosperm storage protein subunit) loci. Electrophoretic profiles are obtained by the relatively straightforward and cost effective procedure SDS-PAGE from crude seed protein extracts. Combinations of *Glu-1* alleles are assigned a 'glutenin' score, which is used to predict bread-making quality in wheat breeding programmes worldwide, even though it only explains up to half of the overall variation in dough quality (Payne *et al.,* 1987, Rogers *et al.,* 1989). The second MAS assay in wheat is diagnostic, as a result of tight genetic linkage, for the presence/absence of a gene introduced from a wild grass, which confers a high level of resistance to eyespot, a stem disease. The MAS assay is performed on an aqueous extract of a single seed, separated by isoelectric focussing and detected by a histological stain for the endopeptidase isozymes (Summers *et al.,* 1987). However, in the last few years, the number of loci for which DNA-based assays have been generated has increased dramatically, and the majority of these use PCR as a technology platform (Gupta *et al.*, 1999).

Although the take-up of MAS in wheat has been slow, its potential is now measurable (Eagles *et al.,* 2001). In 2001, a consortium of US public sector breeders ('MASwheat': http://maswheat.ucdavis.edu/index.htm) was formed with the aim to 'transfer new developments in wheat genomics and biotechnology to wheat production'. The programme is targeting some 19 traits, spread across disease/pest resistance and end-use quality, and is seeking the marker assisted transfer of these target genes through MAS technology into a collection of 99 lines and cultivars, distributed among the most relevant market classes of bread and pasta wheat. The critical issue remains, however, the unit assay cost. Using a capillary sequencing platform, the cost for the generation of a single microsatellite data point (including only the cost of the PCR reaction itself and excluding any costs associated with the acquisition of DNA template) has been estimated to be in the region of US$0.40 (Koebner and Summers, 2003), while a more extensive calculation made by Dreher *et al.* (2003) puts its full economic cost at between US$1 and

US$2. At such a level of cost, the widespread application of MAS in a low end-value crop such as wheat cannot be justified. Its uptake for the moment therefore remains restricted to low volume applications, such as genotype building by backcrossing, and to the development of niche genotypes such as waxy wheats, which can command a price premium. However, the ongoing development of automated assay platforms, extending to both the pre-assay (particularly DNA extraction) and post-assay (data point collection) stages, will inevitably reduce the unit cost per assay and thus will allow an increase in the number of assays possible. One can anticipate therefore that, at least in those large breeding programmes which enjoy economies of scale and are able to make the necessary capital investment, MAS assays will increasingly become feasible. Indeed some large commercial breeding companies are already making the investment to build up MAS capability for wheat breeding. It is instructive to note that exactly the same considerations are seen to be relevant for the widely heralded exploitation of human DNA polymorphisms to predict differential drug response, an area now carrying the label 'pharmacogenomics'. Although the average cost of a human SNP assay fell from US$1.00 to US$0.10 over a recent 12 month period, it was suggested that a further reduction to US$0.01 per assay would be required before wide-scale usage of the technology could become feasible (Roses, 2002).

4.3. Other Cereals: Barley and Rice

Barley breeding closely resembles that of wheat in both structure and economics. It is a self-pollinating species, and like wheat, its end value is low, so that the selection methods used by barley breeders are very similar to those used by wheat breeders. Varieties worldwide to date have been almost exclusively released as pure breeding inbred selections. Only rarely have barley F_1 hybrids been used for commercial cultivation. The first ('Hembar') was released in the USA in 1969-70, through the successful use of balanced teritiary trisomics; more recently barley F_1 hybrids have been launched in the UK, entering official trials in 2000, with commercial seed expected to be available in autumn 2003 (http://www.newfarmcrops.co.uk). An extensive collection of microsatellite assays has been assembled by Ramsay *et al.* (2000) and Varshney *et al.* (unpublished, personal communication). Despite the somewhat downbeat sentiments of Thomas (2003) to the effect that molecular breeding 'is not widely used, other than as a marker for BaYMV (barley yellow mosaic virus) resistance', more has been achieved in barley than in wheat. Thus, the two genes *rym4* and *rym5* are both marked by microsatellites, and MAS is in wide use by European breeders (e.g., Tuvesson *et al*., 1998). An aggressive approach has been taken in Australia, where the

South Australian Barley Improvement Program has deployed markers to eliminate defects in elite varieties and has advanced a "Sloop type" variety with CCN resistance to near commercial release in less than eight years (compared to the average time to release for malting varieties in Australia of 15 years) (Annual Report, CRCMPB, 2001-2002). The US variety 'Tango', released in 2000, is claimed to be the first commercially released barley variety using molecular MAS. It contains two QTL for adult resistance to stripe rust, a character difficult to handle by CPS (Toojinda *et al.,* 1998), which had been transferred into the 1970s variety 'Steptoe' via two cycles of RFLP-aided backcrossing (Hayes *et al.*, 2003). Although 'Tango' has a good level of rust resistance, its yield is less than that of its long outclassed recurrent parent, and hence is seen primarily as a genetically characterised source of resistance, rather than as a variety in its own right. Just as for wheat, most of the proposed targets for MAS in barley relate to genes for disease resistance, although for many of these diseases, efficient phenotypic screens are available. A QTL target for MAS breeding, unique to barley, is malting quality, as this trait attracts a substantial price premium (Han *et al.*, 1997; Ayoub *et al.,* 2003; Hayes *et al.*, 2003).

Although rice is also a self-pollinating species, there has been a sustained effort to develop commercial F_1 hybrids, particularly in the USA, China, India and SE Asia. This has led to an increased use of F_1 hybrid varieties (particularly in China). Interestingly, this has not been the trend in Japan, where because of a strong consumer preference for *japonica* types, the use of *japonica* x *indica* hybrids is restricted. Genetic maps and molecular marker collections are also well developed. The status of rice as a genomic model is set to promote the application of MAS in breeding, since much of the gene cloning activity in monocot species will be driven by the availability of the whole rice genome sequence and a worldwide investment in technologies associated with large scale gene isolation. Much of the progress in rice MAS to date has centred on the pyramiding of disease resistance genes, particularly against blight (Sanchez *et al.,* 2000) and blast (Singh *et al.,* 2001), and both simultaneously (Narayanan *et al.,* 2002). The year 2002 also witnessed the release of two Indonesian cultivars ('Angke' and 'Conde'), in which MAS was used to introduce *xa5* into a background containing *xa4* (Toenniessen *et al.,* 2003). The *Xa21* gene is particularly significant since it has been cloned (Song *et al.*, 1995). This has facilitated the development of perfect within-gene markers for its tracking in segregating materials and its ready transfer between genotypes by transgenesis. Consequently, the selection of *Xa21* by MAS features in a number of programmes. Two other major targets for MAS in rice include resistance to gall midge, controlled by a number of single genes (Katiyar *et al.,* 2001) and resistance to brown planthopper, where both qualitative (Jena *et al.,* 2003) and quantitative (Xu *et al.,* 2002) inheritance

have been identified. A number of quality (Dong *et al.,* 2003; Zhou *et al.,* 2003; Ramalingam *et al.,* 2002) and breeders' quantitative traits (such as root depth: Shen *et al.,* 2001) with known QTL are currently also being considered.

5. SUMMARY AND OUTLOOK

In 1999, Young set out his 'cautiously optimistic vision' for marker-assisted selection. Five years on, the situation is starting to crystallise. The marker technology itself is no longer limiting. With respect to marker availability, SNPs will soon represent a source of plentiful within-gene markers, and these will shortly become available for all the major cereals. The 'big biology' spawned by the genomics revolution has brought miniaturisation and automation to biological assays, so that levels of throughput relevant to the plant breeding process are becoming attainable. The issue that remains is the affordability of large-scale MAS, particularly in low value crops. Because cereals are primarily broad-acre commodity products, their value is generally low, and this will slow the widespread adoption of MAS, except where F_1 hybrid seed is a viable proposition. However, as economies of scale and improvements in technology drive down the assay price, the penetration of MAS into commercial cereal breeding will undoubtedly grow. For maize, this stage has already been reached, while for rice, the medium term prospects are good. However, for wheat and barley, MAS is likely to remain less central to the breeding process, and will be deployed only for specific purposes. These include the accelerated selection of a few traits that are difficult to manage via CPS (exemplified by the endopeptidase assay for eyespot resistance in wheat, and CCN resistance in barley); for the selection of recessive alleles in backcrossing programmes; for the pyramiding of disease resistance genes; and for guiding the choice of parents to be used in crossing programmes.

6. ACKNOWLEDGMENTS

Particular thanks are due to Marlin Edwards (Monsanto Company, St Louis, USA) for providing a commercial viewpoint on MAS exploitation in maize breeding, and to the two anonymous reviewers for suggesting improvements to the original version of the manuscript.

7. REFERENCES

Ayoub M., Armstrong E., Bridger G., Fortin M.G., Mather D.E. Marker-based selection in barley for a QTL region affecting alpha-amylase activity of malt. Crop Sci 2003; 43: 556-561

CRCMPB Annual Report 01-02 Putting plant breeding into fast forward. Cooperative Research Centre for Molecular Plant Breeding, University of Adelaide, Australia, p51

Dib C., Fauré S., Fizaumes C., Samson D., Drouot N., Vignal A., Millasseau P., Marc S., Hazan J., Seboun E., *et al.* A comprehensive genetic map of the human genome based on 5,264 microsatellites. Nature 1996; 380: 152-154

Dong Y.J., Tsuzuki E., Kamiunten H., Terao H., Lin D.Z., Matsuo M., Zheng Y.F. Identification of quantitative trait loci associated with pre-harvest sprouting resistance in rice (*Oryza sativa* L.). Field Crops Res 2003; 81: 133-139

Dreher K. Morris M., Khairallah M., Ribaut J.M., Pandey S., Srinivasan G. Is Marker-Assisted selection cost-effective compared to conventional plant breeding methods? The case of quality protein maize. In: Proc 4th Annu Conf Intern Consor on Agricultural Biotechnology Research (ICABR), "The Economics of Agricultural Biotechnology," Ravello, Italy, 2000

Dreher K., Khairallah M., Ribaut J.M., Morris M. Money matters (I): costs of field and laboratory procedures associated with conventional and marker-assisted maize breeding at CIMMYT. Mol Breed 2003; 11: 221-234

Eagles H.A., Bariana H.S., Ogbonnaya F.C., Rebetzke G.J, Hollamby G.J, Henry R.J., Henschke P.H., Carter M. Implementation of markers in Australian wheat breeding. Aust Jour Agric Res 2001; 52: 1349-1356

Eathington S.R., Dudley J.W., Rufener G.K. Usefulness of marker-QTL associations in early generation selection. 1997; Crop Sci 37: 1686-1693

Evola S.V., Burr F.A., Burr B. The suitability of restriction fragment length polymorphisms as genetic markers in maize. Theor Appl Genet 1986; 71: 765-771.

Glaszmann J.C. Isozymes and classification of Asian rice varieties. Theor Appl Genet 1987; 74: 21-30.

Gupta P.K., Varshney R.K., Sharma P.C., Ramesh B. Molecular markers and their applications in wheat breeding. Plant Breed 1999; 118: 369-390.

Han F., Romagosa I., Ullrich S.E., Jones B.L., Hayes P.M., Wesenberg D.M. Molecular marker-assisted selection for malting quality traits in barley. Mol Breed 1997; 3: 427-437

Hayes P.M., Corey A.E., Mundt C., Toojinda T., Vivar H. Registration of 'Tango' barley. Crop Sci 2003; 43: 729-731

Hayes P.M., Castro A., Marquez-Cedillo L., Corey A., Henson C., Jones B.L., Kling J., Mather D., Matus I., Rossi C. *et al.* Genetic diversity for quantitatively inherited agronomic and malting quality traits. In: R. von Bothmer, Th. Van Hintum, H. Knüpffer, K. Sato (eds) Diversity in Barley (*Hordeum vulagre*), Elsevier Science B.V., Amsterdam, The Netherlands, 2003, pp. 201-226

Helentjaris T., Slocum M., Wright S., Schaefer A., Nienhuis J. Construction of genetic linkage maps in maize and tomato using restriction fragment length polymorphisms. Theor Appl Genet 1986; 72: 761-769

Jena K.K., Pasalu I.C., Rao Y.K., Varalaxmi Y., Krishnaiah K., Khush G.S., Kochert G. Molecular tagging of a gene for resistance to brown planthopper in rice (*Oryza sativa* L.). Euphytica 2003; 129: 81-88

Johnson R. "Marker-assisted selection." In: J. Janick (ed) *Long Term Selection: Crops, Animals, and Bacteria.* Plant Breeding Reviews 2003; volume 24, part 2, in press

Katiyar S., Verulkar S., Chandel G., Zhang Y., Huang B., Bennett J. Genetic analysis and pyramiding of two gall midge resistance genes (*Gm-2* and *Gm-6t*) in rice (*Oryza sativa* L.). Euphytica 2001; 122: 327-334

Koebner R., Summers R. 21st century wheat breeding: selection in plots or detection in plates? Trends in Biotech 2003; 21: 59-63

Koebner R.M.D., Powell W., Donini P. The contribution of current and forthcoming DNA molecular marker technologies to wheat and barley genetics and breeding. In: J. Janick (ed) Plant Breeding Reviews 2001; 21 pp. 181-220

Lindstrom E.W. A genetic linkage between size and color factors in tomato. Science 1924; 60: 182-183

Narayanan N.N., Baisakh N., Cruz C.M.V., Gnanamanickam S.S., Datta K., Datta S.K. Molecular breeding for the development of blast and bacterial blight resistance in rice cv. IR50. Crop Sci 2002; 42: 2072-2079

Payne F. The effect of artificial selection on bristle number in *Drosophila ampelophila* and its interpretation. Proc Natl Acad Sci USA 1918; 4: 55-58.

Payne P.I., Nightingale M.A., Krattiger A.F., Holt L.M. The relationship between HMW glutenin subunit composition and the breadmaking quality of British-grown wheat varieties. Jour Sci Food Agric 1987; 40: 51-65

Ramalingam J., Basharat H.S., Zhang G. STS and microsatellite marker-assisted selection for bacterial blight resistance and waxy genes in rice, *Oryza sativa* L. Euphytica 2002; 127: 255-260

Ramsay L., Macaulay M., Ivanissevich S.D., MacLean K., Cardle L., Fuller J., Edwards K.J., Tuvesson S., Morgante M., Massari A., *et al.* A simple sequence repeat-based linkage map of barley. Genetics 2000; 156: 1997-2005

Rogers W.J., Payne P.I., Harinder K. The HMW glutenin subunit and gliadin composition of German-grown wheat varieties and their relationship with bread making quality. Plant Breed 1989; 103: 89-100

Roses A.D. Pharmacogenetics place in modern medical science and practice. Life Sci 2002; 70: 1471-1480

Sanchez A.C., Brar D.S., Huang N., Li Z., Khush G.S. Sequence tagged site marker-assisted selection for three bacterial blight resistance genes in rice. Crop Sci 2000; 40: 792-797

Sax K. The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. Genetics 1923; 8: 552-560

Shen L., Courtois B., McNally K.L., Robin S., Li Z . Evaluation of near-isogenic lines of rice introgressed with QTLs for root depth through marker-aided selection Theor Appl Genet 2001; 103: 75-83

Singh S., Sidhu J.S., Huang N., Vikal Y., Li Z., Brar D.S., Dhaliwal H.S., Khush G.S. Pyramiding three bacterial blight resistance genes (*xa5*, *xa13* and *Xa21*) using marker-assisted selection into *indica* rice cultivar PR106. Theor Appl Genet 2001; 102: 1011-1015

Song W.Y., Wang G.L., Chen L.L., Kim H.S., Pi L.Y., Holsten T., Gardner J., Wang B., Zhai W.X., Zhu L.H., Fauquet C., Ronald P.A. Receptor kinase-like protein encoded by the rice disease resistance gene, *Xa21*. Science 1995; 270: 1804-1806

Stuber C.W., Goodman M.M., Moll R.H. Improvement in yield and ear number resulting from selection at allozyme loci in a maize population. Crop Sci 1982; 22: 737-740.

Stuber C.W., Moll R.H., Goodman M.M., Shaffer H.E., Weir B.S. Allozyme frequency changes associated with selection for increased grain yield in maize. Genetics 1980; 95: 225-236

Summers R., Koebner R.M.D., Hollins T.W., Förster J., Macartney D.P. The use of an isozyme marker in breeding wheat (*Triticum aestivum*) resistant to the eyespot pathogen (*Pseudocercosporella herpotrichoides*). In: T.E. Miller and R.M.D. Koebner (eds.) *Proc. 7th Intern Wheat Genet Symp* , Institute of Plant Science Research, Cambridge, UK, 1988, pp. 1195-1197

Thomas W.T.B. Prospects for molecular breeding of barley. Ann Appl Biol 2003; 142: 1-12

Toenniessen G.H., O'Toole J.C., DeVries J. Advances in plant biotechnology and its adoption in developing countries. Curr Opin Plant Biol 2003; 6: 191-198

Toojinda T., Baird E., Booth A., Broers L., Hayes P., Powell W., Thomas W., Vivar H., Young G. Introgression of quantitative trait loci (QTLs) determining stripe rust resistance in barley: an example of marker-assisted line development. Theor Appl Genet 1998; 96: 123-131

Tuvesson S., von Post L., Ohlund R., Hagberg P., Graner A., Svitashev S., Schehr M., Elovsson R. Molecular breeding for the BaMMV/BaYMV resistance gene *ym4* in winter barley. Plant Breed 1998; 117: 19-22

Xu X.F., Mei H.W., Luo L.J., Cheng X.N., Li Z.K. RFLP-facilitated investigation of the quantitative resistance of rice to brown planthopper (*Nilaparvata lugens*). Theor Appl Genet 2002; 104: 248-253

Young N.D. A cautiously optimistic vision for marker-assisted selection. Mol Breed 1999; 5: 505-510.

Zhou P.H., Tan Y.F., He Y.Q., Xu C.G., Zhang Q. Simultaneous improvement for four quality traits of Zhenshan 97, an elite parent of hybrid rice, by molecular marker-assisted selection. Theor Appl Genet 2003; 106: 326-331

Chapter 11

MAP-BASED GENE ISOLATION IN CEREAL GENOMES

Nils Stein* and Andreas Graner

Institute of Plant Genetics and Crop Plant Research (IPK), Corrensstrasse 3, 06466 Gatersleben, Germany
*Author for correspondence: stein@ipk-gatersleben.de

1. INTRODUCTION

Gene isolation can be followed by two different strategies, the '*reverse genetics*' (from gene to phenotype) and the '*forward genetics*' approach (from phenotype to gene) (Takahashi *et al.*, 1994). In reverse genetics, we do not know the phenotype that is influenced by the gene of interest, although it is characterised by a genomic DNA sequence, mRNA/cDNA, or a protein sequence. The gene can be directly used for transformation studies to elucidate its function and a possibly caused phenotype. On the other hand, a mutant phenotype may be induced by disrupting the gene function by either insertional mutagenesis like T-DNA/transposon-tagging or gene silencing via antisense RNA, cosuppression or RNA interference (RNAi) (Matzke *et al.*, 2001). In forward genetics, a mutant phenotype for the trait of interest (obtained by mutagenesis or from natural variation) is identified first and the mutated gene, along with a large number of markers, is genetically mapped; this map is then used for gene isolation. This forward genetics approach of gene isolation is described as map based cloning (MBC) (Peters *et al.*, 2003). In crop plants, genes underlying important agronomic traits are largely unknown and are only characterized by a superior/inferior phenotype. Therefore, MBC is of increasing relevance for current and future crop improvement programmes. For an understanding of the genetics and molecular basis of crop-specific traits and to allow efficient manipulation of the corresponding crop-specific genes for breeding, it may be necessary to isolate and characterize these

P.K. Gupta and R.K. Varshney (eds.), Cereal Genomics, 331-360.

genes from individual crops. This is specifically important, since a gene for a specific trait isolated from a model species or one crop may not be suitable for another crop as is particularly true for rapidly evolving disease resistance genes (R genes). As an example, R genes may remain non-functional if transferred between species of different plant families (‘*restricted taxonomic functionality*’) (reviewed in McDowell and Woffenden, 2003) thus excluding Arabidopsis as a major source of R genes for crops other than Brassicaceae.

The approach for isolation of genes that is solely based on their map location (MBC) can be used for any gene, whose phenotypic effect can be followed in populations segregating for alternative variants of the gene. Although, widely used in human and animal systems, the use of MBC for gene isolation in plant systems started only about ten years ago (Gibson and Somerville, 1993; Leyser and Chang, 1996). After the whole sequencing of the Arabidopsis genome (TAGI, 2000) and after major improvements in molecular marker technology, positional cloning became more or less routine in Arabidopsis (for review see: Jander *et al.*, 2002; Peters *et al.*, 2003). However, if applied to other plant systems in the past, the major limitation with the use of this approach was the non-availability of necessary tools (e.g. large insert genomic libraries) and also the size and nature of the target genome (e.g. amount of repetitive DNA, target locus lying in low recombinogenic region). This is particularly true for cereal genomes, although in cereals also, MBC has been successfully used for isolation of a number of genes (Table 1). In the present chapter, an attempt will be made, (i) to highlight different aspects of MBC in cereals; (ii) to summarize importance and success of MBC in cereals with emphasis on the tribe Triticeae, and finally (iii) to outline the impact of this research on the isolation of quantitative trait loci (QTL).

2. MAP BASED CLONING (MBC): THE BASIC TECHNOLOGY

Several reviews are available, where the basic principle and strategy of MBC for plant genes has been discussed (Leyser and Chang, 1996; Jander *et al.*, 2002; Peters *et al.*, 2003). It basically involves the following three steps: (i) initial genetic mapping and fine mapping of the target gene, (ii) physical mapping, identification and isolation of the candidate gene, and (iii) proof of biological function of the candidate gene via transformation or mutant analysis (Fig. 1). In the following, these three major steps will be described in more detail with emphasis on cereal specific aspects.

Table 1. Genes isolated from some cereal species via map-based cloning

Species	Gene	Function	Reference
Rice[1]			
O. Sativa	*Xa21*	receptor protein kinase	Song *et al.* (1995)
O. sativa	*Xa1*	NBS-LRR	Yoshimura *et al.* (1998)
O. sativa	*Pib*	NBS-LRR	Wang *et al.* (1999)
O. sativa	*d1*	α–subunit of GTP-binding protein	Ashikari *et al.* (1999)
O. sativa	*Pi-ta*	NBS-LRR	Bryan *et al.* (2000)
Wheat (2*x*)			
T. monococcum	*Vrn1*	homology to MADS-box proteins	Yan *et al.* (2000)
Wheat (6*x*)			
T. aestivum	*Lr10*	CC-NBS-LRR	Stein *et al. (2*000), Feuillet *et al. (*2003)
T. aestivum	*Lr21*	NBS-LRR	Huang *et al.* (2003a)
T. aestivum	*Pm3*	CC-NBS-LRR	Yahiaoui *et al.* (2003)
T. aestivum	*Q*[2]	*APETALA* (*AP2*)-like	Faris *et al.* (2003)
Barley			
H. vulgare	*mlo*	seven-transmembrane protein, homology to G-protein coupled receptors	Buschges *et al.* (1997), Devoto *et al.* (2003)
H. vulgare	*rar1*	zinc-binding protein (CHORD domain)	Lahaye *et al.* (1998), Shirasu *et al.* (1999)
H. vulgare	*ror2*[3]	syntaxin	Collins *et al.* (2003)
H. vulgare	*Mla1*[4] *Mla6* *Mla7* *Mla10* *Mla12* *Mla13*	CC-NBS-LRR	Wei et al. (1999, 2002), Halterman et al. (2001, 2003), Zhou et al. (2004), Shen et al. (2003), Halterman and Wise (2004)
H. vulgare	*Rpg1*	receptor kinase	Brueggeman *et al.* (2002)

[1]the displayed rice genes represent a selection and serve as examples of genes isolated via MBC from rice, [2]best candidate on a contig co-segregating with the Q-locus - functional proof is pending, [3]synteny-based positional cloning, no chromosome walking was necessary in barley, [4]isolation of the Mla locus by chromosome walking provided the necessary information for homology-based isolation of diverse alleles of the same gene.

However, it has to be noted, that individual steps of the strategy can be highly reduced or even omitted depending on the availability of resources for the respective species. This will be discussed further in context of cereal examples later in this chapter.

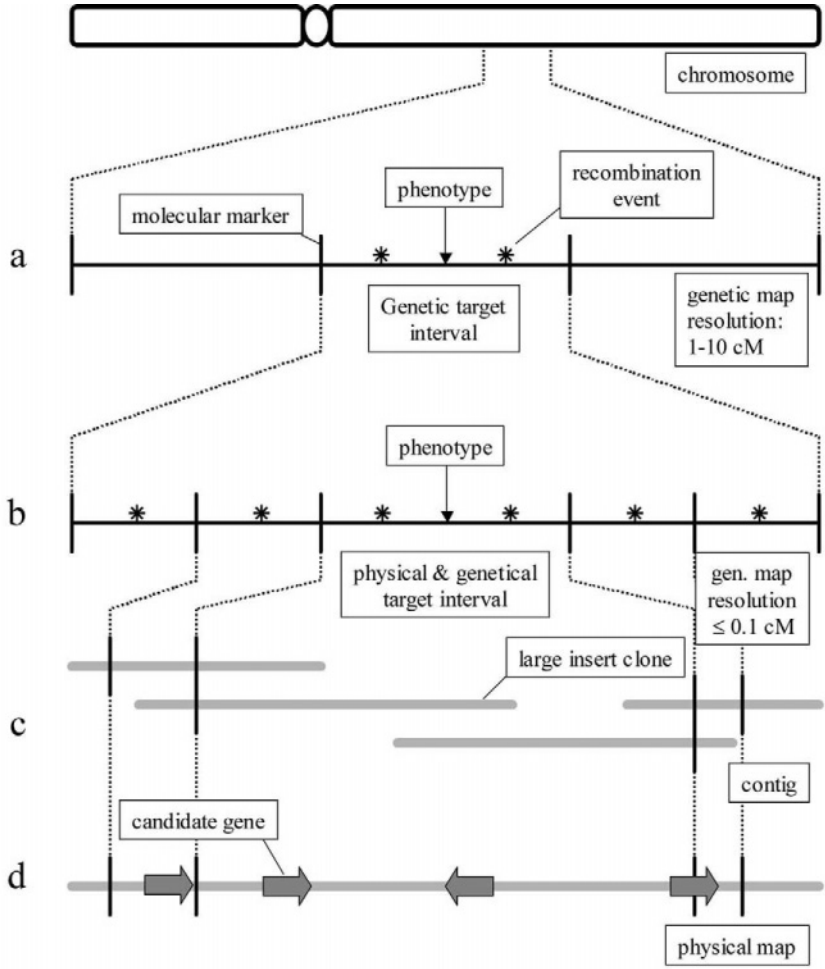


Figure 1. Scheme of map-based gene isolation. a) The initial step in a map-based cloning project is the rough genetic localization of the target gene either by means of morphological or molecular markers and genetic segregation analysis. This usually provides a genetic resolution around the target gene (phenotype) of 1 – 10 cM which is not sufficient for determination of markers linked closely enough to be used as starting points for chromosome walking. b) In the following step a large number of meiotic events has to be analysed for selection of plants that exhibit a recombination event in close proximity to the gene of interest to achieve a genetic resolution of less than 0.1 cM. Marker saturation of the genetic target interval has to be achieved to provide a suitable and promising starting point for the first step of chromosome walking. c) A chromosome walk is usually initiated by screening a large insert genomic library of the target species with a marker that is genetically very closely flanking or cosegregating with the gene of interest. A first anchor clone will be identified. This is then further exploited as a source of new molecular markers that serve to orientate the clone according to the genetic map. New markers have to be developed from this initial clone to screen the library for the extension of the physical map towards the gene. In an unpredictable number of repeated rounds of these steps, finally, a contiguous overlapping series of clones will be identified that exhibit both closest flanking markers separated by recombination breakpoints from the gene. d) This physical map is then used for candidate gene identification. The biological function of candidate genes has to be assayed either by complementation analysis via stable or transient transformation of the candidate gene into wild-type background or by the characterization of independent mutants in the candidate gene.

In the first step of MBC, initially the gene has to be mapped through segregation analysis using morphological or molecular markers (Fig. 1a), and this is followed by high resolution mapping to locate the position of

the gene more precisely. Therefore, a large segregating population (typically the equivalent of 2000 F_2 plants or greater) has to be screened for meiotic recombination events in close proximity of the gene. The target region is thus saturated with additional molecular markers (e.g. RAPD, AFLP), often using bulked-segregant analysis (Michelmore *et al.*, 1991) (Fig.1 b). This should allow physical delimitation of the gene within one large- insert clone using flanking markers. Such a strategy was called '*chromosome landing*' and is especially useful for species with a large genome to avoid labour-intensive chromosome walking across long distances of repetitive DNA (Tanksley *et al.*, 1995). The population size required to achieve this goal also depends strongly on the ratio between physical and genetic distances in the respective target interval. Unfortunately, there is an uneven distribution of recombination frequencies over the grass genome (Kunzel *et al.*, 2000; Akhunov *et al.*, 2003) and local recombination rates may easily vary up to tenfold (Stein *et al*., 2000). Some genomic regions may even lack recombination over long physical distances and therefore may not be amenable to map-based cloning (Qi and Gill, 2001; Neu *et al.*, 2002). Recombination frequencies may also be dependent on the parental combination of the segregating population. Therefore it is recommended not to rely on a single mapping population in map-based cloning attempts in large-genome cereals.

The genetic map, constructed in the initial step of gene isolation as above, serves as the navigation support for tracking down the gene of interest (Fig. 1c). After high-resolution genetic mapping and marker saturation, the closest marker to the gene will be used to screen a large insert genomic library to identify a clone that may carry the gene. Initially, these libraries were based on the yeast artificial chromosome (YAC) vector system (Burke *et al.*, 1987), but more recently bacterial artificial chromosomes (BAC, Shizuya *et al.*, 1992) have been used due to the ease of handling and less sensitivity of clone inserts to recombination events and chimaerism (Zhang *et al.*, 1996). New DNA probes have to be derived from the large genomic clone that carries the mapped marker closest to the gene of interest. This can be achieved by strategies like direct BAC end sequencing (Kelley *et al.*, 1999; Zhao *et al.*, 2001), subcloning (plasmid rescue) of BAC ends (Woo *et al.*, 1994; Ripoll, 2000), representational difference analysis (RDA, Schutte *et al.*, 1995) or random subcloning and low-pass shotgun sequencing for the identification of low-copy DNA sequences (Nurminsky and Hartl, 1996; Stein *et al.*, 2000). Furthermore, hybridisation of genomic DNA to colony filters of random subclones or fingerprint blots of candidate BACs helps to identify low-copy DNA sequences. These low-copy DNA sequences will allow the development of new DNA probes for genetic mapping and anchoring of the physical map to the high-resolution

genetic map. They will also be used for re-screening of the library leading to the identification of new clones that extend the physical map (Fig. 1c). Altogether, this iterative and stepwise procedure of clone identification, development of markers, genetic mapping, and again clone identification is called '*chromosome walking*'.

The newly identified clones have to be aligned to the original starting BAC or an already established contig. In small scale, this is easiest achieved by Southern hybridization (utilizing the original screening probe, e.g. BAC end, random subclone) to obtain restriction fingerprints of the candidate BACs. Overlapping clones may be identified through common fragments, and fragment confirmation will be achieved through comparative sequencing of the amplicons of the screening probe from all candidate clones. Ambiguous overlaps may also be sorted out with the help of reciprocal hybridisation with candidate overlapping BACs to fingerprint blots. This will allow identification of all overlapping fragments.

In principle, methods generally used for the production of BAC physical maps/contigs for large-scale sequencing projects (see Chapter 13 by Yu and Wing in this book) can also be used for small-scale physical mapping needed for MBC. These techniques include, high throughput fluorescent restriction fingerprinting of BAC clones (e.g. Ding *et al.*, 1999) in combination with computer-supported data analysis and assembly [fingerprinted contig (FPC) analysis: Soderlund *et al.*, 1997], which allows automated identification of a maximum number of restriction fragments shared by two individual clones. Finally, a contig will be extended across the gene from a genetically proximal to a distal flanking marker and a physical map of the target region will be available (Fig. 1d). This contig is supposed to carry the target gene, unless the desirable allele of the gene was absent in the species used for constructing the BAC library. This may be of particular relevance for rapidly evolving genes including those coding for disease resistance. As an example, a candidate gene for the leaf rust resistance locus *Lr10* was absent in most susceptible wheat cultivars (Stein *et al.*, 2000; Scherrer *et al.*, 2002; C. Feuillet, personal communication). Similarly, recent results from comparative analysis of orthologous physical regions among different elite maize inbred cultivars revealed differences in the overall gene content and microcollinearity (Fu and Dooner, 2002). Therefore, physical mapping should preferably be done in a genotype that is known to contain the desirable allele of the gene of interest. This aspect is documented by the wealth of rice BAC libraries that have been constructed for the sole purpose of gene isolation (i.e. Nakamura *et al.*, 1997; Peng *et al.*, 1998; Wang *et al.*, 2001; and Table 2).

Table 2. Genome size, structure and genomic resources of some model and crop plant species

Crop Species	Ploidy	Genome size[1] (bp)	Repetitive DNA[2]	Genomic sequence available	Physical map (contig-based) available	Large insert library available	EST sequences[3] available in public domain
A. thaliana	2	0.15 x 10^9	10%[4]	+	+	+	204,396
O. sativa	2	0.4 x 10^9	15-25%[5]	+	+	++	283,935
S. bicolor	2	0.8 x 10^9	-	-	(+)[6]	+	161,813
P. glaucum	2	2.5 x 10^9	-	-	-	+	2,528
Z. mays	2	3.0 x 10^9	78%	(+)[7]	(+)[8]	+	393,719
H. vulgare	2	5.5 x 10^9	76%	-	-	+	356,848
S. cereale	2	9.4 x 10^9	92%	-	-	-	9,194
A. sativa	6	11.7 x 10^9	83%	(?)[9]	-	-	574
T. aestivum	6	17.9 x 10^9	83%	(?)[10]	-	+	549,926

[1]Bennett and Leitch (1995) except for *P.g.*: Bennett and Smith (1976), [2] according to Flavell *et al.* (1974) if not otherwise indicated, [3]dbEST release 030504, http://www.ncbi.nlm.nih.gov/dbEST/dbEST summary.html, [4]Leutweiler *et al.* (1984), [5]Bennetzen (2000); Sasaki *et al.* (2002); Feng *et al.* (2002), [6]construction of a genome-wide physical map is in progress, Draye *et al.* (2001); detailed comparison of syntenic physical map of sorghum chromosome 3 and rice chromosome 1, Klein *et al.* (2003), [7]a maize genome sequencing project has been launched by the US National Science Foundation in September, 2002, Chandler and Brendel (2002), [8]Cone *et al.* (2002), [9]Bernal *et al.* (2001), [10]sequencing of the gene-rich regions of the wheat genome is in the planning phase: IGROW, http://www.aspb.org/publicaffairs/stakeholders/igrow.pdf

It is important to note that the first two steps of a map-based cloning strategy may be followed in parallel instead of one-after-the-other. Against the original suggestion of starting a chromosome walk only after two closely flanking markers have been identified, it can be advantageous to start the step of physical mapping already as soon as a '*good*' marker (i.e. cosegregating or closely linked to the gene in a reasonably sized population) could be identified. In the context of uneven distribution of recombination in the plant genome, the effort of increasing the population size for high-resolution mapping should always be seen in comparison to the expected effort of starting a chromosome walk. This is especially important for genes conferring a phenotype that is either difficult to score at an early stage of the plant development or can be scored only under a laborious experimental setup.

The final step of MBC involves identification of a candidate gene and the proof of its biological function. Different approaches may be followed to reach this goal. One possibility is to use the large insert clone, which presumably carries the gene, for direct complementation analysis via transformation. A special binary BAC (BIBAC) vector system has been introduced for transfer of large DNA fragments into plants (Hamilton, 1997). This strategy may provide evidence that the candidate gene is present in the contig. However, the gene sequence will have to be identified in subsequent cloning steps. Alternatively, the cosegregating contig can be sequenced to low coverage or full length in order to identify open reading frames (ORF). Candidate genes may be identified by functional annotation based on sequence comparison to protein sequence databases and finally used for transformation. A third strategy is the mutagenesis of the target gene by chemical mutagens, mutagenic irradiation or T-DNA/transposon insertion. Sequence analysis of the candidate gene in independent mutants will reveal presence of functional mutations. A genome-wide approach for the targeted screening of induced mutations in a genome has also been suggested (TILLING, McCallum *et al.*, 2000) and international efforts for the establishment of TILLING-panels in different crop species are underway that may be efficiently exploited in future for the purpose of map-based gene isolation.

3. MAP-BASED GENE ISOLATION IN CEREALS: ISSUES OF GENOME STRUCTURE AND SIZE

All cereal species belong to the grass family (Poaceae) but they are split into the subfamilies, *Oryzoideae* (rice), *Pooideae* (barley, wheat, rye, oats)

and *Panicoideae* (maize, sorghum, pearl and foxtail millet) (Strasburger, 1983). Although, they share a significant level of genome conservation in terms of gene-order and gene sequences, they are quite diverse concerning their genome size, and in the composition of genes and different classes of repetitive elements (Table 2) (for review see: Bennetzen and Freeling, 1993; Moore *et al.*, 1995; Keller and Feuillet, 2000; Laurie and Devos, 2002, and Chapter 5 by Paterson in this book). Estimates based on gene prediction after complete sequencing of rice chromosomes 1 and 4 suggest around 60,000 genes for the whole rice genome (Feng *et al.*, 2002; Sasaki *et al.*, 2002). In contrast, the 12 times larger barley genome seems to carry a similar number of 40,000 to 70,000 genes, as estimated on the basis of available ESTs and genomic sequences (Zhang *et al.*, unpublished data). Compared to rice, the massive genome expansion in other cereals with large genomes is probably the consequence of multiplication and nested insertion of transposons, as has been repeatedly demonstrated by detailed and comparative sequence analysis (i.e. SanMiguel *et al.*, 1996; SanMiguel and Bennetzen, 1998; Dubcovsky *et al.*, 2001; Wicker *et al.*, 2001). These differences in the make up of their genomes led to two major conclusions about MBC in cereals: (i) the abundance of repetitive sequences in complex plant genomes hampers chromosome walking, and (ii) cereal species each with a small genome contain transcriptomes of similar size, compared to cereals with larger genomes; due to their genome collinearity, a small genome cereal (e.g; rice) therefore may serve as a model for map-based gene isolation in large genome cereals (Bennetzen and Freeling, 1993). However, macrocollinearity (collinearity on the genetic map-level) seems to be better preserved than collinearity at the sequence level (microcollinearity) (see Chapter 19 by Sorrells in this book). As a result, in cereals, genomics tools and resources required for map-based gene isolation were developed on the basis of their economic importance, but in inverse proportion to the genome complexity of the corresponding species (Table 2). In other words – for map-based gene isolation, resources are far advanced in rice, they are improving for maize, they have yet to be improved for wheat and barley and are largely non-existent for rye and oats.

3.1. Model Genomes for Collinearity-based Gene Isolationin Cereals

Based on the model of genome collinearity, orthologous genes should share the same physical order in different cereal genomes, although they would be separated by long stretches of repetitive DNA in species with large genomes. In order to exploit collinearity, the initial step for MBC in a

cereal with a large genome would be the localization of the putatively orthologous region in the rice genome. As soon as the flanking markers for the gene of interest are located in the target as well as in the model genome, a contig of the model genome region can be established and sequenced, so that candidate genes may be identified from the model species (collinearity-based candidate gene approach). Alternatively, a model genome contig can be the source of new molecular markers that will be employed to scan the large-insert library of the target genome, and thus help to bridge long physical distances in the target genome (collinearity-based chromosome walking/landing approach).

Rice is the model genome for collinearity-based gene isolation in cereals (see also Chapter 18 by Sasaki and Antonio in this book). In this species, map-based gene isolation is simplified due to the availability of a genome-wide physical map (Saji *et al.*, 2001, physical map explorer at: http://rgp.dna.affrc.go.jp/giot/INE.html or http://www.gramene.org/). Identification of a candidate gene can be initiated immediately as soon as a collinear marker interval has been defined by comparative mapping in a target species (e.g. wheat, barley). For instance, collinearity between the barley genomic regions containing the stem rust resistance genes *Rpg1* and *rpg4* and their syntenic regions in rice encouraged the use of rice as an '*intergenomic cloning vehicle*' (Kilian *et al.*, 1997). EST resources (Table 2) can also be exploited efficiently for marker saturation of the target region. For example, during a comparative mapping study of the barley leaf rust resistance locus *Rph16*, as many as 30 closely linked ESTs were selected on the basis of sequence similarity to rice open reading frames (ORF) from a putatively orthologous rice BAC/PAC contig (Perovic *et al.*, unpublished data). In due course of time, this should facilitate isolation of barley gene *rph16*.

The candidate gene approach may sometimes also fail, as suggested by the reports of variation in the content and order of orthologous genomic sequences from several cereal species (for review: Bennetzen, 2000; Bennetzen and Ramakrishna, 2002; Feuillet and Keller, 2002). In particular, for the identification of candidate genes for race-specific disease resistance loci, which are less conserved between species and prone to genomic rearrangements (Leister *et al.*, 1998), the earlier optimism regarding the use of the model genome strategy diminished recently (Brueggeman *et al.*, 2002; Bennetzen, 2000; Bennetzen and Ma, 2003). Nevertheless, the candidate gene approach has been successful in cloning genes involved in basic processes. For instance, the approach has been successfully utilized for the isolation of wheat gene *Vrn1* (response to

vernalisation, transcription factor, Yan *et al.*, 2003) and the barley gene *ror2* (basal penetration resistance, syntaxin, Collins *et al.*, 2003).

Models other than the rice genome have also been suggested for collinearity-based gene isolation, since there is increasing evidence for inconsistency of macro- and microcollinearity between rice and other cereal genomes. For instance, due to its very similar genome organisation, sorghum (genome size sorghum/rice = 2:1) has been found to be a system, which is comparable to rice and is regarded as ultimate model for collinearity based gene isolation in the Panicoideae, especially for maize (Draye *et al.*, 2001). Therefore, major efforts are being made to build up a physical map of sorghum leading already to a detailed alignment of the physical maps of sorghum chromosome 3 and rice chromosome 1 (Klein *et al.*, 2003). This will provide a fast forward tool towards the identification of orthologous candidate genes for important traits in maize. Another tiny grass *Brachypodium distachyon*, which is closely related but does not belong to the tribe Triticeae, possesses several desirable attributes including a small genome size comparable to Arabidopsis, short duration (less than 4 months) and amenability to tissue culture and transformation (Draper *et al.*, 2001). A BAC library for the species seems to be available and accessible to the public (Moore, 2003), but it has not been extensively used so far. Diploid and tetraploid wheat may also be used to circumvent the difficulties that the polyploid nature and large proportion of repetitive DNA poses while cloning genes from hexaploid bread wheat (*Triticum aestivum*; Stein *et al.*, 2000; also see later in this chapter).

3.2. Progress and Success in Map-Based Gene Isolation in Cereals

3.2.1. Rice

Map-based cloning for gene isolation in rice has now become routine due to the availability of whole genome sequences and other genomics resources (see Chapter 13) (Goff *et al.*, 2002; Yu *et al.*, 2001). *In silico* data-mining of this information will allow identification of DNA polymorphisms at almost any location of the rice genome, providing an ultimate resource for marker development. High density transcript linkage maps (Wu *et al.*, 2002), large numbers of EST sequences (Table 2), a comprehensive set of full length cDNA sequences (Kikuchi *et al.*, 2003) and numerous large insert libraries (Table 3) are already in place or will soon become available to the public. These will provide novel

Table 3. Large insert libraries of cereal crops

Species[1]	Accession	Number of clones	Average insert size	Genome coverage	Library type	Cloning site	Reference[2]
Rice							
O. sativa	Nipponbare	6,932	350 kb	6.0 x	YAC	E*co*RI, H*ind*III	Umehara *et al.* (1995)
O. sativa	IRBB56	55,296	132 kb	14.0 x	BAC	H*ind*III	Wang *et al.* (1995)
O. sativa	Lemont	7,296	150 kb	2.6 x	BAC	H*ind*III	Zhang *et al.* (1996)
O. sativa	Teqing	14,208	130 kb	4.4 x	BAC	H*ind*III	Zhang *et al.* (1996)
O. sativa	Shimokita	21,504	155 kb	7.0 x	BAC	H*ind*III	Nakamura *et al.* (1997)
O. sativa	Minghui63	26,000	150 kb	9.0 x	BAC	H*ind*III	Peng *et al.* (1998)
O. sativa	Kasalath	65,000	106 kb	16.0 x	BAC	H*ind*III	Miyazaki and Saito (1999)
O. sativa	Nipponbare	71,040	112 kb	18.0 x	PAC	S*au*3AI	Baba *et al.* (2000)
O. sativa	Kasalath	47,232	133 kb	14.6 x	BAC	M*bo*I	Baba *et al.* (2000)
O. sativa	Nipponbare	23,040	130 kb	6.7 x	BIBAC	H*ind*III	Tao *et al.* (2002)
O. sativa	Nipponbare	23,040	170 kb	8.7 x	BAC	B*am*HI	Tao *et al.* (2002)
O. sativa	Nipponbare	23,040	150 kb	7.7 x	BAC	H*ind*III	Tao *et al.* (2002)
O. sativa	Nipponbare	23,040	156 kb	8.0 x	BAC	E*co*RI	Tao *et al.* (2002)
Sorghum							
S. bicolor	BT×623	13,440	157 kb	2.6 x	BAC	H*ind*III	Woo *et al.* (1994)
S. bicolor	BT×623	12,576	140 kb	2.2 x	BAC	E*co*RI	Klein *et al.* (2000)
S. propinquum	n.n.	38,016	126 kb	6.6 x	BAC	H*ind*III	Lin *et al.* (1999)
Millet							
P. glaucum	Tift 23DB	159,100	90 kb	4.7 x	BAC	B*am*HI	Allouis *et al.* (2001)

Table 3. Continued

Maize							
Z. mays	UE95	79,000	145 kb	3.0 x	YAC	E*co*RI	Edwards *et al*. (1992)
Z. mays	F2	86,858	90 kb	3.2 x	BAC	H*ind*III	O'Sullivan *et al*. (2001)
Z. mays	B73	247,680	136 kb	6.0 x	BAC	H*ind*III	Tomkins *et al*. (2002)
Z. mays	B73	131,200	160 kb	7.0 x	BAC	E*co*RI	Yim *et al*. (2002)
Z. mays	B73	125,700	167 kb	7.0 x	BAC	M*bo*I	Yim *et al*. (2002)
Z. mays	W22	20,000	106 kb	0.7 x	BAC	N*ot*I	Fu *et al*. (1999)
Barley							
H. vulgare	Franka	100,000	200 kb	4.0 x	YAC	E*co*RI, M*lu*I	Kleine *et al*. (1997), Schmidt *et al*. (2001)
H. vulgare	Morex	313,344	106 kb	6.3 x	BAC	H*ind*III	Yu *et al*. (2000)
H. vulgare	Haruna Nijo	294,912	115 kb	6.8 x	BAC	H*ind*III	Sato *et al*. (pers. comm.)
Wheat (2*x*, *A genome*)							
T. boeticum	Boiss	170,000	104 kb	3.0 x	BAC	B*am*HI	Chen *et al*. (2002)
T. monococcum	DV92	276,000	115 kb	5.6 x	BAC	H*ind*III	Lijavetzky *et al*. (1999)
Aegilops (2*x*, *D genome*)							
Ae. tauschii	Aus 18913	144,000	119 kb	3.7 x	BAC	H*ind*III	Moullet *et al*. (1999)
Ae. tauschii	AL8/78	54,000	167 kb	2.2 x	BAC	E*co*RI	Zhang *et al*. (unpublished)
Ae. tauschii	AL8/78	59,000	189 kb	2.2 x	BAC	H*ind*III	Zhang *et al*. (unpublished)
Ae. tauschii	AL8/78	52,000	190 kb	3.2 x	BAC	H*ind*III	Zhang *et al*. (unpublished)
Ae. tauschii	AL8/78	59,000	149 kb	2.8 x	BAC	B*am*HI	Zhang *et al*. (unpublished)
Ae. tauschii	AL8/78	76,000	174 kb	2.4 x	BAC	B*am*HI	Zhang *et al*. (unpublished)

Table 3. Continued

Wheat (4*x*)							
T. dicoccoides	Langdon	516,000	130 kb	5.0 x	BAC	H*ind*III	Cenci *et al*. (2003)
Wheat (6*x*)							
T. aestivum	Glenlea	656,640	80kb	3.1 x	BAC	B*am*HI, H*ind*III	Nilmalgoda *et al*. (2003)
T. aestivum	Chinese Spring	950,000	54 kb	3.1 x	BAC	H*ind*III	Ogihara *et al*. (unpublished)
T. aestivum	Renan	478,840	150 kb	3.2 x	BAC	H*ind*III	Chalhoub *et al*. (unpublished)
T. aestivum	Renan	285,312	132 kb	2.2 x	BAC	E*co*RI	Chalhoub *et al*. (unpublished)
T. aestivum	Renan	236,160	122 kb	1.5 x	BAC	B*am*HI	Chalhoub *et al*. (unpublished)

[1]The list displays only a selection of libraries for rice. For the other cereal species also, the excistence of additional libraries produced for specific gene isolation purposes is very likely,
[2]Information taken from http://agronomy.ucdavis.edu/Dubcovsky/BAC-library/ITMIbac/ITMIBAC.htm if not otherwise indicated

opportunities for the map-based gene isolation. Initially, race-specific disease resistance genes like *Xa21* (Song *et al.*, 1995), conferring resistance to bacterial blight, or *Pib* (Wang *et al.*, 1999), providing resistance to rice blast resistance, were being attempted (Table 1). This was a consequence of their economic importance and the relative simplicity of the genetic system. Recently, more genes have been isolated for additional traits including agronomic traits like culm length (*d1*, Ashikari *et al.*, 1999). Further target loci have been physically delimited and the identified candidate genes await verification of their biological function. In addition to monogenic traits, the far advanced genomics resources in rice allow for the map-based isolation of QTLs also (see later).

3.2.2. Maize, Sorghum, Millets

Until recently, MBC was rarely used for gene isolation in maize, sorghum and millets. In maize, there has been no real need for MBC since gene isolation relied almost exclusively on procedures that utilize any of the well characterized endogenous transposon tagging systems (for review see Maes *et al.*, 1999; Walbot, 2000). No gene has been isolated so far from sorghum and other millets through MBC. However, a few attempts towards the sorghum genes *rf1* and *rf4* (restoration of fertility) are in progress (Klein *et al.*, 2001; Wen *et al.*, 2002). The situation is likely to change in the future since major efforts are underway to establish comprehensive genomics tools in maize and sorghum. For both these species full genome physical maps are under construction (Draye *et al.*, 2001; Cone *et al.*, 2002), and high-density genetic maps are already available (Coe *et al.*, 2002; Menz *et al.*, 2002). In both these crops, several BAC libraries are also now available (Table 3), and in maize a sequencing project for the gene rich regions (GRRs) of the genome has been initiated (Chandler and Brendel, 2002). These resources will be extensively utilized in future for the map-based cloning of QTLs (see later).

3.2.3. Oats

Oats, due to its hexaploid genome and its limited economic importance, is unlikely to be subjected to MBC in the near future. For instance, no genomics resources are available, except the more or less comprehensive genetic maps (see Chapter 3 by Varshney *et al*. in this book), and no BAC library for oats has been published or announced so far.

3.2.4. Triticeae

Map-based cloning may be used for gene isolation in the Triticeae species due to non-availability of other alternatives (e.g. T-DNA or transposon-tagging), although large genome size and high proportion of repetitive DNA discourage the use of this approach. Only recently the development of transposon-tagging using the maize Ac/Ds transposon system has been initiated in wheat and barley (Takumi *et al.*, 1999; Koprek *et al.*, 2000; Scholz *et al.*, 2001). However, it has not reached the stage of routine application, since it is restricted to a few genotypes with superior transformation properties. The development of the tagging populations has also not yet been completed. In contrast, for map-based cloning, new resources like numerous Triticeae YAC and BAC libraries became available (Table 3). Furthermore, the success in the map-based isolation of the *mlo* resistance gene (Buschges *et al.*, 1997) paved the way for further gene isolations from barley and wheat; more than a dozen genes in these two crops have been isolated using the approach of MBC (Table 1). However, almost each of these genes has been isolated in a slightly different manner reflecting the whole spectrum of possible variations of the MBC strategy. In the following, some of the examples for these different strategies will be discussed in more detail.

3.2.4.1. Chromosome landing

In barley, *mlo* and *rar1* (Buschges *et al.*, 1997; Shirasu *et al.*, 1999), were first isolated via the suggested strategy of chromosome landing (Tanksley *et al.*, 1995). Large F_2 progenies of 4000 gametes for *mlo* and 8000 gametes for *rar1*, respectively, were screened for recombination events within the target interval and marker saturation was achieved by AFLP (Vos *et al.*, 1995) in combination with bulked segregant analysis (BSA) employing susceptible and resistant pools of recombinant plants. Cosegregating markers were used to screen a barley YAC library and within a single step a physical map of the target locus could be constructed. In case of *mlo*, a single large YAC was found to span the whole target region including both flanking markers, whereas for *rar1* a contig of five overlapping YACs was established. Despite final success of gene isolation in both attempts, the problems of chromosome walking with YACs became obvious by rearrangements and chimaerism in four out of five of the YACs constituting the *rar1* contig (Lahaye *et al.*, 1998).

3.2.4.2. Chromosome walking and homology-based gene isolation

In barley, the complex powdery mildew resistance locus *Mla* was isolated from barley by a classical chromosome walking approach (Wei *et al.*,

1999, 2002). Three steps of walking were necessary to obtain a mixed contig of YAC and BAC clones that covered the locus. High-resolution mapping was performed in a population representing 3600 gametes and marker saturation was achieved by BSA utilizing a variety of marker sytems (RFLP, RAPD, AFLP, S-SAP). The final steps of chromosome walking were conducted in a BAC library derived from the susceptible barley cultivar Morex. It contained a cluster of eleven members of three families of related resistance gene homologues, *RGH1, RGH2* and *RGH3*. The sequence information from the susceptible cultivar facilitated the homology-based isolation of the *Mla1, Mla6, Mla12,* and *Mla13* genes from cosmid libraries of the corresponding resistant barley varieties (Halterman *et al.*, 2001; Zhou *et al.*, 2001; Halterman *et al.*, 2003; Shen *et al.*, 2003). Comparative sequence analysis revealed that these genes, as well as two additional specificities, *Mla7* and *Mla10,* are likely to represent alleles of only one member of the *Mla-RGH1*-family (Wei *et al.*, 2002; D. Halterman and R. Wise, USA, personal communication). Generally, the situation may occur, where an allele of interest is not present in the cultivar used to construct the large-insert library. In this case, chromosome walking in the available BAC library constructed from the susceptible cultivar provided the necessary information for homology-based gene isolation from resistant cultivars.

3.2.4.3. Collinearity-based marker saturation and chromosome walking in the target genome

The exploitation of collinearity between the genome of the crop of interest and that of the model cereal species has been suggested as a shortcut for efficient gene isolation. For instance, barley genomic regions containing two stem rust resistance genes *Rpg1* and *rpg4* (Brueggeman *et al.*, 2002; Druka *et al.*, 2000) and a leaf rust resistance gene *Rph7* (Brunner *et al.*, 2003) exhibited collinearity for the suggested orthologous region in rice (Kilian *et al.*, 1997; Han *et al.*, 1999). Markers from the orthologous rice intervals could be directly used for efficient marker saturation in the corresponding region of the barley genome. Alternatively, on the basis of significant sequence similarity to rice genes from the respective rice map intervals, barley cDNA clones/ESTs could be selected for designing markers. In all three cases, a physical BAC contig could be established across the target locus by several steps of chromosome walking in barley, but in none of the three cases, sequence analysis of the collinear rice contig (provided by the physical map of RGP) suggested any promising candidate gene. This approach, however allowed successful isolation of the gene *Rpg1*, which was shown to be a novel disease resistance gene with homology to receptor kinases; no orthologous gene for *Rpg1* was found in

the rice genome (Brueggeman *et al.*, 2002). Similar data for the other two loci is not yet available.

An approach similar to the above was used for the isolation of the vernalization response gene *Vrn1* from *Triticum monococcum* (Yan *et al.*, 2003). Complete marker/gene collinearity was observed for the putative orthologous regions on *T. monococcum* chromosome $5A^m$ and rice chromosome 3 and a BAC contig of the target region was constructed from a *T. monococcum* BAC-library. It was collinear to two BACs representing the orthologous locus in rice. However, both physical maps showed a gap between the same two collinear genes. Screening of a sorghum BAC library revealed a collinear BAC that bridged the gap in the other two species leading to a consensus physical map across three cereal species. It turned out that the most promising candidate gene for *Vrn1* is an orthologue in all three species. This provides an example that is opposite to the ones discussed above for stem and leaf rust resistance loci in barley, where a candidate gene was absent or has not yet been identified in the model rice genome.

3.2.4.4. Collinearity-based gene isolation

Recently, the barley gene *ror2* (involved in basal penetration resistance to *Blumeria graminis* f.sp. *hordei*) could be isolated merely on the basis of barley-rice collinearity (Collins *et al.*, 2003). The barley microsatellite marker sequences flanking the gene were used for the identification of the syntenic rice BAC contig sequence, provided from '*The Institute of Genome Research*' (TIGR, http://www.tigr.org/). Rice genes within the contig were used for the identification of potential Triticeae orthologues by BLAST analysis (Altschul *et al.*, 1990). Genetic mapping of the potential Triticeae orthologues in a barley population of 756 F_2 plants uncovered a cosegregating EST with sequence homology to the Arabidopsis gene *pen1* (non-host penetration resistance) and could be identified as the barley gene *ror2*. This is an example, where collinearity-based use of rice genome sequences allowed direct identification of not only a rice orthologue of the barley gene *ror2*, but also that of a functional orthologue of this gene in the Arabidopsis genome.

3.2.4.5. Subgenome insert libraries for gene isolation

In hexaploid wheat, in addition to the problem due to the high proportion of repetitive DNA, the hexaploid nature itself poses another level of complexity to the strategy of MBC. The three homoeologous subgenomes A, B and D are highly collinear and most of the functional loci occur as triplicate genes. Thus, screening of a large insert library will yield two thirds of clones, which are not related to a target locus in a specific

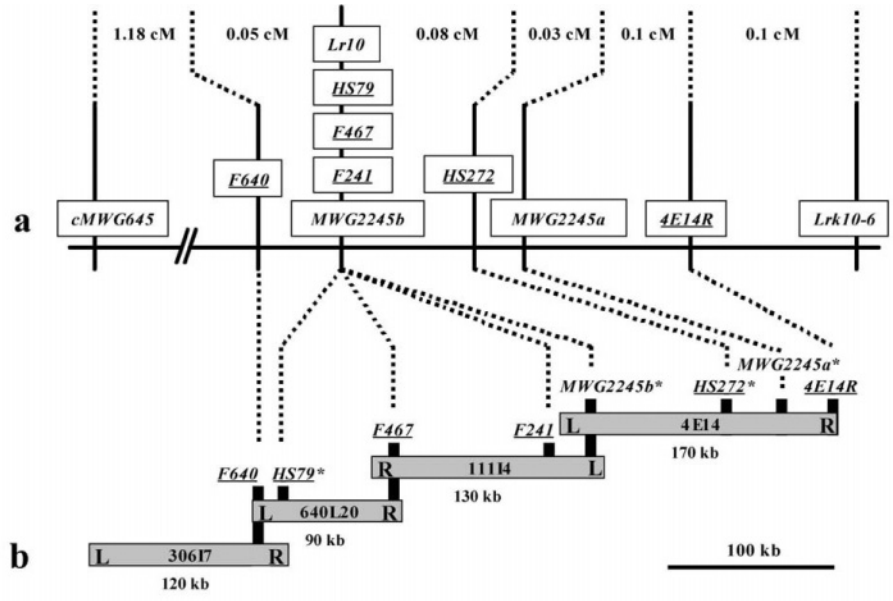


Figure 2. A schematic representation of subgenome chromosome walking at the leaf rust resistance locus *Lr10* in wheat. a) High resolution mapping of the leaf rust resistance gene *Lr10* was achieved in a F_2 population derived from a cross between a resistant (Th*Lr10*) and a susceptible (Frisal) hexaploid wheat. b) A co-segregating marker MWG2245 has been used to screen the *T. monococcum* BAC library to start a chromosomal walk. A 450 kb BAC contig could be established yielding molecular markers that were flanking the genetic position of *Lr10* in hexaploid wheat. All mappable markers derived from the established contig mapped to the target region of chromsome 1AS in complete collinearity between *aestivum* and *monococcum*. Finally, two candidate genes could be identified from the *monococcum* BAC contig. (figure taken from Stein *et al.*, 2000)

subgenome. In order to tackle this problem, large insert libraries were constructed on the basis of diploid and tetraploid wheat species that contain genomes with close relation to the progenitors of hexploid wheat (Table 3). Gene isolation in hexaploid wheat can be achieved by using these libraries for subgenome chromosome walking. A first successful example of this approach involved map-based cloning of the *Lr10* leaf rust resistance locus in bread wheat. The gene *Lr10* is located on chromosome 1AS in *T. aestivum*, for which a high-resolution map was established using 3120 F_2 plants (Fig. 2a). A three step chromosome walk in a *T. monococcum* BAC library initiated from a closely linked RFLP marker allowed to establish a contig, which contained the flanking markers and two candidate resistance genes (Stein *et al.*, 2000; Wicker *et al.*, 2001). Markers cosegregating with the gene were derived from the initial contig and additional markers were developed from low-copy sequences obtained after low-pass shotgun sequencing of neighboring BAC clones. All markers derived from the

monococcum contig mapped to collinear segments of the *T. aestivum* genetic map. The *T. aestivum* orthologues of the two candidate genes were subsequently isolated. One of the candidate genes, *Rga1*, turned out to be *Lr10* as confirmed after sequence analysis of mutant alleles and complementation via transformation into a susceptible genotype (Feuillet *et al.*, 2003).

In a second example, the *Q* locus of *T. aestivum*, conferring free-threshing and square-headed spikes, was physically delimited by the same strategy (Faris *et al.*, 2003). A much smaller population of only 465 F_2 plants was used for favourable genomic location of the gene. A proximal recombination could not be identified on the current contig. However, the target region presumably carrying the candidate gene could be delimited on the contig through the use of induced mutations. This should eventually lead to the isolation of the *Q* locus.

In a third example, the powdery mildew resistance gene *Pm3b* could be isolated from *T. aestivum* using a subgenomic BAC library. Since chromosome walking in *T. monococcum* was not successful due to a gap in the BAC library, a BAC library of the tetraploid relative, *T. turgidum* ssp. *durum* (Cenci *et al.*, 2003) was used, and a contig covering *Pm3b* could be constructed. Resistance conferred by transient expression has been monitored in epidermis of detached wheat leaves of a susceptible *T. aestivum* cultivar after biolistic bombardment with the homoeologue of the identified candidate gene and subsequent powdery mildew infection (Yahiaoui *et al.*, 2003).

Another gene conferring resistance to wheat leaf rust could be isolated through the use of a *Aegilops tauschii* (D genome) subgenomic cosmid library (Huang *et al.*, 2003a). *Lr21* was previously introgressed into *T. aestivum* via a synthetic wheat derived from a cross between *T. turgidum* and the resistant *Ae. tauschii* accession TA1649. A closely linked RFLP probe (1 recombination in 520 F_2) was used for the screening of an *Ae. tauschii* (TA1649) cosmid library. A single cosmid clone harbouring the closely linked RFLP fragment could be isolated. The *Lr21* gene, spanning 4318 bp and encodes a 1080-amino-acid protein containing a conserved nucleotide-binding site (NBS) domain, 13 imperfect leucine-rich repeats (LRRs), and a unique 151-amino-acid sequence missing from known NBS-LRR proteins at the N terminus. The whole cosmid was used for complementation via stable transformation and resistance was achieved. Thus, the identified RGH was confirmed to represent *Lr21*.

4. MAP-BASED CLONING OF QTLS

Many agronomically important traits like yield or a response to biotic and abiotic stresses show a continuous variation of the phenotype in the progeny of a cross. Thus, the underlying genes are referred to as QTLs. In contrast to their assumed importance for the success and progress in plant breeding, the isolation of QTLs for these traits was not possible until recently. However, the fact that QTLs can be dissected into single Mendelian factors (Paterson *et al.*, 1988) implied that map-based cloning should be applicable to QTL isolation also. A major progress towards this aim was possible through the use of new population models for increased genetic resolution, improved reproducibility of individual QTLs, and finally through the successful introgression/fixation of QTLs into a defined genetic background [intermated recombinant inbred populations: Liu *et al.*, 1996; advanced backcross QTL (AB-QTL) analysis: Tanksley and Nelson, 1996; exotic introgression-NIL population: Eshed and Zamir, 1994; for review, see Remington *et al.*, 2001]. The first physical map across a plant QTL was constructed in tomato for the trait fruit weight (*fw*2.2, Alpert and Tanksley, 1996) by using the AB-QTL strategy. Recent progress towards the map-based isolation of QTLs has also been made in rice: three photoperiod sensitivity QTLs (*Hd1*, *Hd3a*, *Hd6*) have been cloned (Yano *et al.*, 2000; Takahashi *et al.*, 2001; Kojima *et al.*, 2002) and several more can be expected in the near future (e.g. Lin *et al.*, 2003). This progress became possible, firstly due to avialablilty of rice genome sequences, and secondly due to the systematic development of the genetic resources (Yano, 2001). In a number of crops, and in each case for a number of important traits the AB-QTL technique is being used for simultaneous identification and transfer of QTLs from exotic germplasm, and concurrently leading to the production of NIL populations. In future, these resources will be used for map-based cloning of agronomically important QTLs (maize: Ho *et al.*, 2002, barley: Pillen *et al.*, 2003, wheat: Huang *et al.*, 2003b).

As an example for QTL isolation in crops other than rice, *Vgt1* responsible for transition from vegetative to the reproductive phase in maize is one possible target (Salvi *et al.*, 2002). However, since the overall genomic resources in cereals other than rice are limited, successful cloning of QTLs in these cereals (except maize) will take time. Nevertheless, collinearity between grass species may be efficiently utilized in case of QTLs that are conserved across species. Major QTLs for domestication traits like seed size, seed dispersal and photoperiod sensitivity seem to be conserved at least among rice, maize and sorghum (Paterson, 1995). Conserved QTLs can also be found for stress response among Triticeae species (Cattivelli *et al.*, 2002). Within the gibberellin signalling pathway also, functionally

orthologous genes have been identified in wheat, maize and Arabidopsis (Peng *et al.*, 1999). Therefore, in the long term, QTLs involved in highly conserved/basic processes have a reasonable chance of becoming the target for isolation. However, accumulating evidence for the abundance of chromosomal rearrangements and the recent demonstrations of the lack of microcollinearity among cereals may discourage, or at least make it difficult to use the map based cloning approach for isolation of loci controlling important quantitative traits.

5. SUMMARY AND OUTLOOK

Map-based cloning is a good strategy for gene isolation, if insertion mutagenesis-based forward and reversed genetic tools are not available for the genome of interest. Despite its universal potential, MBC is challenging in species with large genomes (e.g. wheat, maize, barley), and in context of cereal species can presently be considered routine only in the small-genome crop and model-plant species, rice. In recent years, MBC has been successfully achieved in almost all cereals for one or more genes, due to the recent availability of whole genome sequences particularly in rice, and other genomic resources in all cereals. Still, even in an optimal progression of chromosome walking in Triticeae species the procedure remains laborious. However, resources have improved dramatically over the past five years and reached a level that any attempt of MBC is facilitated significantly. Especially for synteny-based gene isolation or for marker saturation, the completion of the rice genomic sequence provides a very important milestone of improved resources. Furthermore, the public availability of cereal genome research data via numerous databases has made essential information easly accessible. Progress in structural and functional genomics will facilitate further the map-based cloning of important genes in all cereals, although the isolation of QTLs may stay challenging. The importance of map-based cloning is increasing especially for sorghum and maize with the availability of dense physical maps and the genomic sequences (maize), so that several genes are likely to be isolated from these crops by positional cloning in the present decade.

6. ACKNOWLEDGEMENTS

We acknowledge the help of Drs C. Feuillet, T. Lahaye and F. Ordon for useful comments on the manuscript.

7. REFERENCES

Akhunov E.D., Goodyear A.W., Geng S., Qi L.-L., Echalier B., Gill B.S., Miftahudin Gustafson J.P., Lazo G., Chao S., *et al.* The organization and rate of evolution of wheat genomes are correlated with recombination rates along chromosome arms. Genome Res 2003; 13: 753-763

Allouis S., Qi X., Lindup S., Gale M.D., Devos K.M. Construction of a BAC library of pearl millet, *Pennisetum glaucum*. Theor Appl Genet 2001; 102: 1200-1205

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

Altschul S.F., Gish W., Miller W., Myers E.W., Lipman D.J. Basic local alignment search tool. Jour Mol Biol 1990; 215: 403-410

Ashikari M., Wu J.Z., Yano M., Sasaki T., Yoshimura A. Rice gibberellin-insensitive dwarf mutant gene *Dwarf 1* encodes the alpha-subunit of GTP-binding protein. Proc Natl Acad Sci USA 1999; 96: 10284-10289

Baba T., Katagiri S., Tanoue H., Tanaka R., Chiden Y., Saji S., Hamada M., Nakashima M., Okamoto M., Hayashi M., *et al.* Construction and characterization of rice genomic libraries: PAC library of *japonica* variety, Nipponbare and BAC library of *indica* variety Kasalath. NIAR Bull 2000; 14: 41-49

Bennett M.D., Smith J.B. Nuclear DNA amounts in angiosperms. Phil Trans R Soc Lond B 1976; 274: 227-274

Bennett M.D., Leitch I.J. Nuclear DNA amounts in angiosperms. Ann Bot 1995; 76: 113-176

Bennetzen J.L., Freeling M. Grasses as a single genetic system: genome composition, collinearity and compatibility. Trends Genet 1993; 9: 259-261

Bennetzen J.L. Comparative sequence analysis of plant nuclear genomes: Microcolinearity and its many exceptions. Plant Cell 2000; 12: 1021-1029

Bennetzen J.L., Ramakrishna W. Numerous small rearrangements of gene content, order and orientation differentiate grass genomes. Plant Mol Biol 2002; 48: 821-827

Bennetzen J.L., Ma J. The genetic colinearity of rice and other cereals on the basis of genomic sequence analysis. Curr Opin Plant Biol 2003; 6: 128-133

Bernal A., Ear U., Kyrpides N. Genomes online database (GOLD): a monitor of genome projects world-wide. Nucl Acids Res 2001; 29: 126-127

Brueggeman R., Rostoks N., Kudrna D., Kilian A., Han F., Chen J., Druka A., Steffenson B., Kleinhofs A. The barley stem rust-resistance gene *Rpg1* is a novel disease-resistance gene with homology to receptor kinases. Proc Natl Acad Sci USA 2002; 99: 9328-9333

Brunner S., Keller B., Feuillet C. A large rearrangement involving genes and low-copy DNA interrupts the microcollinearity between rice and barley at the *Rph7* locus. Genetics 2003; 164: 673-683

Bryan G.T., Wu K.S., Farrall L., Jia Y.L., Hershey H.P., McAdams S.A., Faulk K.N., Donaldson G.K., Tarchini R., Valent B. A single amino acid difference distinguishes resistant and susceptible alleles of the rice blast resistance gene *Pi-ta*. Plant Cell 2000; 12: 2033-2045

Burke D.T., Carle G.F., Olson M.V. Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. Science 1987; 236: 806-812

Buschges R., Hollricher K., Panstruga R., Simons G., Wolter M., Frijters A., vanDaelen R., vanderLee T., Diergaarde P., Groenendijk J., *et al.* The barley *mlo* gene: A novel control element of plant pathogen resistance. Cell 1997; 88: 695-705

Cattivelli L., Baldi P., Crosatti C., Di Fonzo N., Faccioli P., Grossi M., Mastrangelo A.M., Pecchioni N., Stanca A.M. Chromosome regions and stress-related sequences

involved in resistance to abiotic stress in Triticeae. Plant Mol Biol 2002; 48: 649-665

Cenci A., Chantret N., Kong X., Gu Y., Anderson O.D., Fahima T., Distelfeld A., Dubcovsky J. Construction and characterization of a half million clone BAC library of durum wheat (*Triticum turgidum* ssp. *durum*). Theor Appl Genet 2003; 107: 931-939

Chandler V.L. Brendel V. The maize genome sequencing project. Plant Physiol 2002; 130: 1594-1597

Chen F.G., Zhang X.Y., Xia G.M., Jia J.Z. Construction and characterization of a bacterial artificial chromosome library for *Triticum boeoticum*. Acta Bot Sin 2002; 44: 451-456

Collins N.C., Thordal-Christensen H., Lipka V., Bau S., Kombrink E., Qiu J.-L., Huckelhoven R., Stein M., Freialdenhoven A., Somerville S.C., *et al.* SNARE protein mediated disease resistance at the plant cell wall. Nature 2003; in press

Coe E., Cone K., McMullen M., Chen S.-S., Davis G., Gardiner J., Liscum E., Polacco M., Paterson A., Sanchez-Villeda H., *et al.* Access to the maize genome: an integrated physical and genetic map. Plant Physiol 2002; 128: 9-12

Cone K.C., McMullen M.D., Bi I.V., Davis G.L., Yim Y.S., Gardiner J.M., Polacco M.L., Sanchez-Villeda H., Fang Z.W., Schroeder S.G., *et al.* Genetic, physical, and informatics resources for maize on the road to an integrated map. Plant Physiol 2002; 130: 1598-1605

Devoto A., Hartmann H.A., Piffanelli P., Elliott C., Simmons C., Taramino G., Goh C.S., Cohen F.E., Emerson B.C., Schulze-Lefert P. *et al.* Molecular phylogeny and evolution of the plant-specific seven-transmembrane MLO family. Jour Mol Evol 2003; 56: 77-88

Ding Y., Johnson M.D., Colayco R., Chen Y.J., Melnyk J., Schmitt H., Shizuya H. Contig assembly of bacterial artificial chromosome clones through multiplexed fluorescence-labeled fingerprinting. Genomics 1999; 56: 237-246

Draper J., Mur L.A.J., Jenkins G., Ghosh-Biswas G.C., Bablak P., Hasterok R., Routledge A.P.M. *Brachypodium distachyon*. A new model system for functional genomics in grasses. Plant Physiol 2001; 127: 1539-1555

Draye X., Lin Y.R., Qian X.Y., Bowers J.E., Burow G.B., Morrell P.L., Peterson D.G., Presting G.G., Ren S.X., Wing R.A., *et al.* Toward integration of comparative genetic, physical, diversity, and cytomolecular maps for grasses and grains using the sorghum genome as a foundation. Plant Physiol 2001; 125: 1325-1341

Druka A., Kudrna D., Han F., Kilian A., Steffenson B., Frisch D., Tomkins J., Wing R., Kleinhofs A. Physical mapping of the barley stem rust resistance gene *rpg4*. Mol Gen Genet 2000; 264: 283-290

Dubcovsky J., Ramakrishna W., SanMiguel P.J., Busso C.S., Yan L.L., Shiloff B.A., Bennetzen J.L. Comparative sequence analysis of colinear barley and rice bacterial artificial chromosomes. Plant Physiol 2001; 125: 1342-1353

Edwards K.J., Thompson H., Edwards D., de Saizien A., Sparks C., Thompson J.A., Greenland A.J., Eyers M., Schuch W. Construction and characterisation of a yeast artificial chromosome library containing three haploid maize genome equivalents. Plant Mol Biol 1992; 19: 299-308

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

Faris J.D., Fellers J.P., Brooks S.A., Gill B.S. A bacterial artificial chromosome contig spanning the major domestication locus *Q* in wheat and identification of a candidate gene. Genetics 2003; 164: 311-321

Feng Q., Zhang Y.J., Hao P., Wang S.Y., Fu G., Huang Y.C., Li Y., Zhu J.J., Liu Y.L., Hu X. *et al.* Sequence and analysis of rice chromosome 4. Nature 2002; 420: 316-320

Feuillet C., Keller B. Comparative genomics in the grass family: molecular characterization of grass genome structure and evolution. Ann Bot 2002; 89: 3-10

Feuillet C., Travella S., Stein N., Albar L., Nublat A., Keller B. Map-based isolation of the leaf rust disease resistance gene *Lr10* from the hexaploid wheat (*Triticum aestivum* L.) genome. Proc Natl Acad Sci USA 2003; 100:15253-15258

Flavell R.B., Bennett M.D., Smith J.B., Smith D.B. Genome size and the proportion of repeated nucleotide sequence DNA in plants. Biochem Genet 1974; 12: 257-269

Freeling M. Grasses as a single genetic system. Reassessment 2001. Plant Physiol Biochem 2001; 125: 1191-1197

Fu H.H., Dooner H.K. Intraspecific violation of genetic colinearity and its implications in maize. Proc Natl Acad Sci USA 2002; 99: 9573-9578

Fu J.M., Qu X.P., Li C.Y., Jia J.H., Jin D.M., Wang Q., Yang R.C., Wang B. Construction and characterization of a bacterial artificial chromosome library of rice 5460F. China Sci Bull 1999; 44: 1587-1592

Gibson S., Somerville C. Isolating plant genes. Trends Biotech 1993; 11: 306-313

Goff S.A., Ricke D., Lan T.H., Presting G., Wang R.L., Dunn M., Glazebrook J., Sessions A., Oeller P., Varma H., *et al.* A draft sequence of the rice genome (*Oryza sativa* L. ssp *japonica*). Science 2002; 296: 92-100

Halterman D.A., Wise R.P. A single amino acid substitution in the sixth leucine-rich repeat of barley MLA6 and MLA13 alleviates dependence on RAR1 for disease resistance signaling. Plant Jour 2004; in press

Halterman D.A., Wei F., Wise R.P. Powdery mildew-induced *Mla* mRNAs are alternatively spliced and contain multiple upstream open reading frames. Plant Physiol 2003; 131: 558-567

Halterman D.A., Zhou F., Wei F., Wise R.P. Schulze-Lefert. P. The MLA6 coiled-coil NBS-LRR protein confers *AvrMla6*-dependent resistance specificity to *Blumeria graminis* f. sp. *hordei* in barley and wheat. Plant Jour 2001; 25: 335-348

Hamilton C.M. A binary-BAC system for plant transformation with high-molecular-weight DNA. Gene 1997; 200: 107-116

Han F., Kilian A., Chen J.P., Kudrna D., Steffenson B., Yamamoto K., Matsumoto T., Sasaki T., Kleinhofs A. Sequence analysis of a rice BAC covering the syntenous barley Rpg1 region. Genome 1999; 42: 1071-1076

Ho J.C., McCouch S.R., Smith M.E. Improvement of hybrid yield by advanced backcross QTL analysis in elite maize. Theor Appl Genet 2002; 105: 440-448

Huang L., Brooks S.A., Li W., Fellers J.P., Trick H.N., Gill B.S. Map-based cloning of leaf rust resistance gene *Lr21* from the large and polyploid genome of bread wheat. Genetics 2003a; 164: 655-664

Huang X.Q., Cöster H., Ganal M.W., Röder M.S. Advanced backcross QTL analysis for the identification of quantitative trait loci alleles from wild relatives of wheat (*Triticum aestivum* L.). Theor Appl Genet 2003b; 106: 1379-1389

Jander G., Norris S.R., Rounsley S.D., Bush D.F., Levin I.M., Last R.L. Arabidopsis map-based cloning in the post-genome era. Plant Physiol 2002; 129: 440-450

Keller B., Feuillet C. Colinearity and gene density in grass genomes. Trends Plant Sci 2000; 5: 246-251

Kelley J.M., Field C.E., Craven M.B., Bocskai D., Kim U.J., Rounsley S.D.Adams M.D. High throughput direct end sequencing of BAC clones. Nucl Acid Res 1999; 27: 1539-1546

Kikuchi S., Satoh K., Nagata T., Kawagashira N., Doi K., Kishimoto N., Yazaki J., Ishikawa M., Yamada H., Ooka H., *et al.* Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice. Science 2003; 301: 376-379

Kilian A., Chen J., Han F., Steffenson B., Kleinhofs A. Towards map-based cloning of the barley stem rust resistance genes *Rpgl* and *rpg4* using rice as an intergenomic cloning vehicle. Plant Mol Biol 1997; 35: 187-195

Klein P.E., Klein R.R., Cartinhour S.W., Ulanch P.E., Dong J.M., Obert J.A., Morishige D.T., Schlueter S.D., Childs K.L., Ale M. *et al.* A high-throughput AFLP-based method for constructing integrated genetic and physical maps: Progress toward a sorghum genome map. Genome Res 2000; 10: 789-807

Klein P.E., Klein R.R., Vrebalov J., Mullet J.E. Sequence-based alignment of sorghum chromosome 3 and rice chromosome 1 reveals extensive conservation of gene order and one major chromosomal rearrangement. Plant Jour 2003; 34: 605-621

Klein R.R., Klein P.E., Chhabra A.K., Dong J., Pammi S., Childs K.L., Mullet J.E., Rooney W.L., Schertz K.F. Molecular mapping of the *rf1* gene for pollen fertility restoration in sorghum (*Sorghum bicolor* L.). Theor Appl Genet 2001; 102: 1206-1212

Kleine M., Michalek W., Diefenthal T., Dargatz H., Jung C. Construction of a M*lu*I-YAC library from barley (*Hordeum vulgare* L.) and analysis of YAC insert terminal regions. Genome 1997; 40: 896-902

Kojima S., Takahashi Y., Kobayashi Y., Monna L., Sasaki T., Araki T., Yano M. Hd3a, a rice ortholog of the Arabidopsis FT gene, promotes transition to flowering downstream of *Hd1* under short-day conditions. Plant Cell Physiol 2002; 43: 1096-1105

Koprek T., McElroy D., Louwerse J., Williams-Carrier R., Lemaux P.G. An efficient method for dispersing Ds elements in the barley genome as a tool for determining gene function. Plant Jour 2000; 24: 253-263

Kunzel G., Korzun L., Meister A. Cytologically integrated physical restriction fragment length polymorphism maps for the barley genome based on translocation breakpoints. Genetics 2000; 154: 397-412

Lahaye T., Shirasu K., Schulze-Lefert P. Chromosome landing at the barley *Rar1* locus. Mol Gen Genet 1998; 260: 92-101

Laurie D.A., Devos K.M. Trends in comparative genetics and their potential impacts on wheat and barley research. Plant Mol Biol 2002; 48: 729-740

Leister D., Kurth J., Laurie D.A., Yano M., Sasaki T., Devos K., Graner A., Schulze-Lefert P. Rapid reorganization of resistance gene homologues in cereal genomes. Proc Natl Acad Sci USA 1998; 95: 370-375

Leutweiler L.S., Hough-Evans B.R., Meyerowitz E.M. The DNA of *Arabidopsis thaliana.* Mol Gen Genet 1984; 194: 15-23

Leyser O., Chang C. Chromosome walking. In: G.D. Foster, D. Twell D. (eds.) *Plant Gene Isolation*, John Wiley and Sons, Ltd, Chichester, 1996, pp. 248-271

Lijavetzky D., Muzzi G., Wicker T., Keller B., Wing R., Dubcovsky J. Construction and characterization of a bacterial artificial chromosome (BAC) library for the A genome of wheat. Genome 1999; 42: 1176-1182

Lin H.X., Liang Z.W., Sasaki T., Yano M. Fine mapping and characterization of quantitative trait loci *Hd4* and *Hd5* controlling heading date in rice. Breed Science 2003; 53: 51-59

Lin Y.R., Zhu L.H., Ren S.X., Yang J.S., Schertz K.F., Paterson A.H. A *Sorghum propinquum* BAC library, suitable for cloning genes associated with loss-of-function mutations during crop domestication. Mol Breed 1999; 5: 511-520

Liu S.C., Kowalski S.P., Lan T.H., Feldmann I.A., Paterson A.H. Genome-wide high-resolution mapping by recurrent intermating using *Arabidopsis thaliana* as a model. Genetics 1996; 142: 247-258

Maes T., De Keukeleire P., Gerats T. Plant tagnology. Trends Plant Sci 1999; 4: 90-96

Matzke M.A., Matzke A.J., Pruss G.J., Vance V.B. RNA-based silencing strategies in plants. Curr Opin Genet Dev 2001; 11: 221-227

McCallum C.M., Comai L., Greene E.A., Henikoff S. Targeted screening for induced mutations. Nature Biotech 2000; 18: 455-457

McDowell J.M., Woffenden B.J. Plant disease resistance genes: recent insights and potential applications. Trends Biotech 2003; 21: 178-183

Menz M.A., Klein R.R., Mullet J.E., Obert J.A., Unruh N.C., Klein P.E. A high-density genetic map of *Sorghum bicolor* (L.) Moench based on 2926 AFLP (R), RFLP and SSR markers. Plant Mol Biol 2002; 48: 483-499

Michelmoore R.W., Paran I., Kesseli R.V. Identification of markers linked to disease-resistance genes by bulked segregant analysis: A rapid method to detect markers in specific genomic regions by using segregating populations. Proc Natl Acad Sci USA 1991; 88: 9828-9832

Miyazaki C., Saito A. Construction and characterization of a bacterial artificial chromosome library of the indica rice Kasalath. Breed Science 1999; 49: 193-201

Moore G., Devos K.M., Wang Z., Gale M.D. Grasses line up and form a circle. Curr Biol 1995; 5: 737-739

Moore G. The application of comparative genomics to wheat. In: *International Plant and Animal Genome XI Conference*, *San Diego, CA, USA, 2003*, Abstract no W216

Moullet O., Zhang H.B., Lagudah E.S. Construction and characterisation of a large DNA insert library from the D genome of wheat. Theor Appl Genet 1999; 99: 305-313

Nakamura S., Asakawa S., Ohmido N., Fukui K., Shimizu N., Kawasaki S. Construction of an 800-kb contig in the near-centromeric region of the rice blast resistance gene *Pi-ta(2)* using a highly representative rice BAC library. Mol Gen Genet 1997; 254: 611-620

Neu C., Stein N., Keller B. Genetic mapping of the *Lr20-Pm1* resistance locus reveals suppressed recombination on chromosome arm 7AL in hexaploid wheat. Genome 2002; 45: 737-744

Nilmalgoda S.D., Cloutier S., Walichnowski A.Z. Construction and characterization of a bacterial artificial chromosome (BAC) library of hexaploid wheat (*Triticum aestivum* L.) and validation of genome coverage using locus-specific primers. Genome 2003; 46: 870-878

Nurminsky D.I., Hartl D.L. Sequence scanning: A method for rapid sequence acquisition from large-fragment DNA clones. Proc Natl Acad Sci USA 1996; 93: 1694-1698

O'Sullivan D.M., Ripoll P.J., Rodgers M., Edwards K.J. A maize bacterial artificial chromosome (BAC) library from the European flint inbred lined F2. Theor Appl Genet 2001; 103: 425-432

Paterson A., Lander E.S., Hewitt J.D., Peterson S., S.E., L., Tanksley S.D. Resolution of quantitative traits into mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature 1988; 335: 721-726

Paterson A.H. Molecular dissection of quantitative traits - Progress and prospects. Genome Res 1995; 5: 321-333

Peng J.R., Richards D.E., Hartley N.M., Murphy G.P., Devos K.M., Flintham J.E., Beales J., Fish L.J., Worland A.J., Pelica F., *et al.* 'Green revolution' genes encode mutant gibberellin response modulators. Nature 1999; 400: 256-261

Peng K.M., Zhang H.B., Zhang Q.F. A BAC library constructed to the rice cultivar "Minghui 63" for cloning genes of agronomic importance. Acta Bot Sin 1998; 40: 1108-1114

Peters J.L., Cnudde F., Gerats T. Forward genetics and map-based cloning approaches. Trends Plant Sci 2003; 8: 484-491

Pillen K., Zacharias A., Leon J. Advanced backcross QTL analysis in barley (*Hordeum vulgare* L.). Theor Appl Genet 2003; 107: 340-352

Qi L.L., Gill B.S. 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. Theor Appl Genet 2001; 103: 998-1006

Remington D.L., Ungerer M.C., Purugganan M.D. Map-based cloning of quantitative trait loci: progress and prospects. Genet Res 2001; 78: 213-218

Ripoll P.J., O'Sullivan D.M., Edwards K.J., Rodgers M. Technique for cloning and sequencing the ends of bacterial artificial chromosome inserts. Biotechniques 2000; 29: 271-274

Saji S., Umehara Y., Antonio I.A., Yamane I., Tanoue H., Baba T., Aoki H., Ishige N., Wu J., Koike K., *et al.* A physical map with yeast artificial chromosome (YAC) clones covering 63% of the 12 rice chromosomes. Genome 2001; 44: 32-37

Salvi S., Tuberosa R., Chiapparino E., Maccaferri M., Veillet S., van Beuningen L., Isaac P., Edwards K., Phillips R.L. Toward positional cloning of Vgt1, a QTL controlling the transition from the vegetative to the reproductive phase in maize. Plant Mol Biol 2002; 48: 601-613

SanMiguel P., Tikhonov A., Jin Y.K., Motchoulskaia N., Zakharov D., MelakeBerhan A., Springer P.S., Edwards K.J., Lee M., Avramova Z., *et al.* Nested retrotransposons in the intergenic regions of the maize genome. Science 1996; 274: 765-768

SanMiguel P., Bennetzen J.L. Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. Ann Bot 1998; 82: 37-44

Sasaki T., Matsumoto T., Yamamoto K., Sakata K., Baba T., Katayose Y., Wu J.Z., Niimura Y., Cheng Z.K., Nagamura Y., *et al.* The genome sequence and structure of rice chromosome 1. Nature 2002; 420: 312-316

Scherrer B., Keller B., Feuillet C. Two haplotypes of resistance gene analogs have been conserved during evolution at the leaf rust resistance locus *Lr10* in wild and cultivated wheat. Funct Integr Genom 2002; 2: 40-50

Schmidt D., Roder M.S., Dargatz H., Wolf N., Schweizer G.F., Tekauz A., Ganal M.W. Construction of a YAC library from barley cultivar Franka and identification of YAC-derived markers linked to the *Rh2* gene conferring resistance to scald (*Rhynchosporium secalis*). Genome 2001; 44: 1031-1040

Scholz S., Lorz H., Lutticke S. Transposition of the maize transposable element Ac in barley (*Hordeum vulgare* L.). Mol Gen Genet 2001; 264: 653-661

Schutte M., da Costa L.T., Moskaluk C.A., E., R., Guan X., de Jong P.J., Bittner M., Meltzer P.S., Trent J.M., Kern S.E. Isolation of YAC insert sequences by representational difference analysis. Nucl Acid Res 1995; 23: 4127-4133

Shen Q., Bieri S., Zhou F., Haizel T., Shirasu K., Schulze-Lefert P. Recognition specificity and RAR1/SGT1 dependency in barley *Mla* disease resistance alleles to the powdery mildew fungus. Plant Cell 2003; 15: 732-744

Shirasu K., Lahaye T., Tan M.W., Zhou F.S., Azevedo C., Schulze-Lefert P. A novel class of eukaryotic zinc-binding proteins is required for disease resistance signaling in barley and development in *C. elegans*. Cell 1999; 99: 355-366

Shizuya H., Birren B., Kim U.-J., Mancino V., Slepak T., Tachiiri Y., Simon M. Cloning and stable maintenance of 300-kilobase fragments of human DNA in *Escherichia coli* using and F-factor-based vector. Proc Natl Acad Sci USA 1992; 89: 8794-8797

Soderlund C., Longden I., Mott R. FPC: a system for building contigs from restriction fingerprinted clones. Comput Appl Biosci 1997; 13: 523-535

Song W.Y., Wang G.L., Chen L.L., Kim H.S., Pi L.Y., Holsten T., Gardner J., Wang B., Zhai W.X., Zhu L.H., *et al.* A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa21*. Science 1995; 270: 1804-1806

Stein N., Feuillet C., Wicker T., Schlagenhauf E., Keller B. 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.). Proc Natl Acad Sci USA 2000; 97: 13436-13441

Strasburger E. "*Lehrbuch der Botanik*", D. von Denffer, Ziegler H., Ehrendorfer F., Bresinsky A. (eds.), Gustav Fischer Verlag, Stuttgart, 1983

TAGI, The Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 2000; 408: 796-815

Takahashi Y., Shomura A., Sasaki T., Yano M. *Hd6* a rice quantitative trait locus involved in photoperiod sensitivity, encodes the alpha subunit of protein kinase CK2. Proc Natl Acad Sci USA 2001; 98: 7922-7927

Takahashi J.S., Pinto L.H., Vitaterna M.H. Forward and reverse genetic approaches to behavior in the mouse. Science 1994; 264: 1724-1733

Takumi S., Murai K., Mori N., Nakamura C. Trans-activation of a maize Ds transposable element in transgenic wheat plants expressing the Ac transposase gene. Theor Appl Genet 1999; 98: 947-953

Tanksley S.D., Ganal M.W. Martin G.B. Chromosome landing - a paradigm for map-based gene cloning in plants with large genomes. Trends Genet 1995; 11: 63-68

Tanksley S.D., Nelson J.C. Advanced backcross QTL analysis: A method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. Theor Appl Genet 1996; 92: 191-203

Tao Q., Wang A., Zhang H.B. One large-insert plant-transformation-competent BIBAC library and three BAC libraries of *japonica* rice for genome research in rice and other grasses. Theor Appl Genet 2002; 105: 1058-1066

Tomkins J.P., Davis G., Main D., Yim Y., Duru N., Musket T., Goicoechea J.L., Frisch D.A., Coe E.H., Wing R.A. Construction and characterization of a deep-coverage bacterial artificial chromosome library for maize. Crop Sci 2002; 42: 928-933

Umehara Y., Inagaki A., Tanoue H., Yasukochi Y., Nagamura Y., Saji S., Otsuki Y., Fujimura T., Kurata N., Minobe Y. Construction and characterization of a rice yac library for physical mapping. Mol Breed 1995; 1: 79-89

Vos P., Hogers R., Bleeker M., Reijans M., Vandelee T., Hornes M., Frijters A., Pot J., Peleman J., Kuiper M., *et al.* AFLP - a new technique for DNA-fingerprinting. Nucl Acid Res 1995; 23: 4407-4414

Walbot V. Saturation mutagenesis using maize transposons. Curr Opin Plant Biol 2000; 3: 103-107

Wang G.L., Holsten T.E., Song W.Y., Wang H.P., Ronald P.C. Construction of a rice bacterial artificial chromosome library and identification of clones linked to the *Xa-21* disease resistance locus. Plant Jour 1995; 7: 525-533

Wang W., Zhai W., Luo M., Jiang G., Chen X., Li X., Wing R.A., Zhu L. Chromosome landing at the bacterial blight resistance gene *Xa4* locus using a deep coverage rice BAC library. Mol Genet Genom 2001; 265: 118-125

Wang Z.X., Yano M., Yamanouchi U., Iwamoto M., Monna L., Hayasaka H., Katayose Y., Sasaki T. The *Pib* gene for rice blast resistance belongs to the nucleotide binding and leucine-rich repeat class of plant disease resistance genes. Plant Jour 1999; 19: 55-64

Wei F., Gobelman-Werner K., Morroll S.M., Kurth J., Mao L., Wing R., Leister D., Schulze-Lefert P. Wise R.P. The *Mla* (powdery mildew) resistance cluster is associated with three NBS-LRR gene families and suppressed recombination within a 240-kb DNA interval on chromosome 5S (1HS) of barley. Genetics 1999; 153: 1929-1948

Wei F., Wing R., Wise R.P. Genome dynamics and evolution of the *Mla* (powdery mildew) resistance locus in barley. Plant Cell 2002; 14:1903-1917

Wen L., Tang H.V., Chen W., Chang R., Pring D.R., Klein P.E., Childs K.L., Klein R.R. Development and mapping of AFLP markers linked to the sorghum fertility restorer gene *rf4*. Theor Appl Genet 2002; 104: 577-585

Wicker T., Stein N., Albar L., Feuillet C., Schlagenhauf E., Keller B. Analysis of a contiguous 211 kb sequence in diploid wheat (*Triticum monococcum* L.) reveals multiple mechanisms of genome evolution. Plant Jour 2001; 26: 307-316

Woo S.S., Jiang J.M., Gill B.S., Paterson A.H., Wing R.A. Construction and characterization of a bacterial artificial chromosome library of *Sorghum bicolor*. Nucl Acid Res 1994; 22: 4922-4931

Wu J., Maehara T., Shimokawa T., Yamamoto S., Harada C., Takazaki Y., Ono N., Mukai Y., Koike K., Yazaki J., *et al.* A comprehensive rice transcript map containing 6591 expressed sequence tag sites. Plant Cell 2002; 14: 525-535

Yahiaoui N., Srichumpa P., Dudler R.Keller B. Genome analysis at different ploidy levels allows cloning of the powdery mildew resistance gene *Pm3b* from hexaploid wheat. Plant Jour 2003; in press

Yan L., Loukoianov A., Tranquilli G., Helguera M., Fahima T. Dubcovsky J. Positional cloning of the wheat vernalization gene *VRN1*. *Proc Natl* Acad Sci USA 2003; 100: 6263-6268

Yano M., Katayose Y., Ashikari M., Yamanouchi U., Monna L., Fuse T., Baba T., Yamamoto K., Umehara Y., Nagamura Y., *et al. Hd1*, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the *Arabidopsis* flowering time gene *CONSTANS*. Plant Cell 2000; 12: 2473-2483

Yano M. Genetic and molecular dissection of naturally occurring variation. Curr Opin Plant Biol 2001; 4: 130-135

Yim Y.S., Davis G.L., Duru N.A., Musket T.A., Linton E.W., Messing J.W., McMullen M.D., Soderlund C.A., Polacco M.L., Gardiner J.M., *et al.* Characterization of three maize bacterial artificial chromosome libraries toward anchoring of the physical map to the genetic map using high-density bacterial artificial chromosome filter hybridization. Plant Physiol 2002; 130: 1686-1696

Yoshimura S., Yamanouchi U., Katayose Y., Toki S., Wang Z.-X., Kono I., Kurata N., Yano M., Iwata N., Sasaki T. Expression of *Xa1*, a bacterial blight-resistance gene in rice is induced by bacterial inoculation. Proc Natl Acad Sci USA 1998; 95: 1663-1668

Yu J., Hu S.N., Wang J., Li S.G., Wong K.S.G., Liu B., Deng Y., Dai L., Zhou Y., Zhang X.Q. *et al.* A draft sequence of the rice (*Oryza sativa* ssp. *indica*) genome. China Sci Bull 2001; 46: 1937-1942

Yu Y., Tomkins J.P., Waugh R., Frisch D.A., Kudrna D., Kleinhofs A., Brueggeman R.S., Muehlbauer G.J., Wise R.P., Wing R.A. A bacterial artificial chromosome library for barley (*Hordeum vulgare* L.) and the identification of clones containing putative resistance genes. Theor Appl Genet 2000; 101: 1093-1099

Zhang H.B., Choi S.D., Woo S.S., Li Z.K., Wing R.A. Construction and characterization of two rice bacterial artificial chromosome libraries from the parents of a permanent recombinant inbred mapping population. Mol Breed 1996; 2: 11-24

Zhao S.Y., Shatsman S., Ayodeji B., Geer K., Tsegaye G., Krol M., Gebregeorgis E., Shvartsbeyn A., Russell D., Overton L., *et al.* Mouse BAC ends quality assessment and sequence analyses. Genome Res 2001; 11: 1736-1745

Zhou F.S., Kurth J.C., Wei F.S., Elliott C., Vale G., Yahiaoui N., Keller B., Somerville S., Wise R., Schulze-Lefert P. Cell-autonomous expression of barley *Mla1* confers race-specific resistance to the powdery mildew fungus via a *Rar1*-independent signaling pathway. Plant Cell 2001; 13: 337-350

Chapter 12

GENE DISTRIBUTION IN CEREAL GENOMES

Kulvinder S. Gill

Department of Crop and Soil Sciences, Washington State University, Pullman, WA 99164-6420, USA
E-mail: ksgill@wsu.edu

1. INTRODUCTION

Genome size differs more than 10,000 fold among different species of eukaryotes. Even among eukaryotes of similar biological complexity, difference in genome size may be few hundred-fold. A small part of this difference in genome size may be attributed to gene number and, to some extent gene size, which both increase with increase in biological complexity. A much larger proportion of the difference in genome size, however, is due to non-transcribing repeated DNA that is mainly composed of retrotransposons and pseudogenes. For many cereals, only a small fraction of otherwise very large genomes encompasses genes, which are distributed unevenly along the chromosomes. One of the methods used for the study of the organization of different elements (including genes) within the genome is to generate molecular maps. These maps generally include DNA markers, but sometimes also include genes of interest. Recombination frequencies are generally used as measures of distances between markers and genes of interest, so that a precise estimate of recombination at a sub-chromosomal level is mandatory for any study of gene distribution within the genome and for map based cloning of genes. However, it has been observed that recombination frequencies are also unevenly distributed along the chromosomes of cereals and other higher eukaryotes, so that genetic maps may not give a true picture of the distribution of genes on different chromosomes. Therefore, physical maps of chromosomes are often prepared through the use of cytogenetic stocks (aneuploids, deletions) or the techniques of fluorescence in situ hybridization (FISH). A comparison of these physical maps with the genetic maps resolves the uneven distribution of recombination frequencies and

P.K. Gupta and R.K. Varshney (eds.), Cereal Genomics, 361-384.

allows the study of distribution of genes along the chromosomes. This chapter will compare distribution of genes and recombination frequencies among grass genomes of varying sizes and will discuss relationship of gene distribution with genome size. A detailed description of relationship between the distribution of genes and the distribution of recombination frequencies is also presented.

2. DISTRIBUTION OF GENES WITHIN A GENOME

2.1. Genome Size vs Number of Genes in a Genome

Among the sequenced eukaryotic genomes, the genome of *S. cerevisiae* is the smallest followed by those of *Caenorhabditis elegans*, *Arabidopsis thaliana*, *Drosophila melanogaster*, and *Homo sapiens* (human) in an increasing order of genome size (Fig. 1) (Goffeau *et al.*, 1996; TCeSC, 1998; TAGI, 2000; Adams *et al.*, 2000; IHGSC, 2001). The predicted number of genes is about 6,000 in budding yeast, 13,600 in fruit-fly, 18,420 in the worm, 25,000 in Arabidopsis, and 30,000 in human. There is a general tendency of increased number of genes and genome size with the increased biological complexity. However, the difference in the number of genes between the simple unicellular yeast *S. cerevisiae* and the highly evolved and perhaps the most complex organism, *H. sapiens*, is only about five-fold, the corresponding difference for the genome size being about 267-fold (Fig. 1). Similarly, the genome size of organisms with a similar biological complexity, in some groups of higher plants may also differ by as much as 240-fold.

The actual number of functional genes in any organism will probably be less than the predicted numbers. For instance, in yeast, so far 57% of the genes have neither been functionally characterized nor did they display homology to proteins with known biochemical functions (Mewes *et al.*, 1997). Most of the estimates of number of genes are from the open reading frame (ORF) finding computer softwares and from comparison with expressed sequence tags (ESTs). Beside other parameters, minimum gene size can dramatically affect the estimated number of genes. Presence of pseudogenes may further inflate the gene number estimates. This is particularly significant in organisms with larger genomes, where the number of pseudogenes is expected to be much higher (see section on '*Gene-empty regions*'). Gene calling based on comparisons with ESTs can also be erroneous because of the presence of large genes found in human and others. Two non-overlapping EST contigs corresponding to the same gene may also be counted more than once. Another evidence of inflated estimates of the number of genes in an

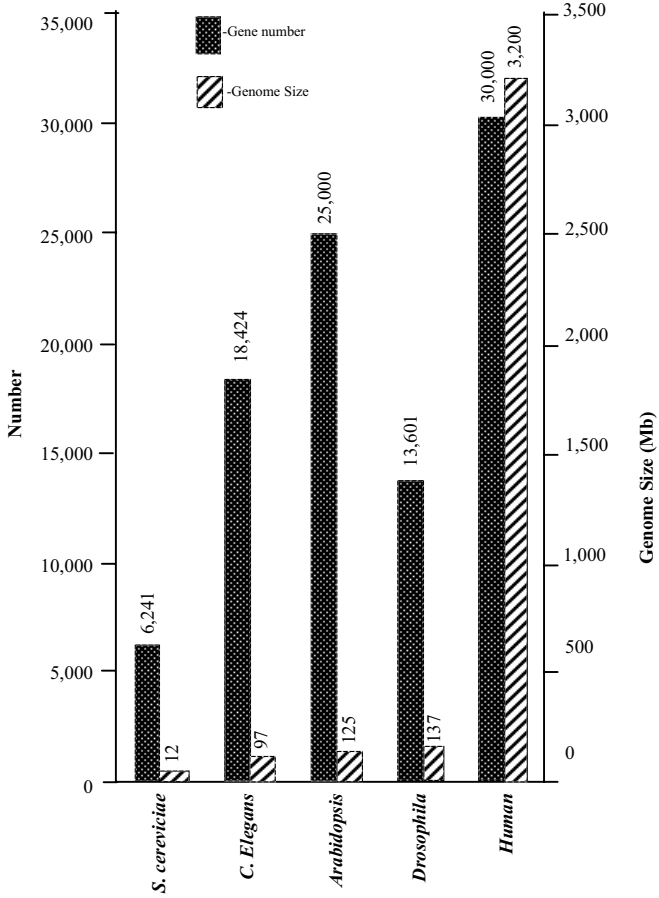


Figure 1. Comparison of genome size and number of genes among some eukaryotes, whose genomes have been sequenced.

organism is the experience that the number of genes goes down with improved bioinformatics tools. For instance, in humans, the initial estimate was 120,000 compared to 30,000, which is the latest and most reliable estimate available so far (Spring, 1997; IHGSC, 2001). Similarly, the estimate in Arabidopsis has gone down over time from 50,000 to 25,000. The number of genes in higher plants, particularly in different members within a family, is expected to be the same although differences are expected due to ploidy level. The genome size among higher plants can, however, vary as much as 240 times, as observed between Arabidopsis (Cruciferae) and *Tulipa* sp. (Liliaceae) (Arumuganathan and Earle, 1991). Even within Poaceae

family, the genome size difference can be as much as 37-fold, as observed between wheat (*Triticum aestivum* L.) (16,000Mb) and rice (430Mb). Even if difference in genome size due to polyploidy is ignored, this difference is about 12-fold.

In silico estimates of number of gene for Arabidopsis is 25,000 and for rice it ranges from 30,000 to 50,000 (TAGI, 2000; Goff *et al.*, 2002; Yu *et al.*, 2002). In comparison with human, the estimates of the number of genes for rice and perhaps also Arabidopsis seem inflated. Furthermore, a significant loss of DNA may follow polyploidization, as observed in wheat, where up to 15% of homoeologous gene loci were lost in F_1 hybrids and/or in the first generation of chromosome doubling (Ozkan *et al.*, 2001; Shaked *et al.*, 2001). Considering these observations, the number of genes in bread wheat should be <25,000 to 50,000.

Difference in gene size may also explain some of the differences in genome size. The average gene size in eukaryotes ranges from 1.7kb in budding yeast to 27kb in human (Mewes *et al.*, 1997; IHGSC, 2001). The average size of gene in Arabidopsis, rice, worm, and fruit-fly is approximately 2kb, 2.5kb, 5kb, and 11kb, respectively (TAGI, 2000; Yu *et al.*, 2002; TCeSC, 1998; Adams *et al.*, 2000). The available data suggest that gene size among grass species does not vary significantly. Assuming 50,000 to 150,000 genes with an average size of 2.5kb, only 0.8 to 2.4% (160 to 480Mb) of the wheat genome is expected to contain genes.

2.2. Gene Distribution in Wheat Genome

Among the poorly sequenced genomes, distribution of genes is best known for the wheat genome. Wheat is a segmental allohexaploid containing three related genomes that originated from a common progenitor (Kihara, 1944; McFadden and Sears, 1946). The gene distribution of wheat was resolved by physically mapping gene markers on an array of single-break deletion lines (Endo and Gill, 1996; Gill *et al.*, 1996a, b; Sandhu *et al.*, 2001). Although subtle differences do exist, gene synteny and collinearity appear to be conserved among the wheat homoeologues at the current resolution. This conservation makes it possible to combine deletion breakpoints of the three homoeologues on a consensus map in order to increase resolution of the physical maps. High-density consensus physical maps containing ~3000 gene markers mapped on 352 deletion lines revealed that about 30% of the wheat genome contains >90% of the genes (Erayman *et al.*, 2003). Most of the wheat genes are present in clusters spanning physically small regions (gene- rich regions, GRRs). A total of 48 GRRs were identified with an

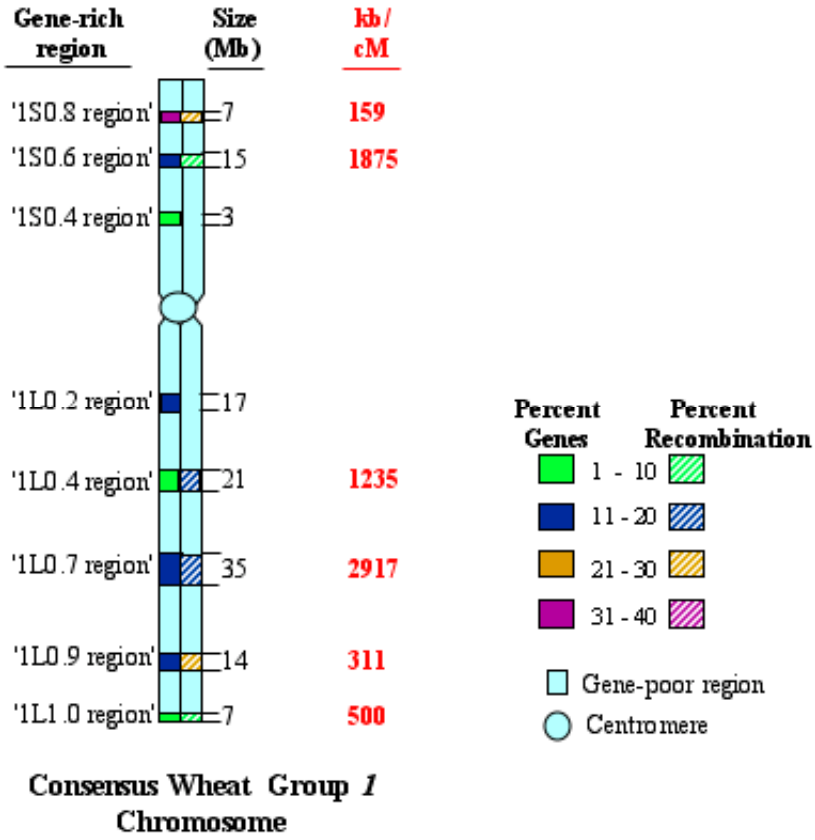


Figure 2. Distribution of genes and recombination on wheat homoeologous group 1 chromosomes (Sandhu and Gill, 2002a). The consensus chromosome and the location and size of the gene-rich regions were drawn to scale, based on the average size of the three homoeologous chromosomes. Actual physical size (black) and the ratio of physical to genetic distance (red) for a region are given on the right hand side of the consensus chromosome. The coloured version of this figure is available on the cover of this book.

average of about seven per chromosome. Of these, 20 were major GRRs spanning 13% of the genome but accounting for ~66% of the genes. The GRRs are interspersed by gene-poor compartments predominantly consisting of retrotransposon-like repetitive DNA sequences (SanMiguel *et al.*, 1996; Barakat *et al.*, 1997; Feuillet and Keller, 1999) and pseudogenes. The GRRs vary in size and have a gene-density with a general trend of increased gene-density towards the distal parts of the chromosome arms. About 60% of the genes are present in the distal one-third of the chromosomes (Gill *et al.*, 1993; Sandhu and Gill, 2002a; Erayman *et al.*, 2003; Akhunov *et al.*, 2003).

An example of gene distribution on wheat chromosomes is shown in Fig. 2 using wheat homoeologous group *1*. On wheat group *1* chromosomes, there are eight GRRs, of which three are on the short arm and five are on the long arm. These encompass ~119Mb of DNA with a size ranging from 3 to 35Mb per GRR. The estimated size of chromosome *1B* is ~800Mb. All GRRs have been demarcated by deletion breakpoints. The GRRs differ for size, number of genes and gene-density. For instance, the '1S0.8 region' is the smallest of all GRRs but has the highest gene-density that is ~12 times higher than that in the '1L1.0 region'. Accuracy of bracketing the GRRs depends upon the number of deletion breakpoints. Currently, with 352 deletion lines, a breakpoint occurs every ~16Mb on the consensus wheat map. Gametocidal gene-based chromosomal breaks however occur preferentially around the GRRs (Gill *et al.*, 1993; Gill, 1996b), so that the breakpoints were twice as frequent (every 7Mb) around the GRRs as one would expect due to random distribution (Erayman *et al.*, 2003). Even with that frequency, GRRs that are each interspersed by <7Mb of gene-poor DNA regions, will not be resolved. Therefore, the GRRs shown in Fig. 2 may be further partitioned into mini gene-rich and gene-poor regions, as and when better resolution becomes available for the gene-containing regions of the wheat genome (see section 2.4).

2.3. Gene Distribution in Genomes of Other Cereals and Other Eukaryotes

The distribution of genes in grass species other than wheat has also been found to be uneven. Even in the smaller grass genome of rice, although ~20% of the genome has not been contiged/ sequenced, despite major efforts (Wu *et al.*, 2002), in the remaining 80% of the rice genome, which has been sequenced, gene distribution is highly uneven (Fig. 3a). Mapping of 6591 unique ESTs on yeast artificial chromosome (YAC)-based contiguous map of rice also revealed that 21% of the genome contained 40% of the genes (Wu *et al.*, 2002). The difference in gene density among various regions was more than 10-fold. In barley, the translocation breakpoint-based physical maps showed that barley chromosomes are also partitioned into gene-rich and gene-poor regions (Kunzel *et al.*, 2000). The location and relative density of genes in barley is very similar to that in wheat although some subtle differences were observed. Similarly, gene-density revealed by sequencing ~200kb around well-characterized genes strongly suggests that the gene distribution is uneven in maize and sorghum also (see section 2.4) (Chen *et al.*, 1998; Feuillet and Keller, 1999; Tikhonov *et al.*, 1999).

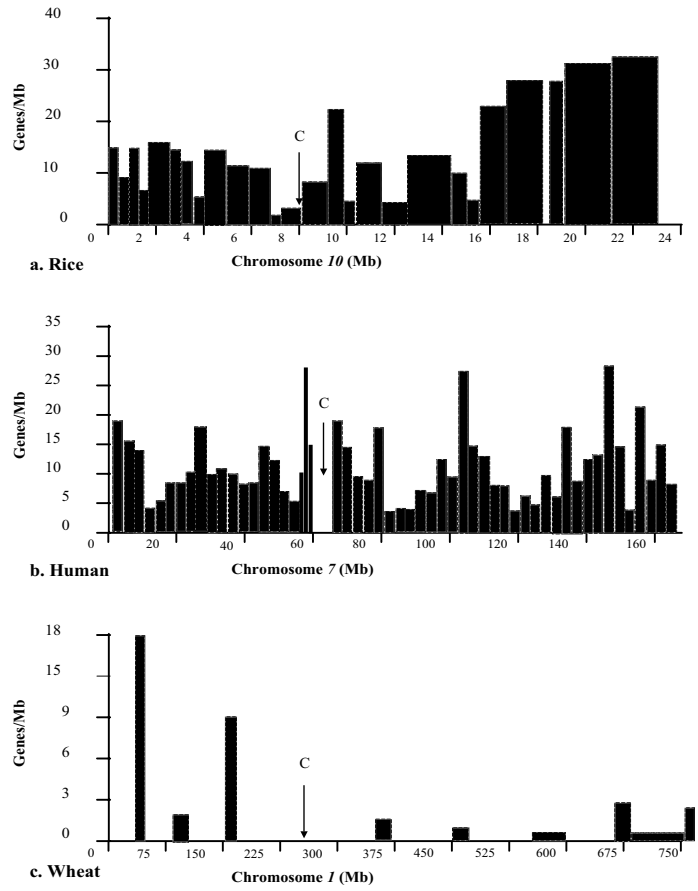


Figure 3. Distribution of genes across chromosomes of rice, human, and wheat. For each species, a chromosome was selected randomly. Gene distribution is based on mapping of 369 ESTs on YAC contigs (Wu *et al.*, 2002; http://rgp.dna.affrc.go.jp/Publicdata.html) in rice, mapping of 1784 genes on contigs in human (http://www.ncbi.nlm.nih.gov), and mapping of 260 gene sequences on chromosome deletion lines in wheat (Sandhu and Gill, 2002a). Sequencing gaps that are present both in rice and human are not on the figure and were assumed to be absent. The centromere is marked by an arrow and shown as 'C'.

In view of the above, it seems obvious now that uneven distribution of genes on chromosomes is a common feature of all eukaryotes. Even in *S. cerevisiae*, gene-density among regions may vary more than 2-fold (Dujon *et al.*, 1997). For example, in chromosome XV, gene-density varied from <0.4 to 0.85 genes per kb. Similarly, in the smaller genome of Arabidopsis, genes are asymmetrically distributed on the chromosomes. About 45% of the genome accounts for 25,000 genes with an average size of ~2kb. The remaining 55% of the genome is 'gene-empty' and is interspersed among genes ranging

in size from 100 bp to about 50 kb (TAGI, 2000). The human genome is relatively large and the partitioning into gene-rich and gene-poor regions is more distinct (IHGSC, 2001; Venter *et al.*, 2001). The 30,000 genes with an average size of 27kb account for about 25% of the human genome. The remaining 75% of the genome consists of retrotransposon-like repetitive DNA interspersed among genes in blocks of varying sizes. Cytologically, the human chromosomes show G-, R-, and T-bands that are structurally different. The T-bands are very gene-rich and the G-bands are gene-poor. As shown in Fig. 3b, the gene-density in chromosome *7* ranged from zero to 30 genes/Mb (http://www.ncbi.nlm.nih.gov). On other chromosomes, the gene-density in some regions was as high as 60 genes/Mb. The wheat genome is about five times larger than that of the human. Although not well characterized at sequence level, the 'gene-empty' regions of wheat have been identified and are shown in Figs. 2 and 3c. Out of the 754Mb of chromosome *1*, 635Mb is 'gene-empty', with the largest gene-empty block of 192Mb around the centromere. In summary, the average and the largest size of 'gene-empty' regions among eukaryotes seem to increase with an increase in the genome size.

2.4. The Gene-Rich and Gene-Empty Regions within a Genome

In the eukaryotes studied so far, genes are interspersed by 'gene-empty' regions, which vary in number and size in proportion to their genome size. About 82% of the yeast genome contains genes (Mewes *et al.*, 1997) and about 90% of the genes are present in clusters of size ranging from 10 to 450kb (Bussey *et al.*, 1997; Dujon *et al.*, 1997). An interspersed 'gene-empty' block can be up to 20kb in size. In Arabidopsis, about 45% of the genome contains genes. The remaining 'gene-empty' part is interspersed among genes, the gene empty block having an average size of ~2.4kb and ranging up to ~4Mb (TAGI, 2000). The largest genome sequenced so far is human genome, where the genes account for only ~25% of the genome. The size of individual 'gene-empty' blocks in this genome can be up to 30Mb with an average of ~80kb (IHGSC, 2001).

Among grasses, rice and barley are diploids belonging to the same family Poaceae. The two genomes however differ in size by 12-fold. In rice, a gene is expected to occur every 16kb, assuming 25,000 randomly distributed genes. Mapping of 6591 unique ESTs on yeast artificial chromosome (YAC) contigs of rice revealed 300-500kb long genomic regions of high gene-density with a gene every ~4kb (Fig. 3a) (Wu *et al.*, 2002; RGP web page-

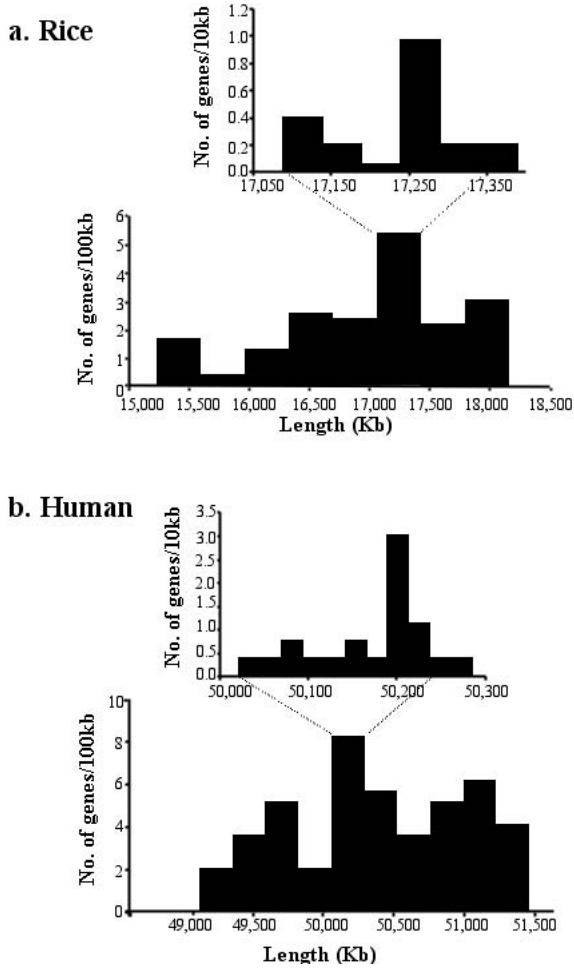


Figure 4. Detailed gene distribution within small regions of rice and human.
The data was drawn from http://rgp.dna.affrc.go.jp/Publicdata.html for rice and from http://www.ncbi.nlm. nih.gov for human, respectively.

http://rgp.dna.affrc.go.jp/Publicdata.html). These regions were more prevalent in the distal parts of the chromosomes and the gene-density even within these regions was highly variable (Fig. 4a). The size of the individual interspersing 'gene-empty' regions ranged from 1 to 30kb. Clustering of genes is more pronounced in larger genomes such as barley, where the presumed 30,000 genes are expected to occur one every 166kb. However, analysis of 1.1Mb sequence from 12 different gene-containing regions showed a gene every 21kb (Panstruga *et al.*, 1998; Feuillet and Keller, 1999; Shirasu *et al.*, 2000; Dubcovsky *et al.*, 2001; Fu *et al.*, 2001, 2002; Rostoks *et*

al., 2002; Wei *et al.*, 2002). The gene distribution within these regions was also highly uneven, with the interspersing 'gene-empty' blocks ranging in size from 0.8kb to 94kb. Like rice genome, localized regions of high gene-density (a gene every 4.4kb) were also observed in barley. The average size of the high gene-density regions in barley seems to be smaller than that of rice. Among the 12 regions sequenced, the largest contig is 261kb around the *Mla* locus (Wei *et al.*, 2002). A 130kb region of this contig contained 23 genes with an average gene-density of one gene every 5.6kb. Even within this region, the size of 'gene-empty' regions ranged from few hundred bps to 11kb. The second region of high gene-density was ~40kb long and contained 10 genes (a gene every 4kb). The largest 'gene-empty' region was 5.5kb. Another 60kb barley contig spanned a 32kb region around the bronze (*bz*) locus with a gene every 3.2kb (Fu *et al.*, 2001). Similar observations were made in other grasses with large genomes, such as *Triticum monococcum*, *Aegilops tauschii*, and maize (Rahman *et al.*, 1997; Feuillet and Keller, 1999; Tikhonov *et al.*, 1999; Wicker *et al.*, 2001).

In summary, gene distribution is uneven on the chromosomes of probably all eukaryotes. Genome expansion among eukaryotes can be partly explained by the increase in the number of genes due to increased biological complexity and polyploidy. Beside polyploidy, size difference among genomes of similar complexity in several groups of plants is mainly due to differential invasion and amplification of retroelements, with genes still occurring in clusters.

As the size of the genome changes from the smaller (Arabidopsis and rice) to the larger genome (wheat, barley, and maize), an overall reduction in the size of the gene-cluster and an expansion in the size of interspersing 'gene-empty' regions is witnessed. For instance, the average size of gene clusters in rice is ~300kb, while in wheat and barley, it seems to be smaller than 50kb. The 'gene-empty' regions seem to amplify preferentially, as is obvious in wheat, where large, essentially 'gene-empty' blocks of up to ~192Mb are common (Figs. 2 and 3c). The current demarcation of the gene-rich and gene-poor regions has perhaps identified these larger 'gene-empty' regions of wheat genome. Taking polyploidy into account, the 30% gene-rich part of the genome is still ~4 times larger than the entire rice genome (Erayman *et al.*, 2003). However, gene distribution within the currently defined GRRs of wheat would probably still 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. In this respect, wheat genome resembles barley genome (see above).

2.5. The Gene-Empty Regions in a Genome

The 'gene-empty' regions of higher eukaryotic genomes are mainly comprised of retrotransposons and pseudogenes. The proportion of the retrotransposons is significantly higher than pseudogenes, especially in the larger genomes.

2.5.1 Retrotransposons

Retrovirus invasion and its subsequent amplification have played a major role in expansion of eukaryotic genomes. The retroelements, once integrated, are usually not excised out, but can amplify via an RNA intermediate, i.e. reverse transcribed before integration (Finnegan, 1992; Grandbastien, 1992; Kubis *et al.*, 1998). The retroelements can be with or without long-terminal repeats (LTRs). The non-LTR retrotransposons can either be long (long interspersed nuclear elements, LINEs) or short (short interspersed nuclear elements, SINEs). In plant genomes, the repeated DNA mainly consists of LTR-type of retrotransposons, and the non-LTR type of retrotransposons form a very small proportion (Flavell *et al.*, 1994; Kumar and Bennetzen, 1999; see also Chapter 4 by Schulman et al. in this book). In contrast to this, in human genome, the non-LTR type LINEs and SINEs form majority of the repetitive fraction (IHGSC, 2001). In smaller genomes such as Arabidopsis, retrotransposons make up a very small percentage (5%) of the genome and are mostly present around the centromeres (TAGI, 2000). In larger genomes such as maize, on the other hand, the repetitive fraction consisting mainly of retrotransposons, ranges from 64% to 73% (Meyers *et al.*, 2001). In most cases, the LTR-type of retrotransposons seem to transpose preferentially into the gene-poor regions or the regions flanking the gene clusters, as witnessed in maize genomic regions around the bronze locus (*bz*), *Adh1* locus, and the locus encoding 22kDa zein (SanMiguel *et al.*, 1996; Llaca and Messing, 1998; Fu *et al.*, 2001). The miniature inverted repeat type of transposable elements (MITEs) are more common in the gene-rich regions, as observed in the *bz* locus (Fu *et al.*, 2002).

2.5.1. Duplications and Pseudogenes

In addition to retrotransposons, gene duplication has also played an important role in eukaryotic genome expansion. Duplications may involve a part of a genome (segmental duplication) or a whole genome (polyploidy), or may be localized involving one or few genes (unequal crossovers or other mechanisms). About 13% of yeast and ~60% of Arabidopsis genes are

duplicated (Henikoff *et al.*, 1997; TAGI, 2000). In human, 3,522 unique genes are present with an average of 5.8 copies to make 20,428 genes. These genes are present on 1,077 blocks carrying at least three genes per block with an average of 19 genes per block (Venter *et al.*, 2001). Out of 1,077 blocks, 159 contain only three genes, 137 contain four, and 781 contain five or more genes.

The human genome is believed to have undergone multiple cycles of polyploidization during cordate evolution (Henikoff *et al.*, 1997). The segmental duplications of varying sizes may be the footprints of ancient polyploidization events. Following polyploidization, the duplicated genomes may either be kept relatively intact as in wheat or reshuffled as observed in maize, rice, and Arabidopsis. Wheat is a relatively young polyploid (Kihara, 1944; McFadden and Sears, 1946) and the duplicated genomes over time may be reshuffled and modified beyond recognition. DNA elimination after polyploidization may be another mechanism of changing the duplicated genomes (Ozkan *et al.*, 2001; Shaked *et al.*, 2001). The duplicated genes may get silenced, retain the same function, acquire specificity to different tissue or environments; or may acquire an all-together different function. Examples do exist for all these possibilities. A good example of silenced duplicated genes is that of resistance gene analogs (RGA). About 600 genomic RGAs structurally resembling cloned functional resistant genes have been characterized from 20 plant species using conserved motif primers (Kanazin *et al.*, 1996; Leister *et al.*, 1996; Yu *et al.*, 1996; Feuillet *et al.*, 1997). Only one (*Dm3*) of these RGAs has been shown to be functional (Shen *et al.*, 2002). These gene-like non-functional copies seem to be prevalent in all eukaryotes. Even in the smallest eukaryote genome, *S. cerevisiae*, ~0.7% are pseudogenes (Mewes *et al.*, 1997). Careful analysis of the human genome sequence revealed 2,909 processed pseudogenes (Venter *et al.*, 2001). The exact number of pseudogenes, however, is difficult to estimate because some genes may be expressing for a very short period of time, making very few and highly unstable transcripts, or may be functional under extreme and unusual conditions. Assuming that such genes are rare, the large gap that exists between the number of functionally characterized genes and that predicted by *in silico* approaches, may be attributed to the presence of pseudogenes.

3. DISTRIBUTION OF RECOMBINATION RATES WITHIN A GENOME

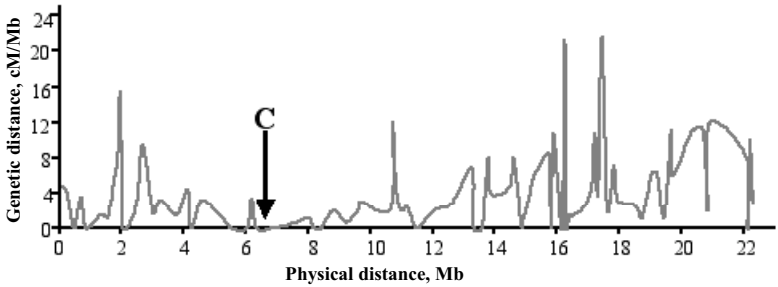
Meiotic recombination is an important cellular phenomenon that, in addition to proper chromosome segregation, plays a pivotal role in evolution.

Distribution of recombination has been observed to be highly uneven on the chromosomes of all eukaryotes studied so far. The current state of knowledge about relationship of the extent and distribution of recombination rates with chromosome size, distribution of genes, and chromosomal landmarks is discussed in the following sections.

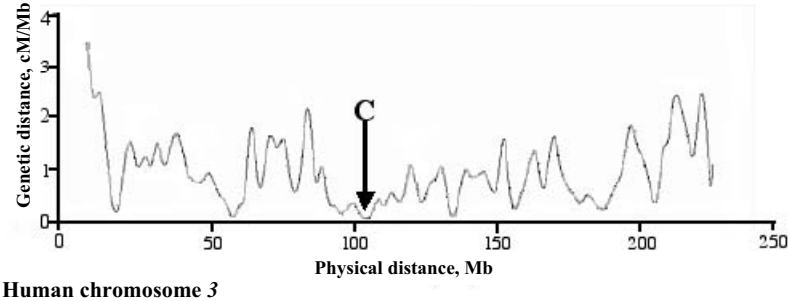
3.1. Extent And Distribution Of Recombination Rates

The extent of recombination per chromosome seems to be independent of chromosome size. For instance, the recombination values for rice and human are 127 and 154cM/chromosome, respectively (Wu *et al.*, 2002; http://rgp.dna.affrc.go.jp/Publicdata.html; http://www.ncbi.nlm.nih.gov), although the average size of a human chromosome is about four-times the size of a rice chromosome. Similarly, wheat chromosomes are >1000-fold larger relative to yeast chromosomes, but average recombination per chromosome is about the same (Gill *et al.*, 1991; Erayman *et al.*, 2003), the average recombination in yeast being ~262cM/chromosome, with an average chromosome size of ~0.75Mb (http://genome-www.stanford.edu/Sacchar omyces). In a mapping populations derived from a cross between two related species, the extent of recombination may also depend upon the evolutionary distance between the two parents of such a mapping population. It is for this reason that lower rates of recombination per chromosome have been reported in inter-specific mapping populations that were developed to maximize polymorphism. Pairing and subsequent recombination between chromosomes of different species is significantly lower than those in inbreds (Kimber, 1970; Dvorak and Chen, 1984).

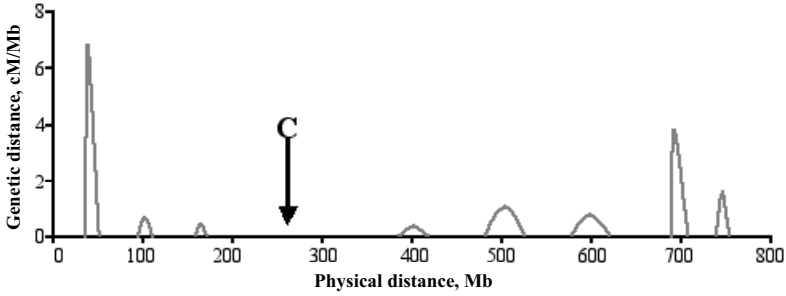
Although recombination rates are independent of chromosome size, as discussed above, uneven distribution of recombination has been observed along chromosomes of all eukaryotes studied so far (Dvorak and Chen, 1984; Bollag *et al.*, 1989; Curtis and Lukaszewski, 1991; Gill *et al.*, 1993; Lichten and Goldman, 1995; Gill *et al.*, 1996a, b; Tranquilli *et al.*, 1999; Sandhu and Gill, 2002a, b; Akhunov *et al.*, 2003). Distribution of recombination is best studied on yeast chromosomes, in which at 1kb resolution, the average frequency of recombination is about 0.35cM per kb with a range of 0 to 2cM per kb (http://genome-www.stanford.edu/Saccharomyces). Among cereals, the rice genome is one of the smallest, and average frequency of recombination is about 0.003cM per kb with a range of 0 to 0.06cM (Fig. 5a) (http://rgp.dna.affrc.go.jp/Publicdata.html). Non-recombinogenic regions were observed in yeast as well as in rice, but the highest rate of recombination for a region appears to be ~35-fold less in rice. It may be because the rice estimates were at ~400kb resolution, whereas the



a. Rice chromosome *10*



b. Human chromosome *3*



c. Wheat chromosome *1*

Figure 5. Distribution of recombination along the chromosomes of rice, human, and wheat. The data used was the same as that of Fig. 3. The centromere is marked by an arrow and shown as 'C'.

recombination hotspots may only be <1kb in length. Due to averaging over a larger region, recombination in hotspots may appear less in rice compared to that in yeast. Average recombination frequency in human is 0.001cM per kb with a range from 0 to 0.004cM per kb (Fig. 5b) (Kong *et al.*, 2002). As mentioned for rice, the difference in recombination between 'hot' and 'cold' spots is not as dramatic as in yeast, probably because the resolution was at 3Mb. Similarly in wheat, at a resolution of ~7Mb the average recombination

frequency is 0.0003cM per kb with a range from 0 to 0.007cM per kb (Figs. 2 and 5c). Distribution of recombination rates on wheat chromosomes is shown in Fig. 2 (Sandhu and Gill, 2002a). Recombination occurred in six of the eight small gene rich regions encompassing ~13% of the chromosome and essentially no recombination was observed in the remaining 87% of the chromosome (Sandhu and Gill, 2002a; Akhunov *et al.*, 2003).

Rate of recombination was very different even among the six regions, ranging from 159kb to 2917kb per cM, but most of the recombination occurred in the distal 50% of the chromosome. These estimates were obtained by averaging recombination over 7 to 35Mb regions (Fig. 2). A comparison of recombination distribution on wheat chromosomes with those of rice and human is shown in Fig. 5. Peaks indicating recombination hotspots are sharper in rice, probably due to a higher resolution in estimation. The corresponding peaks in human and more so in wheat, have a wider base due to a lower resolution. Similarly the recombination cold spots are much larger in wheat compared to rice and human. As in wheat, in rice also, more recombination is observed in the distal regions than that in the proximal regions of chromosomes. This difference between distal and proximal regions of chromosome is not as dramatic in human as in wheat and rice.

3.2. Recombination Coldspots and Hotspots

Chromosomal regions, with essentially no recombination (coldspots), have also been observed in all eukaryotes. Regions lacking recombination measure upto 15kb in yeast (http://genome-www.stanford.edu/Saccharomyces), upto a few Mb in human, and up to 190Mb in wheat (Fig. 2 and Figs. 5b, c). Small regions of very high recombination (hotspot) are also observed in most eukaryotes (Fig. 5). These hotspots can be only a few hundred base pairs in size. There is a major hotspot around *PHO8* gene of yeast located on the distal end of chromosome IV, where recombination is 2cM/kb (http://genome-www.stanford.edu/Saccharomyces). Similar hotspots have been reported in higher plants also. A 194bp region around *wx* locus of rice has a recombination rate of 0.06cM/kb that is 20-fold higher than the genome average (Inukai *et al.*, 2000). Similarly, a 377bp region of maize has a recombination rate of 0.02cM/kb that is about 30 times higher than the genome average (Civardi *et al.*, 1994; Yao *et al.*, 2002). Similar regions of localized high recombination have been reported in barley, wheat, maize, rice, and other plants (Lahaye *et al.*, 1998; Wei *et al.*, 1999; Stein *et al.*, 2000; Tarchini *et al.*, 2000; Brueggeman *et al.*, 2002). Recombination within the

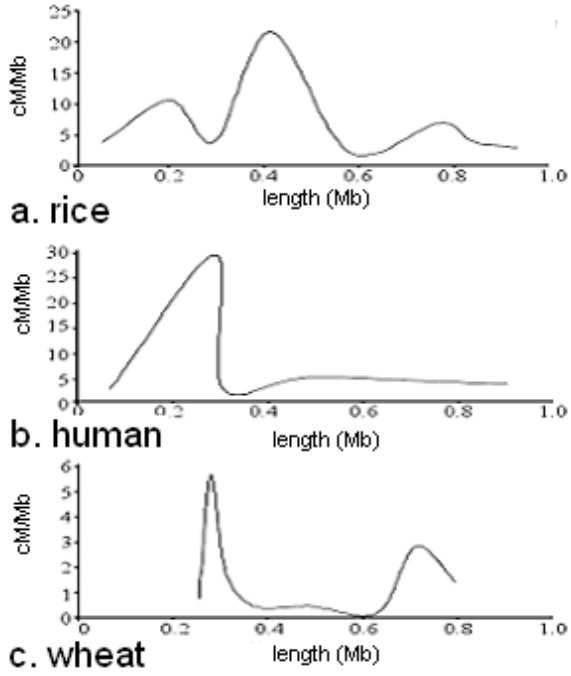


Figure 6. Detailed distribution of recombination within small regions of rice, human, and barley. The data was used from Hubert *et al.*, 1994; Wei *et al.*, 2002; Wu *et al.*, 2002, respectively.

bronze (*bz*) locus of barley is 0.07cM/kb as compared to the genome average that is 0.0002cM/kb (Dooner, 1986).

In addition to extreme cases of cold- and hot- spots, recombination rates are highly variable even within small regions of chromosomes. Variation in recombination rate within - 1Mb region of rice, human, and barley is shown in Fig. 6 (http://rgp.dna.affrc.go.jp/Publicdata.html; http://www.ncbi. nlm. nih.gov; Wei *et al.*, 1999) and can be more than 20-fold within the same genomic region. Similar comparison within the 1Mb region flanking the *Mla* locus of barley shows ~10 fold difference (Wei *et al.*, 1999). Differences in recombination have also been observed in genomic regions measuring less than 1Mb. For instance, a 13-fold difference in recombination rate was observed within the 300kb barley region encompassing *Rpg1*, when studied at 50kb interval (Brueggeman *et al.*, 2002). The 470kb segment from the long arm of chromosome 3 of maize, encompassing the *a1*(anthocyaninless1) and *sh2* (shrunken2) loci, separated by 140kb, has also been extensively studied for the distribution of recombination (Civardi *et al.*,

1994, Yao *et al.*, 2002). It was found that a1 locus was approximately seven times more recombinogenic as compared to the neighboring 140kb region between *a1* and *sh2*. Further characterization of the 140kb *a1-sh2* interval revealed that rate of recombination is different even among small segments within this region (Yao *et al.*, 2002). For instance, recombination in a 0.38kb region was about seven times higher than the flanking regions.

3.3. Genes vs Recombination

The relationship between distribution of genes and recombination is easier to establish in wheat where large 'gene empty' blocks have been identified and demarcated (Fig. 2). Essentially no recombination was observed in the 'gene-empty' regions (Sandhu and Gill, 2002a; Erayman et al, 2003), and all recombination hotspots observed so far are found around GRRs, although different GRRs may differ in rates of recombination. For instance, recombination among the eight GRRs of wheat group *1* chromosomes varied from 0.2 to 70cM/Mb (Fig. 2). Part of this difference may be due to the effect of centromere, since lower rate of recombination is observed in the proximal GRRs. Even at a higher resolution, GRRs showed significant differences for the rate of recombination. A general correlation between the distribution of genes and recombination was observed in rice and human genomes. In many eukaryotes, genes and GRRs have been identified that show essentially no recombination. One such example is a 240kb region spanning the *Mla* genes of barley where no recombination occurred (Wei *et al.*, 1999). Between these extremes, genes and GRRs have been observed, which exhibit a wide range of recombination rates. For instance, recombination within a 1Mb region of rice with a uniform gene distribution ranged from 0 to 1.5cM, when measured at 150kb inerval (http://rgp.dna.affrc.go.jp/Publicdata.html).

Recombination rates also vary within a gene and on the state of expression of a gene. It has been proposed that promoter sequences are the sites for recombination initiation (Cao *et al.*, 1990; Wu and Lichten, 1994) and that the recombination frequency is often highest at the 5' end and the lowest at the 3' end, as witnessed in *ARG4* and *HIS4* genes of yeast (Nicolas *et al.*, 1989; White *et al.*, 1993). It was also found that deletions in the upstream sequences of *ARG4* and *HIS4* led to significant reduction in recombination. Because of the open chromatin structure, the promoter regions bound by transcription factors may be more prone to double strand breaks, and thus may be highly recombinogenic (White *et al.*, 1993; Wu and Lichten, 1994). Preliminary investigations also suggest a correlation between gene expression and recombination, although it needs to be confirmed. For instance,

recombination was induced due to activation of gene expression brought about by altered acetylation pattern of a 'silent' *mat2-mat3* interval (Grewal *et al.*, 1998).

3.4. Recombination Around Centromere and Telomere

Recombination around centromeres of eukaryotes is suppressed (Lambie and Roeder, 1986; Tanksley *et al.*, 1992; Mckee *et al.*, 1993; Lichten and Goldman, 1995; Laurent *et al.*, 1998). Even the small centromeres of *S. cerevisiae* suppress recombination. A 600bp fragment containing centromere III, when integrated at a new site on the distal part of the chromosome, accounted for about a five-fold decrease in recombination (Lambie and Roeder, 1986; Lambie and Roeder, 1988). The effect of centromere on recombination is more dramatic in organisms with larger genomes such as wheat. An approximately 190Mb region around the centromere of group *1* chromosomes shows essentially no recombination (Fig. 2). Blocks of similar size around centromeres of other wheat and barley chromosomes also showed severe suppression (Kunzel *et al.*, 2000; Erayman *et al.*, 2003). Non-recombinogenic megabase-sized regions have also been observed around all centromeres of rice and human chromosomes (Wu *et al.*, 2002; IHGSC, 2001). An approximately 10Mb region around the centromere of human chromosome *5* showed no recombination (Puechberty *et al.*, 1999). In wheat, the suppressive effect on recombination extends to ~35% of the chromosomal region surrounding centromeres. The effect of centromere in human and rice seems to be more localized (Fig. 5).

The recombination in telomeric regions is also suppressed as shown in yeast and tomato (http://genome-www.stanford.edu/Saccharomyces; Tanksley *et al.*, 1992). For instance, double-strand break survey of ~25kb subtelomeric region of yeast showed very little recombination (Klein *et al.*, 1996; Ross *et al.*, 1996). Genetic markers mapped every ~7kb showed a gradient of recombination decreasing towards the telomere (Su *et al.*, 2000). Up to 10-fold lower rate of recombination was also observed in the telomeric regions of tomato and potato (Tanksley *et al.*, 1992). The effect of telomeres on recombination is not well known in other eukaryotes.

3.5. Other Factors Affecting Recombination Rates

Additional factors, which affect recombination include differences in the

nutritional state, sex, cytoplasm, chromosomal arms, etc. (Laurie and Hulten, 1985; Bondareva, 2000; Kong *et al.*, 2002). For instance, difference in recombination rate between human males and females is well known. The high-density microsatellite marker-based genetic linkage maps of human also showed that recombination is 1.6 times more frequent in the females compared to males (Dib *et al.*, 1996; Kong *et al.*, 2002). Even the location of crossovers is different between males and females. Crossovers are more frequent near the centromeres in females (Kong *et al.*, 2002).

4. SUMMARY AND OUTLOOK

An overview presented in this chapter suggests that among all eukaryotes, and particularly, in cereal genomes, the genes as well as recombination rates are distributed unevenly not only within a genome but also within individual chromosomes. It has been demonstrated that within a genome, there are gene rich regions (GRRs) and gene empty regions, and that there are also recombination hot spots and recombination cold spots. A survey of genomes of higher plants, particularly those of cereals, also suggests that there is a reduction in the average size of the gene cluster, with an increase in the size of genome. For different regions of the genome, a correlation also exists in the recombination rates and the density of genes, although recombination rates may also differ in different GRRs. Within individual chromosomes, the gene density and recombination rates increase as we move away from the gene poor and low recombination regions around centromere, although the recombination is suppressed again at the end of chromosomes, i.e. in the telomeric regions.

5. REFERENCES

Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D., Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F., *et al.* The genome sequence of *Drosophila melanogaster*. Science 2000; 287:2185-2195

Akhunov E.D., Goodyear, A.W., Geng, S., Qi, L.-L., Echalier, B., Gill, B.S., Miftahudin, Gustafson, J.P., Lazo, G., Chao, S., *et al.* The organization and rate of evolution of wheat genomes are correlated with recombination rates along chromosome arms. Genome Res 2003; 13: 753-763

Arumuganathan K., Earle E.D. Nuclear DNA content of some important plant species. Plant Mol Biol Rep 1991; 9: 208-218

Barakat A., Carels N., Bernardi G. The distribution of genes in the genomes of Gramineae. Proc Natl Acad Sci USA 1997; 94: 6857-6861

Bollag R.J., Waldman A.S., Liskay R.M. Homologous recombination in mammalian cells. Annu Rev Genet 1989; 23: 199-225

Bondareva S.N. Thesis: Extent of recombination and its detection using F_2 vs. backcross population; Dept of Agronomy. University of Nebraska-Lincoln, USA, 2000

Brueggeman R., Rostoks N., Kudrna D., Kilian A., Han F., Chen J., Druka A., Steffenson B., Kleinhofs A. The barley stem rust-resistance gene *Rpg1* is a novel disease-resistance gene with homology to receptor kinases. Proc Nat Acad Sci USA 2002; 99: 9328-9333

Bussey H, Storms R.K., Ahmed A., Albermann K., Allen E., Ansorge W., Araujo R., Aparicio A., Barrell B., Badcock K., *et al.* The nucleotide sequence of *Saccharomyces cerevisiae* chromosome XVI. Nature 3 1997; 87: 103-105

Cao L., Alani E., Kleckner N. A pathway for generation and processing of double-strand breaks during meiotic recombination in *S. cerevisiae*. Cell 1990; 61: 1089-1101

Chen M., SanMiguel P., Bennetzen J.L. Sequence organization and conservation in *sh2/a1*-homologous regions of sorghum and rice. Genetics 1998; 148: 435-443

Civardi L., Xia Y., Edwards K.J., Schnable P.S., Nikolau B.J. The relationship between genetic and physical distances in the cloned *a1-sh2* interval of the *Zea mays* L. genome. Proc Natl Acad Sci USA 1994; 91: 8268-8272

Curtis C.A., Lukaszewski A.J. Genetic linkage between C-bands and storage protein genes in chromosome 1B of tetraploid wheat. Theor Appl Genet 1991; 81: 245-252

Dib C., Faure S., Fizames C., Samson D., Drouot N., Vignal A., Millasseau P., Marc S., Hazan J., Seboun E., Lathrop M., Gyapay G., Morissette J., Weissenbach J. A comprehensive genetic map of the human genome based on 5,264 microsatellites. Nature 1996; 380: 152-154

Dooner H.K. Genetic fine structure of *bronze* locus in maize. Genetics 1986; 113: 1021-1036

Dubcovsky J., Ramakrishna W., SanMiguel P.J., Busso C.S., Yan L., Shiloff B.A., Bennetzen J.L. Comparative sequence analysis of colinear barley and rice bacterial artificial chromosomes. Plant Physiol 2001; 125: 1342-1353

Dujon B., Albermann K., Aldea M., Alexandraki D., Ansorge W., Arino J., Benes V., Bohn C., Bolotin-Fukuhara M., Bordonne R., *et al.* The nucleotide sequence of *Saccharomyces cerevisiae* chromosome XV. Nature 1997; 387: 98-102

Dvorak J., Chen K.C. Phylogenetic relationships between chromosomes of wheat and chromosome 2E of *Elytrigia elongata*. Can Jour Genet Cytol 1984; 26: 128-132

Endo T.R., Gill B.S. The deletion stocks of common wheat. J Hered 1996; 87: 295-307

Erayman M., Sandhu D., Sidhu D., Dilbirligi M., Gill K.S. Gene-rich regions of wheat. Plant Cell 2003; (submitted)

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

Feuillet C., Schachermayr G., Keller B. Molecular cloning of a new receptor-like kinase gene encoded at the Lr10 disease resistance locus of wheat. Plant J 1997; 11: 45-52

Finnegan D. Transposable elements. Curr Opin Genet Dev 1992; 4: 838-844

Flavell A.J., Pearce S., Kumar A. Plant transposable elements and the genome. Curr Opin Genet Dev 1994; 4: 838-844

Fu H., Park W., Yan X., Zheng Z., Shen B., Dooner H.K. The highly recombinogenic *bz* locus lies in an unusually gene-rich region of the maize genome. Proc Natl Acad Sci USA 2001; 98: 8903-8908

Fu H., Zheng Z., Dooner H.K. Recombination rates between adjacent genic and retrotransposon regions in maize vary by 2 orders of magnitude. Proc Natl Acad Sci USA 2002; 99: 1082-1087

Gill B.S., Friebe B., Endo T.R. Standard karyotype and nomenclature system for description of chromosome bands and structural aberrations in wheat (*Triticum aestivum*). Genome 1991; 34: 830-839

Gill K.S., Gill B.S., Endo T.R. A chromosome region-specific mapping strategy reveals gene rich telomeric ends in wheat. Chromosoma 1993; 102: 374-381

Gill K.S., Gill B.S., Endo T.R., Boyko E. Identification and high-density mapping of gene-rich regions in chromosome group 5 of wheat. Genetics 1996a; 143: 1001-1012

Gill K.S., Gill B.S., Endo T.R., Taylor T. Identification and high-density mapping of gene-rich regions in chromosome group 1 of wheat. Genetics 1996b; 144: 1883-1891

Goff S.A., Ricke D., Lan T.H., Presting G., Wang R.L., Dunn M., Glazebrook J., Sessions A., Oeller P., Varma H., *et al.* A draft sequence of the rice genome (*Oryza sativa* L. ssp. japonica). Science 2002; 296: 92-100

Goffeau A., Barrell B.G., Bussey H., Davis R.W., Dujon B., Feldmann H., Galibert F., Hoheisel J.D., Jacq C., Johnston M., *et al.* Life with 6000 genes. Science 1996; 274:562-567

Grandbastien M. Retroelements in higher plants. Trends Genet 1992; 8: 103-108

Grewal S.I., Bonaduce M.J., Klar A.J. Histone deacetylase homologs regulate epigenetic inheritance of transcriptional silencing and chromosome segregation in fission yeast. Genetics 1998; 150: 563-576

Henikoff S., Greene A.E., Pietrokovski S., Bork P., Attwood T.K., Hood L. Gene families: The taxonomy of protein paralogs and chimeras. Science 1997; 278: 609-614

Hubert R., MacDonald M., Gusella J., Arnheim N. High resolution localization of recombination hot spots using sperm typing. Nat Genet 1994; 7: 420-424

IHGSC. Initial sequencing and analysis of the human genome. Nature 2001; 409: 860-921

Inukai T., Sako A., Hirano H.Y., Sano Y. Analysis of intragenic recombination at *wx* in rice: Correlation between the molecular and genetic maps within the locus. Genome 2000; 43: 589-596

Kanazin V., Marek L.F., Shoemaker R.C. Resistance gene analogs are conserved and clustered in soybean. Proc Natl Acad Sci USA 1996; 93: 11746-11750

Kihara H. Discovery of the DD-analyser, one of the ancestors of *vulgare* wheats. Agric Hortic 1944; 19: 889-890

Kimber G. The relationships of single alien chromosomes to the homoeologous groups of *Triticum aestivum*. In: Finlay KW, Shephard KW (eds) *Proceedings of the 3rd International Wheat Genetics Symposium*, (pp. 62-68). New York, NY, 1970

Klein S., Zenvirth D., Dror V., Barton A.B., Kaback D.B., Simchen G. Patterns of meiotic double-strand breakage on native and artificial yeast chromosomes. Chromosoma 1996; 105: 276-284

Kong A., Gudbjartsson D.F., Sainz J., Jonsdottir G.M., Gudjonsson S.A., Richardsson B., Sigurdardottir S., Barnard J., Hallbeck B., Masson G., *et al.* A high-resolution recombination map of the human genome. Nat Genet 2002; 31: 241-247

Kubis S., Schmidt T., Heslop-Harrison J. Repetitive DNA elements as a major component of plant genomes. Ann Botany 1998; 82: 45-55

Kumar A., Bennetzen J.L. Plant retrotransposons. Annu Rev Genet 1999; 33: 479-532

Kunzel G., Korzum L., Meister A. Cytologically integrated physical restriction fragment length polymorphism maps for the barley genome based on translocation breakpoints. Genetics 2000; 154: 397-412

Lahaye T., Shirasu K., Schulze-Lefert P. Chromosome landing at the barley *Rar1* locus. Mol Gen Genet 1998; 260: 92-101

Lambie E.J., Roeder G.S. Repression of meiotic crossing over by a centromere (CEN3) in *Sacchromyces cerevisiae.* Genetics 1986; 114: 769-789

Lambie E.J., Roeder G.S. A yeast centromere acts in cis to inhibit meiotic gene conversion of adjacent sequences. Cell 1988; 52: 863-873

Laurent A.M., Puechberty J., Prades C., Roizes G. Informative genetic polymorphic markers within the centromeric regions of human chromosome 17 (D17S2205) and 11 (D11S4975). Genomics 1998; 52: 166-172

Laurie D.A., Hulten M.A. Further studies on bivalent chiasma frequency in human males with normal karyotypes. Ann Hum Genet 1985; 49: 189-201

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

Lichten M., Goldman A.S.H. Meiotic recombination hotspots. Annu Rev Genet 1995; 29: 423-444

Llaca V., Messing J. Amplicons of maize zein genes are conserved within genic but expanded and constricted in intergenic regions. Plant J 1998; 15: 211-220

McFadden E.S., Sears E.R. The origin of *Triticum spelta* and its free-threshing hexaploid relatives. J Hered 1946; 37: 81-89

Mckee B.D., Lumsden S.E., Das S. The distribution of male meiotic pairing sites on chromosome 2 of *Drosophila melanogaster:* meiotic pairing and segregation of 2-Y transpositions. Chromosoma 1993; 102: 180-194

Mewes H.W., Albermann K., Bahr M., Frishman D., Gleissner A., Hani J., Heumann K., Kleine K., Maierl A., Oliver S.G., Pfeiffer F., Zollner A. Overview of the yeast genome. Nature 1997; 387: 7-65

Meyers B.C., Tingey S.V., Morgante M. Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome. Genome Res 2001; 11: 1660-1676

Nicolas A., Treco D., Svostak S., Svostak J.W. An initiation site for meiotic gene conversion in the yeast *Saccharomyces cerevisiae*. Nature 1989; 338: 35-39

Ozkan H., Levy A., Feldman M. Allopolyploidy-induced rapid genome evolution in the wheat (*Aegilops*-Triticum) group. Plant Cell 2001; 13: 1735-1747

Panstruga R., Buschges R., Piffanelli P., Schulze-Lefert P. A contiguous 60kb genomic stretch from barley reveals molecular evidence for gene islands in a monocot genome. Nucleic Acids Res 1998; 26: 1056-1062

Puechberty J., Laurent A.M., Gimenez S., Billault A., Laurent M.E.B., Calenda A., Marcais B., Prades C., Loannou P., Yurov Y., Roizes G. Genetic and physical analyses of the centromeric and pericentromeric regions of human chromosome 5: Recombination across 5cen. Genomics 1999; 56: 274-287

Rahman S., Abrahams S., Abbott D., Mukai Y., Samuel M., Morell M., Appels R. A complex arrangement of genes at a starch branching enzyme I locus in the D-genome donor of wheat. Genome 1997; 40: 465-474

Ross L., Maxfield R., Dawson D. Exchanges are not equally able to enhance meiotic chromosome segregation in yeast. Proc Natl Acad Sci USA 1996; 93: 4979-4983

Rostoks N., Park Y.J., Ramakrishna W., Ma J., Druka A., Shiloff B.A., SanMiguel P.J., Jiang Z., Brueggeman R., Sandhu D., Gill K.S., Bennetzen J.L., Kleinhofs A. Genomic sequencing reveals gene content, genomic organization, and recombination relationships in barley. Funct. Integr. Genomics 2002; 2: 51-59

Sandhu D., Champoux J.A., Bondareva S.N., Gill K.S. Identification and physical localization of useful genes and markers to a major gene-rich region on wheat group *1S* chromosomes. Genetics 2001; 157: 1735-1747

Sandhu D., Gill K.S. Gene-containing regions of wheat and the other grass genomes. Plant Physiol 2002a; 128: 803-811

Sandhu D., Gill K.S. Structural and functional organization of the '1S0.8 gene-rich region' in the Triticeae. Plant Mol Biol 2002b; 48: 791-804

SanMiguel P., Tikhonov A., Jin Y.K., Motchoulskaia N., Zakharov D., Melake-Berhan A., Springer P.S., Edwards K.J., Lee M., Avramova Z., Bennetzen J.L. Nested retrotransposons in the intergenic regions of the maize genome. Science 1996; 274: 765-768

Shaked H., Kashkush K., Ozkan H., Feldman M., Levy A. Sequence elimination and cytosine methylation are rapid and reproducible responses of the genome to wide hybridization and allopolyploidy in wheat. Plant Cell 2001; 13: 1749-1759

Shirasu K., Schulman A.H., Lahaye T., Schulze-Lefert P. A contiguous 66-kb barley DNA sequence provides evidence for reversible genome expansion. Genome Res 2000; 10: 908-915

Shen K.A., Chin D.B., Arroyo-Garcia R., Ochoa O.E., Lavelle D.O., Wroblewski T., Meyers B.C., Michelmore R.W. Dm3 is one member of a large constitutively expressed family of nucleotide binding site-leucine-rich repeat encoding genes. Mol Plant Microbe Interact 2002; 15:251-261

Spring J. Vertebrate evolution by interspecific hybridization--are we polyploid? FEBS letts 1997; 400: 2-8

Stein N., Feuillet C., Wicker T., Schlagenhauf E., Keller B. 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.). Proc Natl Acad Sci USA 2000; 97: 13436-13441

Su Y., Barton A.B., Kaback D.B. Decreased meiotic reciprocal recombination in subtelomeric regions in *Saccharomyces cerevisiae*. Chromosoma 2000; 109: 467-475

Tanksley S.D., Ganal M.W., Prince J.P., de Vicente M.C., Bonierbale M.W., Broun P., Fulton T.M., Giovannoni J.J., Grandillo S., Martin G.B., *et al.* High density molecular linkage maps of the tomato and potato genomes. Genetics 1992; 132: 1141-1160

Tarchini R., Biddle P., Wineland R., Tingey S., Rafalski A. The complete sequence of 340kb of DNA around the rice Adh1-Adh2 region reveals interrupted colinearity with maize chromosome 4. Plant Cell 2000; 12: 381-391

The Arabidopsis genome initiative (TAGI) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 2000; 408: 796-815

The *C. elegans* sequencing consortium (TCeSC) Genome Sequence of the Nematode *C. elegans*: A Platform for Investigating Biology. Science 1998; 282: 2018-2018

Tikhonov A.P., SanMiguel P.J., Nakajima Y., Gorenstein N.M., Bennetzen J.L., Avramova Z. Colinearity and its exceptions in orthologous *adh* regions of maize and sorghum. Proc Natl Acad Sci USA 1999; 96: 7409-7414

Tranquilli G., Lijavetzky D., Muzzi G., Dubcovsky J. Genetic and physical characterization of grain texture-related loci in diploid wheat. Mol Gen Genet 1999; 262: 846-850

Venter J.C., Adams M.D., Myers E.W., Li P.W., Mural R.J., Sutton G.G., Smith H.O., Yandell M., Evans C.A., Holt R.A., *et al.* The sequence of the human genome. Science 2001; 291: 1304-1351.

Wei F., Gobelman-Werner K., Morroll S.M., Kurth J., Mao L., Wing R.A., Leister D., Schulze-Lefert P., Wise R.P. The *Mla* (powdery mildew) resistance cluster is associated with three NBS-LRR gene families and suppressed recombination within a 240-kb DNA interval on chromosome 5S (1HS) of barley. Genetics 1999; 153: 1929-1948

Wei F., Wing R.A., Wise R.P. Genome Dynamics and Evolution of the *Mla* (Powdery Mildew) Resistance Locus in Barley. Plant Cell 2002; 14: 1903-1917

White M.A., Dominska M., Petes T.D. Transcription factors are required for the meiotic recombination hotspot at the HIS4 locus in *Saccharomyces cerevisiae*. Proc Natl Acad Sci USA 1993; 90: 6621-6625

Wicker T., Stein N., Albar L., Feuillet C., Schlagenhauf E., Keller B. Analysis of a contiguous 211kb sequence in diploid wheat (*Triticum monococcum* L.) reveals multiple mechanisms of genome evolution. Plant J 2001; 26: 307-316.

Wu J., Maehara T., Shimokawa T., Yamamoto S., Harada C., Takazaki Y., Ono N., Mukai Y., Koike K., Yazaki J., *et al.* A comprehensive rice transcript map containing 6591 expressed sequence tag sites. Plant Cell 2002; 14: 525-535

Wu T.C., Lichten M. Meiosis-induced double-strand break sites determined by yeast chromatin structure. Science 1994; 263: 515-518

Yao H., Zhou Q., Li J., Smith H., Yandeau M., Nikolau B.J., Schnable P.S. Molecular characterization of meiotic recombination across the 140-kb multigenic *a1-sh2* interval of maize. Proc Nat Acad Sci USA 2002; 99: 6157-6162

Yu J., Hu S., Wang J., Wong G.K., Li S., Liu B., Deng Y., Dai L., Zhou Y., Zhang X., *et al.* A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). Science 2002; 296: 79-92

Yu Y.G., Buss G.R., Maroof M.A. Isolation of a superfamily of candidate disease-resistance genes in soybean based on a conserved nucleotide-binding site. Proc Natl Acad Sci USA 1996; 93: 11751-11756

Chapter 13

WHOLE GENOME SEQUENCING: METHODOLOGY AND PROGRESS IN CEREALS

Yeisoo Yu and Rod A. Wing*

Arizona Genomics Institute, Department of Plant Sciences,University of Arizona, 303 Forbes Building, Tucson, AZ 85721, USA
*Author for correspondence: rwing@ag.arizona.edu

1. INTRODUCTION

Despite the short 35 years history of DNA sequencing, its progress has been enormous and has significantly influenced the manner, in which we study biological phenomena. In 1977, Sanger and colleagues reported a method to determine the order of DNA sequences using DNA polymerase and dideoxynucleotide triphosphates (ddNTPs) (Sanger *et al.,* 1977). During *in vitro* DNA synthesis with deoxynucleotide triphosphates (dNTPs), DNA polymerase can incorporate ddNTP analogues that cause termination of DNA synthesis at the A, C, G, or T positions. By separating and detecting newly synthesized DNA molecules on acrylamide gels, the exact order of the DNA sequence, base-by-base, can then be determined. This was the beginning of modern DNA sequencing and it is still used today without conceptual modification. This technique was successfully applied to determine the order of DNA sequences in several ambitious projects in early 1980's such as the human mitochondrial genome (Anderson *et al.,* 1981), human adenovirus (Gingeras *et al.,* 1982) and bacteriophage lambda (Sanger *et al.,* 1982). These pioneering projects not only yielded important sequence data to the research community, but also the confidence and potential to sequence entire genomes from living organisms representing individual species of bacteria, plants and animals. However, because DNA sequencing was not yet automated it was still not logistically or economically feasible to accomplish such large scale sequencing projects until the early 1990's. Figure 1 illustrates the progress of DNA sequence submission over the last 20 years.

P.K. Gupta and R.K. Varshney (eds.), Cereal Genomics, 385-423.

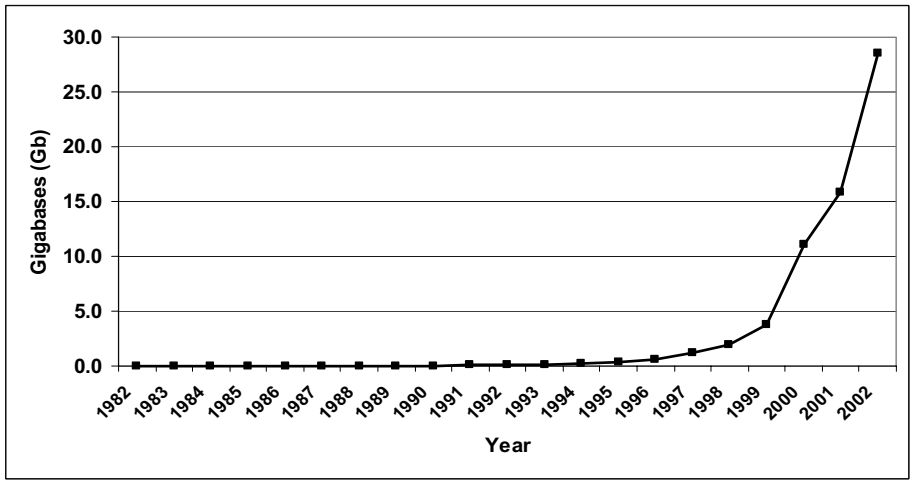


Figure 1. Sequence accumulation in Genbank from 1982 to 2002 (source: Genbank statistics at http://www.ncbi.nlm.nih.gov/Genbank/genbankstats.html)**.**

In 1982, the total amount of sequence submitted to Genbank was about 0.7 million bases (Mb) and progress of sequence deposition increased steadily until the mid 1990's. In 2002, sequence accumulation reached about 28.5 gigabases (Gb), about 10 times the size of the human genome, with about 12.7 Gb of new sequence deposited in 2002 only. This exponential release of sequence data into public databases can mainly be attributed to three factors: 1) advanced sequencing methods; 2) improved sequencing strategies; and 3) funded genome sequencing projects.

2. IMPROVEMENTS IN SEQUENCING TECHNIQUES

2.1. Advanced Sequencing Methods

Modification in labelling methods from radioactive to non-radioactive fluorescent tags on ddNTPs (Smith *et al.,* 1986; Prober *et al.,* 1987) was the first major improvement in DNA sequencing and led the birth of automated DNA sequencing in the late 1980s. Unlike the conventional radioactive labeling method which used autoradiography (X-ray film) for signal detection, fluorescently labeled sequences could be detected using lasers and automatically decoded with computer software. This eliminated error prone manual base calling and led to a quick turn around time from sequencing

reaction to genetic code. Further it made sequencing in a single tube possible, so called "cycle sequencing", because each of the four possible ddNTPs could be labeled with a different fluorescent dye. Development of cycle sequencing methods (Murray, 1989) offered several advantages over conventional isotope methods. First, it requires less amount of template DNA by reusing template in each cycle. Second, high temperature allows one to use double-stranded DNA templates instead of single-stranded ones. Third, high annealing temperature increases the specificity of primer binding. Fourth, high extension temperature decreases secondary structure of template DNA and allows the use of a variety of templates.

These two improved methods successfully increased the throughput and accuracy of sequencing until late 1990s. However, slab gel based sequencing was still a bottleneck for larger-scale production sequencing of many on-going genome projects. The introduction of capillary electrophoresis (Luckey *et al.,* 1990; Swerdlow *et al.,* 1990) accelerated production sequencing by removing labor dependent steps such as manual gel casting, sample loading and lane tracking. Capillary sequencers were distributed to genome sequencing centers in 1999 and increased the overall throughput by more than two-fold.

2.2. Improved Sequencing Strategies

Sanger and co-workers sequenced bacteriophage lambda (49 kb) using a random shotgun sequencing method (Sanger *et al.,* 1982) and thereby demonstrated the concept of sequence assembly of larger genomes with sequence overlap. Since then whole genome shotgun sequencing was thought of as a promising strategy to sequence much larger genomes. However, it was not a practical method at the time because computational software was not sufficient to support the assembly and sequencing throughput was too low. A major breakthrough came in 1995 when the 1.8 Mb genome of *Haemophilus influenzae* was completely sequenced and computationally assembled and annotated using a whole genome shotgun (WGS) method (Fleischmann *et al.*, 1995). Afterward WGS became a popular method for sequencing prokaryotic genomes.

Simultaneously, an alternative approach called the clone-by-clone (CBC) sequencing strategy was also evolving for eukaryotic genome projects due to the advent of the bacterial artificial chromosome (BAC) cloning system (Shizuya *et al.*, 1992) and advanced sequencing techniques. The BAC system permits the efficient cloning of 100-200 kb genomic DNA fragments as stable single copy plasmids in *E. coli*. BACs supplanted the problematic YAC

system and are now the method of choice for the CBC strategy. In 1996, the sequence tagged connector (STC) approach for genome sequencing was proposed to efficiently sequence the human genome (Venter *et al.,* 1996). In this new strategy, BACs could be used to eliminate two tedious mapping steps in conventional CBC sequencing – the construction of YAC and cosmid sequence ready maps. In addition, STCs could reduce sequencing costs by helping to determine a minimum tile of overlapping BAC clones to be sequenced thereby reducing the amount of overlapping sequence between clones. The CBC/STC method was subsequently used for the human (Lander *et al.,* 2001), Arabidopsis (TAGI, 2000) and rice genome (Sasaki *et al.,* 2002; Feng *et al.,* 2002; The Rice Chromosome 10 Sequencing Consortium, 2003) sequencing projects.

In 1997, Weber and Myers re-proposed the idea of sequencing the human genome using a WGS approach after successes with several microbial genomes and *Drosophila*. WGS was proven to be a faster and less expensive method to sequence and assemble prokaryotic genomes, but it was expected to have more difficulties in sequence assembly with larger and more complex eukaryotic genomes due to their higher repeat content. During last three years, because of advances bioinformatics, several working drafts have now been generated using the WGS method including *Drosophila* (Adams *et al.,* 2000), mouse (Waterston *et al.,* 2002), human (Venter *et al.,* 2001) and rice (Goff *et al.,* 2002; Yu *et al.,* 2002).

2.3. Genome Sequencing Projects

The human genome project (HGP) was launched in 1990 with the goal of providing a highly accurate sequence map to the community to better understand human health and disease. It began by concentrating on improving the genetic map and constructing a YAC physical map. A human EST program was started in 1991 to catalogue the expressed genes in the genome. Other model organisms' genome projects started early 1990's including mouse, yeast, *C. elegans,* and *Drosophila*. Because of the labor intensive procedures in sequencing, it was not efficient to produce massive sequence data until mid 1990s. The pilot sequencing phase of the HGP was started in 1996 to revisit the feasibility of large scale sequencing, and generated about 15% of the human genome during three years. Full-scale sequencing started from 1999 for two years and two human genome working drafts containing over 90% of human genome sequence were released in 2001. Meanwhile, genome sequencing of yeast, *C. elegans* and *Drosophila* were finished in 1996, 1999 and 2000 respectively, and the 120Mb genome of the flowering plant Arabidopsis was completely finished using the CBC approach in 2000.

In 1997, a year after the Arabidopsis genome sequencing project started, scientists began to organize the rice community to sequence the 430Mb of genome. In 1999, the CBC shotgun sequencing was initiated with integrated information from a BAC-based fingerprint/STC framework and a 10X public draft was announced December 2002 (http://rgp.dna.affrc.go.jp/rgp/Dec18_NEWS.html).

Whole genome sequencing provides a bounty of information to begin to understand the biology of our complex world. It can be applied to identify gene function and regulation, to study evolutionary relationships among organisms and to understand genome organization. With these consequences in mind, currently nine eukaryotic genomes are completely sequenced, five are in the finishing phase and many more are on the horizon. Which sequencing strategy to use, clone-by-clone or whole genome shogun, is still being debated. More recently, a hybrid method taking advantages from both approaches was successfully applied to the mouse genome project (Waterston *et al.,* 2002). There is no one absolute method over the other, thus one must justify the approach based on sequence accuracy in CBC versus speed in WGS or both. In this chapter, we will describe the two approaches of genome sequencing and their procedures to better understand genomic sequencing.

3. FRAMEWORKS FOR GENOME SEQUENCING PROJECTS

In eukaryotic genome sequencing projects, several components are required *a priori* to the initiation of sequencing. Those components are referred to as frameworks and are composed of genetic maps, large insert genomic libraries, fingerprint databases, physical maps, and STC databases. Integrated framework information plays a key role in management of the overall project, and at the same time provides unified information and material to the research community. Figure 2 shows the general scheme for a genome sequencing framework. In this section we describe a genome sequencing framework used to sequence the rice genome.

3.1. Genetic Maps

High density molecular linkage maps are an essential resource for genetics and genomics. They are essential for constructing high resolution maps for positional cloning, and are also used for regional or genome wide synteny detection among related species. More importantly, molecular genetic

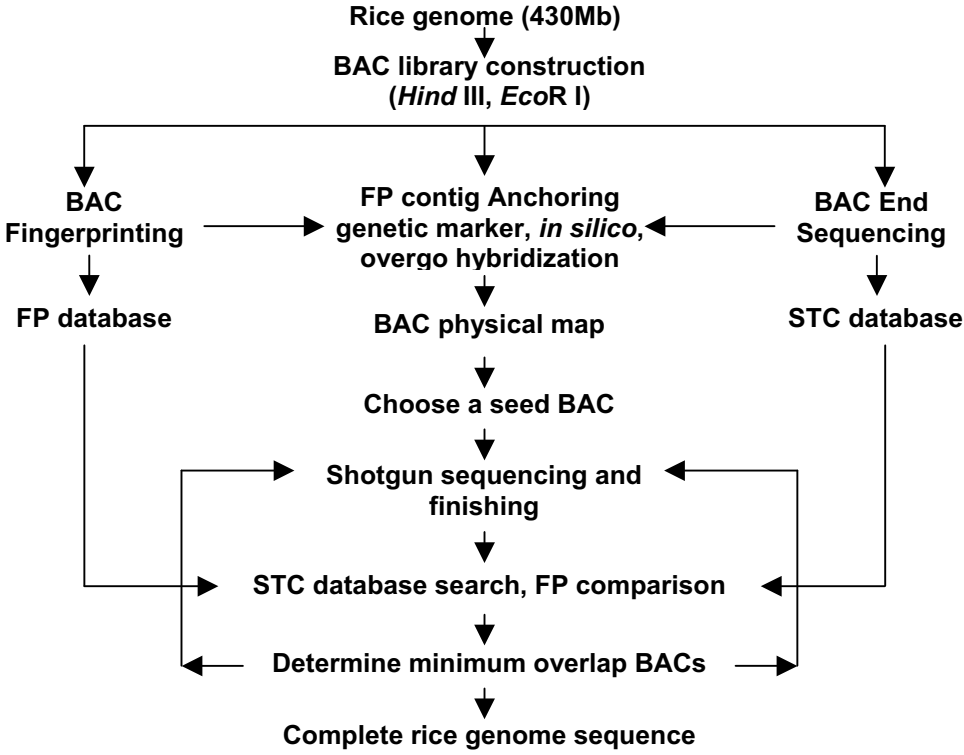


Figure 2. Framework components and their associations in CBC/STC genome sequencing approaches.

markers serve as anchor points to link and order fingerprinted contigs back to the genetic map.

Although several RFLP maps for rice have been developed (Monna et al 1994; Causse *et al.,* 1994; Yoshimura *et al.,* 1997), the Japanese RGP map (Harushima *et al.,* 1998) has been the most extensively used map for the rice genome framework and sequencing projects. The RGP high density genetic map was constructed with 186 F2 plants from a single cross between cultivars Nipponbare (*japonica*) and Kasalath (*indica*). It contains 3,267 RFLPs derived from cDNA clones, random genomic clones, YAC ends and genetic markers from other cereals (RGP, Japan, http://rgp.dna.affrc.go.jp/publicdata/geneticmap2000/index.html; also consult Chapter 3 by Varshney *et al.* in this book).

3.2. Large Insert Genomic Libraries

Large insert genomic clones are utilized as the raw material for positional cloning, physical mapping and DNA sequencing. Therefore a high-quality library is essential for genome research. Cosmids (Collins and Hohn, 1978), YACs (Burke *et al.,* 1987), BACs (Shizuya *et al.,* 1992) and PACs (Ioannou *et al.,* 1994) are the most popular large insert genomic DNA cloning systems.

Cosmids and YACs were the most widely used large insert cloning systems until early 1990s. In the early stage of human and *C. elegans* genome projects, YAC libraries were used for constructing physical maps and cosmid libraries served as the source for sequencing templates. YACs can theoretically maintain an unlimited insert size but YACs are relatively difficult to manipulate and inserts are frequently chimeric or unstable.

The advent of BACs rapidly replaced the roles of YACs and cosmids in genomic research due to the benign characteristics of BAC system. BACs hold up to 350kb inserts, and the inserts are stably maintained and are rarely shown to be chimeric (Shizuya *et al.,* 1992; Woo *et al.,* 1994). BAC-based physical maps of Arabidopsis (Mozo *et al.,* 1999; Marra *et al.,* 1999), mouse (Gregory *et al.,* 2002), human (McPherson *et al.,* 2001) and rice (Chen *et al.,* 2001) demonstrated the usefulness of BACs in providing accurate sequence ready maps for sequencing.

At present in a a number of cereal spcies, many large insert libraries (i.e. YAC, BAC) are available. Details about these libraries can be found in Chapter 11 of this book by Stein and Graner. For genome sequencing it is important to utilize 2-3 large insert BAC/PAC libraries containing inserts derived with different restriction enzymes or fragmentation methods to ensure unbiased BAC coverage across a genome. *Hin*dIII, *Eco*RI and *Bam*HI are the three main cloning enzymes for BAC library construction but some modified cloning techniques allow using one of the existing cloning sites with a different restriction enzyme (e.g. clone into *Bam*H I site after *Mbo* I partial digestion of genomic DNA). To support the rice genome project, our group built two deep-coverage BAC libraries using *Hin*d III and *Eco*R I partially digested high molecular weight DNA derived from rice nuclei. A total of 92,160 BACs were generated representing 10X and 15X genome coverage for the *Hin*dIII (OSJNBa) and *Eco*R1 (OSJNBb) libraries respectively. Table 1 shows the available large insert genomic libraries utilized in rice genome project.

Table 1. List of rice cv. Nipponbare large insert libraries used for rice genome projects

[1] Genome coverage = (Average insert size × Number of clones) / 430Mb; [2]RGP=Rice Genome Research Program, Japan;

Library	Cloning enzyme	Average insert size (kb)	Number of clones	[1]Genome coverage	Institute
YAC	*Eco*RI/ *Not* I	350	7,606	6X	[2]RGP
PAC	*Sau*3AI	112	71,040	18X	RGP
BAC	*Mbo*I	124	48,960	14X	RGP
BAC (OSJNBa)	*Hind* III	130	36,864	11X	[3]CUGI/[4]AGI
BAC (OSJNBb)	*Eco*R I	120	55,296	15X	CUGI/AGI
[5]BAC (OJ)	*Hind* III	122 kb	76,229	22X	Monsanto

[3]CUGI=Clemson University Genomics Institute, USA; [4]AGI=Arizona Genomics Institute, USA; [5]3,391 OJ BACs were shotgun sequenced with about 5X coverage by Monsanto Co.

3.3. Fingerprinting Database

Clone fingerprinting is an efficient method to build contigs of overlapping BAC clones by comparing the restriction fragment patterns among clones. Conventional methods using acrylamide gel and radioactive labeling of enzymatically restricted cosmids were first introduced in the *C. elegans* genome sequencing project in 1986 (Coulson *et al.,*1986). Since then, several improvements were made to make this method suitable for large scale projects. BAC fingerprint databases were successfully used to develop physical and sequence ready maps for Arabidopsis (Mozo *et al.,* 1999; Mara *et al.,* 1999), human (McPherson *et al.,* 2001), rice (Chen *et al.,* 2001) and mouse (Gregory *et al.,* 2002).

Figure 3 illustrates the procedures for generating a BAC-based physical map of the rice genome. Briefly, about 74,000 rice BACs were individually restricted using *Hin*d III and electrophoresised on 1% agarose gels, and imaged analyzed using the software Image (Sanger Centre, http://www.sanger.ac.uk/Software/Image/). Image is used to track lanes and call individual fingerprint bands for each BAC. Then Image processed gel data was input to the software FPC (FingerPrinted Contig, Soderlund *et al.,* 2000) to build the fingerprinted contigs with the assembly parameters of tolerance = 7 and cutoff <e-12. FPC builds contigs based on the number of shared bands in a fingerprint. Following the above procedures, about 74,000

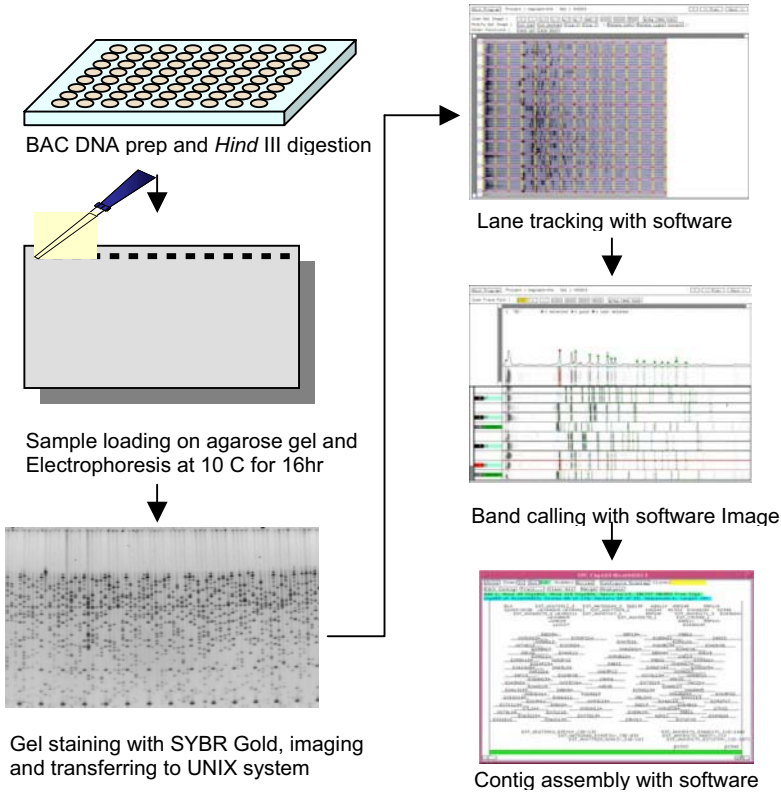


Figure 3. Procedures for BAC fingerprinting, band calling and contig assembly.

rice BACs were assembled into 1,019 contigs and 2,778 singletons without manual editing. The average number of bands was 13.5 in singletons and 28 in non-singleton clones; thus smaller clones were excluded in the contig assembly with a <e-12 cutoff. About 1,700 chloroplast clones assembled by themselves into one contig as expected. When we assume one *Hin*dIII band is 4.8kb in size, a scaffold size (93,000 unique bands) was estimated at 446Mb across the genome from the initial assembly which is very close to the rice genome size of 430 Mb.

3.4. Contig Anchoring - Toward a Physical Map

A molecular genetic map is required to assign chromosomal location of each fingerprinted contig. Library screening with genetically anchored markers is the primary tool used to assign the genetic position of a fingerprinted contig.

For rice two phases of contig anchoring were performed. Phase I primarily utilized two approaches: 1) hybridization of the BAC libraries with inserts or overgoes derived from genetically mapped RFLP markers; and 2) *in silico* marker hybridization with sequences derived from the ends of the BAC clones (Yuan *et al.,* 2000). For phase I, a total of 603 genetic markers and overgo probes were hybridized to the two BAC libraries resulting in 472 contigs (45% of total contigs) that could be anchored to the rice genome.

For phase II, three additional methods were applied to extend the anchoring information to the un-anchored contigs: 1) overgo probes were made from BAC end sequences located at the extreme ends of the anchored contigs and then used to screen the BAC libraries; 2) the FPC utility called "contig end search" was used to find potential neighboring contigs in FPC; and 3) BAC end sequences were used to search the assembled draft sequence of the rice genome donated by Monsanto Company (Barry *et al*., 2000) to identify potential contig ends that could be merged. These three approaches successfully detected neighboring contigs and allowed merging of contigs (Chen *et al.,* 2002). Currently 179 BAC contigs exist in rice physical map, 98 of which are anchored to the genetic map.

3.5. BAC End Sequencing

BAC end sequencing generates DNA sequences from both ends of BAC clones using a deep-coverage library. Coupled with restriction enzyme fingerprinting of BAC clones, BAC end sequences {or sequence tagged connectors (STCs)} are used to detect minimum overlapping clones with finished contiguous sequences of BAC clones (Venter *et al.,* 1996; Mahairas *et al.,* 1999; Siegel *et al.,* 1999; Batzoglou *et al.,* 1999). The sequence overlap between two clones can be as little as a few hundred base pairs, therefore, the

Table 2. Overall quality of rice BAC end sequences generated from two BAC libraries, *Hin*dIII and *Eco*RI

BACs	Total	Successful (%)	[1]Avg. HQ total	[2]Avg. HQ successful	BAC description and vector
OSJNBa	73,548	63,432 (86.2)	337.6bp	387.1bp	rice *Hin*d III BAC, pBeloBAC11
OSJNBb	54,097	47,006 (86.9)	304.9bp	342.9bp	rice *Eco*R I BAC, pIndigo536
Total	127,645	110,438 (86.5)	323.7bp	368.3bp	

[1]Average high quality bases (the bases having Phred ≥20 score) of total BAC end sequences; [2]Average high quality bases (the bases having Phred ≥20 score) of successful BAC end sequences (a sequence has ≥ 100 high quality bases).

STC strategy will decrease sequence redundancy in clone-by-clone genome sequencing. In whole genome shotgun sequencing, BAC end sequences are used to order the shotgun contigs and construct scaffolds across the genome. At the same time, this sequence information allows examination of important elements in a genome at early stages of sequencing (Mahairas *et al.,* 1999; Meyers *et al.,* 1999; Mao *et al.,* 2000; Temnykh *et al.,* 2001).

For the rice genome sequencing project we generated over 110,000 successful STCs from two rice BAC libraries (Table 2) and deposited the data into Genbank. The Rice STC database represents a total of 41Mb of high quality rice sequence and it is estimated that there is one STC located every 3.9kb across the genome. To estimate the STC coverage in rice genome, we further analyzed the STC contents in 3.5Mb of contiguous and non-redundant sequence on chromosomes 3 and 10 using Blast in FPC (referred BSS). The results are summarized in Table 3. A total of 443 STCs were aligned on 3.5Mb of sequence, thus one STC was observed in every 7.9kb. The difference between predicted (3.9kb/STC) and observed (7.9kb/STC) is mainly because: (i) about 17% of the clones (singletons and failed fingerprints) are not in a contig, and those STCs are excluded from the analysis. Therefore the corrected prediction is 4.7kb/STC; (ii) the BLAST cutoff parameters were very stringent (e<-200 and >95% sequence homology) and therefore some potential positive hits with lower confidence values were excluded in the analysis; (iii) the STC density varies across the genome (Table 3), and thus STC under-represented regions might be included in the analysis.

Table 3. Distribution and quality of positive sequence tagged connectors (STCs) in 3.5Mb of finished rice genome sequences

Location	Sequence length	Positive STCs	kb/STC	Average alignment (bp)	Homology
Chr 10-short arm	1Mb	142	7.0	548	97.8%
Chr 10-long arm	1Mb	110	9.1	549	98.0%
Chr 3-short arm	1.5Mb	191	7.9	547	98.0%
Total	3.5Mb	443	7.9	548	97.9%

4. GENOME SEQUENCING STRATEGIES

Two main strategies are presently being used to sequence genomes. One is called “clone-by-clone shotgun sequencing” (or CBC) and the other “whole genome shotgun sequencing” (WGS). The major difference between the two approaches is that a sequence ready physical map is constructed to pick BAC/PAC clones for sequencing in CBC while no map information is initially required in WGS. Thus, CBC requires more time up front to construct a physical map and WGS needs more effort to anchor sequence contigs to the map. Framework information, especially a physical map and STCs are used substantially for clone selection and chromosome walking in CBC sequencing, and used for scaffolding sequence contigs in WGS. In this section we describe the procedures, and discuss advantages and disadvantages of each strategy.

4.1. Clone by Clone Genome Sequencing

The clone-by-clone approach was successfully used to completely sequence the yeast, *C. elegans* and *Arabipopsis* genomes (Goffeau *et al.,* 1996; TceSC, 1998; TAGI, 2000) and is being used to sequence rice by the International Rice Genome Sequencing Project (IRGSP), (Sasaki and Burr, 2000; Sasaki *et al.,* 2002; Feng *et al.,* 2002; The Rice Chromosome 10 Sequencing Consortium, 2003). CBC was considered as a standard procedure for many sequencing projects because it could reduce sequence mis-assemblies caused by repeat sequences in eukaryotic genomes, and could generate nearly complete sequence information across highly repetitive regions like centromeres (TAGI, 2000). However, in the absence of a sequence-ready physical map and an STC database, CBC alone produced significant sequence overlap between clones and overall progress was relatively slow. With the addition of sequence-ready physical maps and STCs, as proposed by Venter *et al.* (1996), CBC became much more efficient by increasing the number of seed clones as entry points into the genome and significantly reducing clone overlap. For rice we adopted the STC approach to sequence chromosomes 3 and 10 (Fig. 2).

Four steps are required and reiterated to sequence a chromosome or genome using the CBC approach: (i) seed BAC selection; (ii) shotgun cloning and random sequencing; (iii) sequence assembly and finishing; and (iv) chromosome walking.

4.1.1. Seed BAC Selection

An initial set of BAC clones evenly spaced across a genome or chromosome are referred to as "seed" BACs. Because seed BACs provide entry points into the genome, their selection requires extreme care. For the short arm of rice chromosome 10, seed BACs were selected as follows. First, rice BAC libraries were screened with several single copy RFLP markers evenly spaced across the short arm of chromosome 10. Positive BAC clones were refigerprinted with *Hin*dIII and mapped to the rice physical map using FPC. This step was done to confirm that the BAC clones selected were correctly addressed. To confirm that the candidate BAC clones were derived from the short arm of chromosome 10, BAC DNA was digested with the same restriction enzyme that was used to map a given RFLP to chromosome 10. The digested DNA was then run on an agarose gel along with similarly digested rice genomic DNA and Southern blotted. The blots were probed with the labeled RFLP markers and analyzed to confirm that the BACs contained the identical RFLP band as that found in the rice genomic DNA. BAC clones passing these strict criteria were then used as seed BACs. It should be noted that since a seed BAC provides two directions of sequence extension, selection of a BAC in the middle of a FPC contig is highly desirable where possible.

4.1.2. Shotgun Cloning and Random Sequencing

To sequence a seed BAC, a shotgun library is constructed by random fragmentation of BAC plasmid DNA into 2-4kb fragments and cloned into a high copy plasmid vector. The inserts are then sequenced to generate about 8X sequence coverage of the BAC clone which are used for the initial assembly.

4.1.2.1. Insert preparation

Because BAC clones are present as single copy plasmids in *E. coli*, large culture volumes are required for DNA preparation. In addition, the single copy nature of the BAC system greatly enhances the chances for the BAC DNA preparation to be contaminated with *E. coli* genomic DNA. Although, CsCl gradients work very well for isolation of highly pure BAC plasmid DNA free from *E. coli* genomic DNA, they are time consuming and not very high throughput. As an alternative, we have successfully adopted a protocol called the "double acetate" method for BAC plasmid preparation. It significantly reduces *E. coli* genomic DNA contamination to less than 2%. The protocol is available at http://www.genome.ou.edu/DblAcetate

ProcV3.html and the cell culture volume can be scaled down to 50ml for shotgun library construction.

4.1.2.2. DNA fragmentation

Two methods for DNA fragmentation are used - partial digestion with restriction enzymes and mechanical shearing. Restriction enzyme fragmentation is convenient but can lead to biased representation of cloned fragments in a shotgun library. Mechanical shearing, by either nebulization, sonication or hydroshearing, is preferred because it generates an unbiased set of DNA fragments for shotgun library construction. After fragmentation, the ends are repaired (blunt-ended) using DNA polymerase, DNA kinase and dNTPs, and DNA fragments in the range of 2-4kb or 5-10kb are size-selected in agarose gels. The size-selected DNA fragments are then ligated with the appropriate vector and transformed into *E. coli* to create an 8-10 fold redundant shotgun library.

4.1.2.3. Choice of vector

Single strand sequencing vectors (e.g. M13) were the vectors of choice for shotgun sequencing until recently. Single strand templates are extremely cost effective to produce and generally yield higher quality sequences. However, they can only be sequenced in one direction, thus requiring twice as many M13 clones to sequence a BAC than if a double stranded vector was used. Recently, plasmid vectors (e.g. pUC or pBluescript) have become very popular for sequencing because: (i) both strands of a clone can be sequenced thereby requiring half as many template preparations as with M13 libraries; (ii) forward and reverse "read-pair" sequence information can be generated which is critical for ordering sequence contigs; and (iii) bridge clones located in sequencing gaps can easily be identified and used directly for gap filling via *in vitro* transposon insertion mutagenesis or shatter library construction. For rice chromosomes 10 and 3 we have exclusively and successfully used the high copy plasmid vector pBluescript KS for BAC shotgun libraries.

4.1.2.4. Library quality

Once a shotgun library is constructed, it is first quality tested by sequencing 96 randomly chosen clones to determine the amount of *E. coli* contamination, empty vectors and small inserts in the library. Libraries found to contain less than 10 percent (typically less than 2%) contamination are considered acceptable for production sequencing.

4.1.2.5. Random shotgun sequencing

Lander and Waterman's mathematical model (1988) permits one to calculate the number and size of the expected gaps in a sequencing project. In the model, the probability of a base not sequenced is expressed $P0=e^{-c}$, where c is

Table 4. Lander and Waterman's shotgun sequence coverage and error probability at a given target size of 150 kb BAC

Coverage	[1]P0 (= an error probability)	[2]P1 (=1-P0)	Shotgun sequencing of 150 kb BAC		
			[3]Sequence required (N)	[4]Gap length (Lg)	[5]No. of gaps (Ng)
1	0.3676	0.6324	250	55.147	91.912
2	0.1352	0.8648	500	20.275	67.582
3	0.0497	0.9503	750	7.454	37.270
4	0.0183	0.9817	1000	2.740	18.269
5	0.0067	0.9933	1250	1.008	8.396
6	0.0025	0.9975	1500	0.370	3.704
7	0.0009	0.9991	1750	0.136	1.589
8	0.0003	0.9997	2000	0.050	0.668
9	0.0001	0.9999	2250	0.018	0.276
10	0.0000	1.0000	2500	0.007	0.113

[1]$P0=e^{-c}$, the probability of a base is not sequenced at a certain coverage. Coverage (c) =L*N/G, where L is an average sequence length, N is the number of sequences required and G is a target size. [2]P1=1-P0, the probability of a base is sequenced at a certain coverage. [3]N=c*G/L (re-formulated from c=L*N/G), where G is 150kb and L is 0.6kb in length. [4]Lg=G*P0, where G is 150kb in length. [5]Ng=N*P0.

coverage. The coverage (c) is calculated by L*N/G, where L is sequence length, N is number of sequences, and G is sequence target size. Table 4 shows the expected sequencing probability and gaps at various sequence coverages. Predicted size and number of gaps are not the same as those observed because of the lack of genome complexity in the model, however it still is a useful guide to help estimate the number of sequence reads required for a given project. In our experience, we typically generate about eight fold sequence coverage with universal primers at the production sequencing phase. When we sequence a 150kb BAC, 8x coverage will require about 2000 sequences (1,000 shotgun clones with two primers – forward and revers) which will cover 99.97% of the BAC with 600bp read length per sequence at 100% sequencing success (Table 4).

4.1.3. Sequence Assembly and Finishing

After production sequencing, about 2,000 random shotgun sequences are processed sequentially using three computer programs for base calling, vector masking and sequence assembly.

4.1.3.1. Base calling

Phred (Ewing and Green, 1998; Ewing *et al.,* 1998) is the most standardized software for base calling. Phred calls bases and assigns quality values to each nucleotide by reading trace files. The quality value (q or phred value) is calculated using $q= -10*\log_{10}(p)$, where p is a probability of an error. For example, if the probability of an error is 0.01 (one error in 100 bases) then the phred value equals 20 (q=20). Thus, phred 30 corresponds to an error probability of 1/1,000bp and phred 40 represents 1/10,000bp error rate.

4.1.3.2. Vector removal

The next step is to remove the vector sequences, cloning and sequencing vectors, from the based called sequences. Cross_match (Green, 1999) is a popular program that uses the Smith-Waterman algorithm (Smith and Waterman, 1981) to compare a sequence against a vector sequence database to generate vector free sequences for assembly.

4.1.3.3. Sequence assembly

Phrap (Green, 1999) is the program used to generate a consensus sequence using vector free sequences with the same Smith-Waterman algorithm. Phrap can assemble projects with up to 64,000 sequences. Anything larger requires a different version of phrap called phrap.manyreads or alternative software like Arachne (Batzogluo *et al.*, 2001), Phusion (Mullikin and Ning, 2003), etc. Phrap also provides information including repeats, single stranded regions, low quality areas, high quality discrepancies and paired read information.

4.1.3.4. Finishing

Finishing is an integrated activity to make a contiguous consensus sequence with less than one error in 10,000bp at any base position. Unlike production sequencing, finishing uses many specialized techniques such as custom primer walking, PCR amplification, alternative sequencing chemistry, *in vitro* transposon mutagenesis and shatter libraries (200-500bp insert library derived from a shotgun clone) to produce a contiguous sequence and resolve problem areas. Consed (Gordon *et al.,* 1998) is the program primarily used for finishing. Consed graphically displays the phrap output or assembly information and, thus finishers can easily view and edit the assembly. Consed v12.0 supports *in silico* comparison of virtual and experimental restriction digestion patterns of BAC clones to confirm that the assembly is correct as a final check in finishing. Autofinish (Gordon *et al.,* 2001) is another feature in Consed that automatically generates a list of experiments to improve sequence quality and close gaps in the target sequence. Finishing is a very labor dependent procedure but can be more streamlined using autofinish.

To streamline the finishing process, one round of autofinish and *in vitro* transposon mutagenesis can be added as a "pre-finishing" step after production sequencing. This pre-finishing stage can focus on improving low quality regions using custom primers and re-sequencing failed reactions, and targeting gaps using custom primers. *In vitro* transposon insertion is a directed method used to close gaps where bridge clones are available. Transposons are inserted randomly in a target plasmid which provides unique priming sites for closing a sequence gap. Typically less than 48 sequences are required to close a sequence gap using transposons. Pre-finishing is extremely useful as a first pass for finishing and can also be efficiently scaled up to pre-finish many projects simultaneously.

4.1.4. Chromosome Walking

Chromosome walking is a reiterative process of clone picking, sequencing, and finishing to extend sequence information from a BAC to neighboring clones. The sequence tagged connector concept (STC: Venter *et al.,* 1996) proposed an efficient way to select minimally overlapping clones from seed BACs. Briefly, finished or nearly finished BACs are used as a query to BLAST search a database containing end sequences from every BAC in a deep coverage BAC library. Positive BACs from the BLAST results are compared and a single BAC having the least amount of sequence overlap with the finished BAC is selected for sequencing. By determining a minimal overlapping BAC at every walking step, the STC approach in theory minimizes sequence redundancy in an overall project. The BSS (BLAST some sequences) function in FPC successfully incorporates the STC approach for clone selection. STC search results, using a built-in BLAST programme in FPC, are parsed and graphically displayed in a BAC contig (Figure 4). One advantage of BSS is that false positive hits related to the repetitive sequence can be eliminated by displaying the hits in the contig where the query clone is located. At the same time, it can use not only finished sequence but also ordered contigs (Phase II sequence) as a query, thus subsequent clones can be selected at a relatively early stage of finishing.

One drawback of the STC approach is that clone selection is very dependent on finishing progress. Because finishing usually requires a much longer period of time to complete than production sequencing, the finishing process can stall clone selection thereby resulting in a lack of clones to production sequence. This becomes particularly problematic in regions of the genome that are highly repetitive, thus sequenced contigs are difficult to order. Moreover, uneven distribution of STCs across the genome due to the reaction

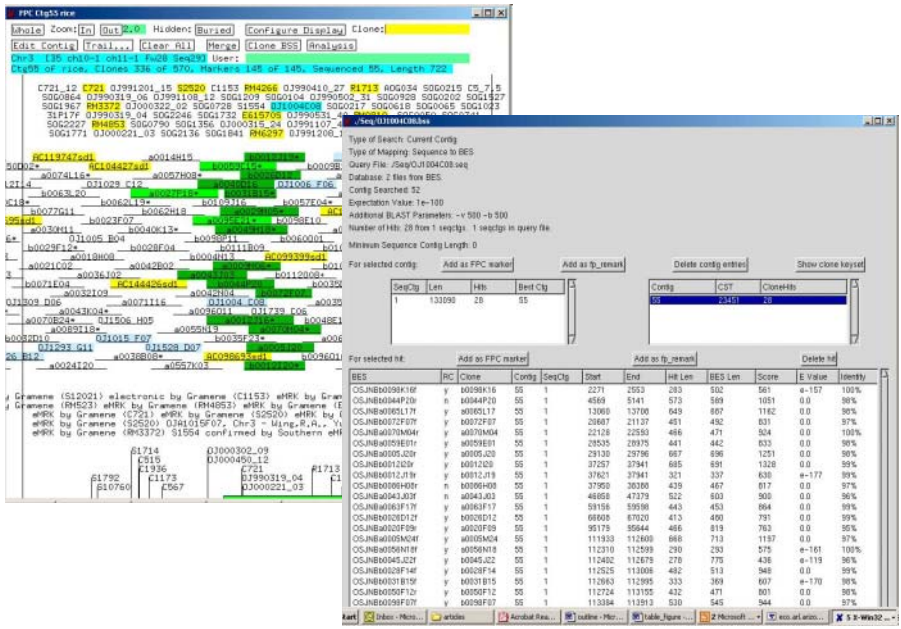


Figure 4. BSS output and display in a BAC fingerprinted contig. A rice BAC, OJ1004C08, was blasted against the rice STC database with the BSS function in FPC. 28 positive hits were identified and displayed in FPC.

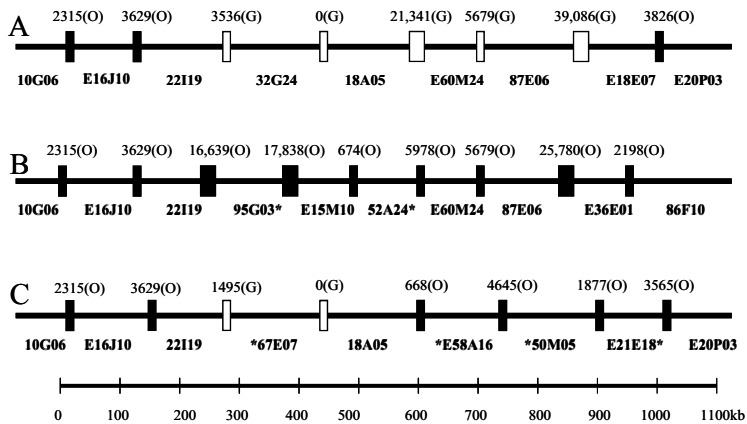


Figure 5. Results of STC walking simulations with three proposed methods.
A; Walk with gap using available STCs only, B; Re-sequence and walk without gap, C; Re-sequence and walk with gap. Number is the size of overlap or gap with (O) as overlap and (G) as gap. * is the degenerated STCs and direction.

failure or biased cloning of BACs also hinder the selection of minimally overlapping BAC clones for sequencing.

Using fingerprint maps alone can be an alternative method to select the next set of clones for sequencing. Fingerprints of three clones (one clone being sequenced, a candidate clone for sequencing, and a bridge clone that shares all of its *Hin*dIII restriction fragments with the two previous clones) are compared and a candidate clone having a minimum number of shared bands with the sequenced clone can be selected for sequencing. By this method, many BACs can be selected independently of finishing progress, but the overlaps are usually larger than with the sequenced based STC approach.

To develop a practical model for the STC approach for genome sequencing, we simulated three walking models using about 1.1Mb of contiguous genomic sequence spanning 43.4-52.4 cM on rice chromosome 1.

The first method can be called "walk with gap" and uses only STC information to walk and choose the next BAC to be sequenced. The walk with gap method uses the following steps: (i) pick a BAC clone within a minimal overlapping distance; (ii) shotgun sequence the BAC; (iii) walk with STCs; and (iv) when no BAC clone is detected within the minimum overlap (10% of an average BAC size), leave it as a gap rather than choose a BAC with a large overlap, and pick the adjacent BAC clone located close to the sequenced BAC on the physical map by comparing fingerprints. Figure 5A shows the result of this simulation. There were four overlaps between clones with a total of 15.4 kb or 3.9 kb each, at the same time, it produces four gaps totaling 64 kb in length. This gap size can be detected easily based on fingerprint data of three BACs, two BACs on each side of the gap (sequenced BAC and candidate BAC) and the third BAC spanning the gap. Depending on the gap size, it could be covered by different experimental procedures such as PCR or subcloning using an overlapping BAC. PCR can be used to fill the first and second gaps in Figure 5A because the gaps are small (the second one is actually not a gap but a shared *Hin*dIII site between two BAC clones, and is impossible to detect by STC walking). The third and fourth gaps can be filled by screening and sequencing small insert libraries (10-40kb in size) (Yang *et al.,* 2003; Jetty and Wing, unpublished).

The second method can be referred to as "Re-sequence and walk without gap". In this method, re-sequence the failed BAC ends near the finished BACs in the contig, which will increase the probability of chosing minimal overlapping BACs. For selecting a BAC without a gap, it is necessary to increase the minimum overlap to about 20% instead of 10%. The results of

this simulation are shown in Figure 5B. For this simulation, the fingerprint size data was used to identify the location of the failed BAC ends instead of re-sequencing (OSJNBa0095G03 and OSJNBa0052A24 had known STC data. See Figure 5B). Three large overlaps of 17, 18 and 26 kb were identified in this simulation but the overlap was about 81 kb in total or 9 kb between BACs. Thus the amount of overlap was 7.3% of total sequences.

The third method is called "Re-sequence and walk with gap". This is a mixture of the previous two methods and the result is shown in Figure 5C. Six overlaps exist with 16.7 kb in total or 2.8 kb per overlap and which represents 1.5% of total sequence length. Two gaps are detected here with 1.5 kb and one of them is a shared *Hin*d III site between two BACs, OSJNBa67E07 and OSJNBa18A05. Therefore, these two gaps could be covered and verified easily by PCR amplification.

These three simulations demonstrate that re-sequencing of failed STCs significantly increased the probability of finding minimally overlapping clones and that gaps generated by walking could be easily closed.

4.2. Whole Genome Shotgun Sequencing

The concept of whole genome sequencing using a shotgun approach for higher eukaryotes (Weber and Myers 1997; Venter at al. 1998) was based on the successful sequencing of several microbial genomes (Fleischmann *et al.,* 1995; Fraser *et al.,* 1995). Later it was extended to *Drosophila*, human, mouse and rice genome sequencing (Adams *et al.,* 2000; Venter *et al.,* 2001; Waterston *et al.,* 2002; Goff *et al.,* 2002; Yu *et al.,* 2002). Although WGS is an inexpensive and speedy way to sequence small microbial genomes (0.8-10 Mb) without the need for a physical map, it is difficult to have an accurate sequence map across complex eukaryote genomes because of sequence mis-assemblies caused by the nature and distribution of repeat sequences. For example, Green (1997) predicted representation as "poor to none" of sequences high in G/C, or heterochromatic regions, when assembling WGS reads from higher eukaryotic genomes.

WGS sequencing can be divided into three phases: (i) production sequencing and assembly; (ii) scaffolding and anchoring to the genome; and (iii) finishing. Phases (i) and (ii) are the most common procedures in WGS and lead to what is called a "draft" sequence of a genome and can be used to identify coding regions in a genome. For example in rice, more than 32,000 genes were identified in the *japonica* WGS draft produced by Syngenta (Goff *et al.,* 2002) and the sub-set of the data showed that about 90% of them were

full length genes. However, when the draft sequence was compared to the finished rice chromosome 10, more than twice the number of annotated genes was missing from the draft sequence. Similar data is observed when other drafts are compared with finished sequence and underscores the importance of accurate and finished genome sequence (Mardis *et al.,* 2002).

4.2.1. Shotgun Sequencing and Assembly

About six to ten fold genome coverage of sequence data is generated from randomly-fragmented genomic clones (~2kb insert clone). Additional genomic libraries with different insert sizes are required in order to cover the entire genome in an unbiased manner and to reduce potential mis-assemblies (2, 10, 50kb and BACs are used for human WGS; 2, 4, 6, 10, 40kb and BACs are used for mouse WGS). Two or three different sets of sequences are assembled using sophisticated software and massive computational power (Figure 6). Paired ends information (mate pairs) is of critical importance for whole genome shotgun sequencing to produce the correct assembly. By examining the location of end sequences in the contig, average clone size is obtained. Any paired ends that far exceed the average clone size can be considered a potential mis-assembly and the assembly can be corrected. Because 10kb clones can span most repeat elements, its paired end information increases the assembly accuracy. However, potential mis-assemblies can occur in longer repetitive regions and in recent sequence duplications (showing >99% sequence similarity).

For the *japonica* rice WGS project (Goff *et al.,* 2002), about 5.5 million sequences (~ 6X genome coverage) generated from 1.5kb insert clones were assembled into more than 42,000 contigs. It represented 390Mb genomic sequences and covered about 93% of rice genome. In the *indica* rice WGS draft, about 4.2X genome coverage of random sequences (3.6 M reads) from 2kb libraries were produced and assembled into 127,550 contigs (361Mb) with an average contig size of 6.7kb (Yu *et al.,* 2002).

4.2.2. Scaffolding and Anchoring to Genome

Paired end sequence information is also used to produce scaffolds in WGS. By observing paired ends, one end sequence is located in one contig and the other is in another contig, computer programs can detect adjacent contigs that could be joined in the assembly. Thus generation of perfect paired ends is extremely valuable in WGS. The distance and direction of the reads (shotgun clones have expected size and both reads face each other) is also used as

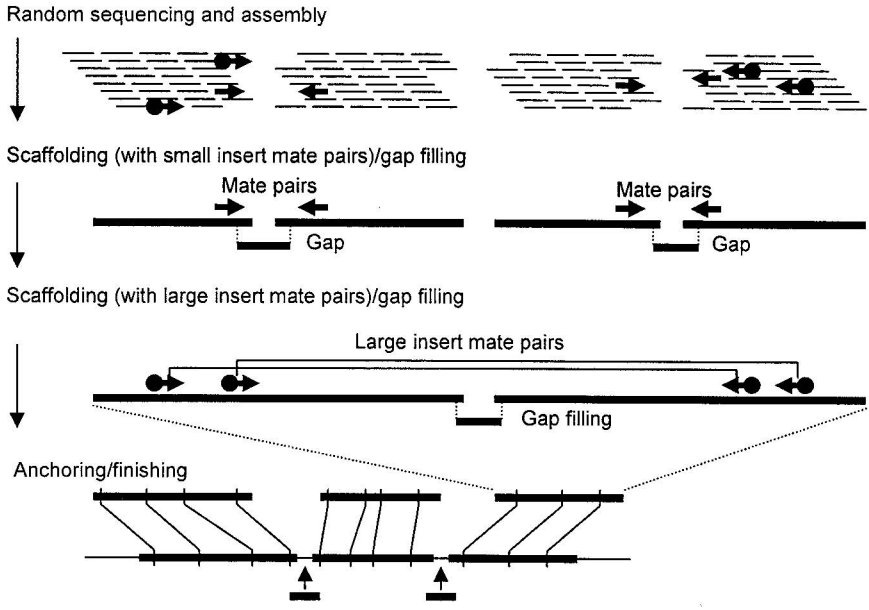


Figure 6. General scheme for whole genome shotgun genome sequencing (WGS).

additional information to make scaffolds and eliminate potential mis-assemblies. In mouse WGS, the N50 scaffold size (size above which 50% of scaffold was found) was reported as 16.9Mb, which far exceeds any WGS assembly previously reported. One of the important contributions mentioned in their report was the high-quality libraries with a variety of insert lengths such as 2, 4, 6, 10, 40 (fosmids) and 150 (BACs) kbs. Because the length of repeats is not uniform in genomes, sequences generated from libraries with various size inserts can potentially eliminate many assembly errors.

Genetic markers are essential to position sequence contigs or scaffolds to chromosomes. Sequence contigs containing genetic markers are first placed on the genetic map as initial anchoring points and all the "hybridization" can be done by *in silico* methods. BAC FPC maps, when available, can be added to the anchoring procedure by comparing the BAC clone order in the map and the BAC end sequences in the sequence contigs as was done with the *japonica* rice WGS (Goff *et al.,* 2002).

4.2.3. Gap Filling and Finishing

The procedures and methods involved gap filling in WGS are same as with the CBC approach. Many clones spanning gaps are identified from the

scaffolding step. Thus custom primer walking, PCR amplification and *in vitro* transposon methods are used to connect contigs where clones span the gap (sequence gap). In regions where clones are not represented (physical gaps), the gaps are amplified by PCR using genomic DNA followed by sequencing off the PCR amplicon, or cloning the PCR product and sequencing it. Low quality areas can be resolved with similar procedures.

4.2.4. Finishing Phase of Genome Sequencing

4.2.4.1. Sizing physical gaps and closure

Physical gaps are areas of the genome where sequence information is extremely difficult to obtain either due to poor clone representation or sequence complexity. In general it is very difficult to obtain complete clone coverage across telomeric, centromeric and highly heterchromatic regions and thus these areas can remain as physical gaps in both CBC and WGS methods permanently or for long periods of time. To estimate the size and nature of such gaps, fluorescence *in situ* hybridization (FISH) of whole chromosomes or extended fibers with probes flanking the gaps is commonly performed. Extended DNA fiber FISH (Fransz *et al.,* 1996; Cheng *et al.,* 2002) is the method of choice because its resolution of detection is down to a few kilobases, so that it can be used to measure physical gaps very precisely. For example, fiber FISH was successfully used to determine the size of seven physical gaps (30, 119, 69, 192, 104, 124 and 30kb in this order, and two telomere gaps (80 and 30kb) on rice chromosome 10 (The Rice Chromosome 10 Sequencing Consortium, 2003).

One of the only ways to fill such physical gaps is to access additional genomic libraries or construct new libraries with randomly sheared DNA fragments. For rice chromosome 10 we collaborated with the Rice Genome Research Program (RGP) in Japan to screen a new *Mb*o I BAC library with flanking sequences on either side of the physical gaps, and positive clones spanning two gaps were obtained. To help close gaps in the 24 rice telomeres, 96 telomere positive clones were identified from two rice 10kb genomic libraries (Yang *et al.,* 2003), four of which could be mapped on ch3L, ch7S, ch7L, and 10S (unpublished). These experiments clearly demonstrate the value of having a multitude of genomic libraries available for gap filling in the final stages of finishing a chromosome or genome.

4.2.4.2. Annotation

Annotation is defined as the integrated activity to assign genes, repeats and other elements of interests in the finished sequence using biological and

Table 5. List of gene prediction programmes used for rice gene annotation

Programme	Plant Training set/ purpose	Developer and URL
[1,2,3]FGENESH	Arabidopsis, tobacco, monocot	V.V. Solovyev http://www.softberry.com/berry.phtml
[1,2]Genscan/ Genscan+	Arabidopsis, maize	C. Burge http://genes.mit.edu/GENSCAN.html
[2]GeneMark.hmm	Arabidopsis, rice, maize	M. Borodovsky and A. Lukashin http://opal.biology.gatech.edu/GeneMark/eukhmm.cgi
[2]GlimmerR	Arabidopsis, rice	S.L. Salzberg *et al.* http://www.tigr.org/tdb/glimmerm/glmr_form.html
[1]RiceHMM	Rice	K. Sakata *et al.* http://rgp.dna.affrc.go.jp/RiceHMM/index.html
[1]MZEF	Exon prediction	M. Zhang http://argon.cshl.org/genefinder/
[1]Splice Predictor	Splice site prediction	L. Xing and V. Brendel http://bioinformatics.iastate.edu/cgi-bin/sp.cgi
[2]Gene Splicer	Splice site prediction	M. Pertea and S. Salzberg http://www.tigr.org/tdb/GeneSplicer/index.shtml
[1,2]tRNAscan-SE	tRNA prediction	S. Eddy http://www.genetics.wustl.edu/eddy/tRNAscan-SE/
[1,2]Repeat Masker	Repeat finding	A.F.A. Smit and P. Green http://ftp.genome.washington.edu/cgi-bin/RepeatMasker

[1]Programs used for Rice Genome Automated Annotation System (RiceGAAS); [2]Programs used for TIGR whole rice genome automated annotation database; [3]Program used for Rice Genome Browser in Gramene.

computational knowledge. Three primary steps are included in annotation; database searching, gene prediction, and editing.

(a) Database searching. Three database searches are performed, one each against protein, EST, and repeat databases. Protein and EST database searches help to provide evidence to support or refute gene prediction models. The databases that AGI routinely search are comprised of Genbank

non-redundant protein, Swiss-prot, Genbank plant EST and the TIGR Gene Indices. Most data can be downloaded from public databases and used locally. The repeat database is used to find the repeat elements in the sequence. Annotation features in repeats are transposon, retrotransposon, simple sequence repeats (SSR), centromere, telormere and MITEs (miniature inverted transposible elements).

(b) Gene prediction. Three to four gene prediction programs are used to predict the genes and make gene models. A list of the gene prediction programs is shown in Table 5. Usually many full length genes from a specific organism are used to train a program to increase the accuracy of gene prediction. For rice gene prediction, GenScan+, GenMark.hmm, RiceHMM, GlimmerR and FGENESH were trained with rice or monocot gene datasets. The best gene model is decided upon using a comparison of multiple gene prediction programmes output and similarity searches. t-RNA scan-SE is used to identify tRNA genes in the sequence.

(c) Editing and submission. Results from database searches and prediction programs are graphically displayed in the interface software and evidence is added to decide the optimal gene model for annotation. Final gene annotation is divided in three categories: (i) putative gene- any gene model having the significant homology with a known protein sequence; (ii) unknown gene- a gene model not having the protein homology but having the significant EST similarity; and (iii) hypothetical protein- a gene model not having positive hits on database entries but showing the same predictions with atleast three gene prediction programs. Untranslated region, exon and intron information and (start and stop information) is summarized and tabulated along with the protein sequences. Finally, all the annotated features (genes and repeats) are transferred to software Sequin (http://www.ncbi.nlm.nih.gov/Sequin/index.html) for a final check and submitted to Genbank.

5. PROGRESS- CEREAL GENOME SEQUENCING

5.1. Rice Genome Sequencing

5.1.1. Rice Genome Sequencing Progress

Rice is the most important food crop in the world. Its compact genome, evolutionary relationship with other cereals and sophisticated molecular genetic tools has made sequencing the rice genome a top priority for plant science. To meet this priority, the International Rice Genome Sequencing

Project (IRGSP) was launched in 1998 to generate a high quality genome sequence map using the cultivar Nipponbare (*Oryza sativa* ssp. *japonica*).

The IRGSP is currently comprised of eight active members: Japan, the United States of America, China, Taiwan, Korea, India, France, and Brazil. The IRGSP adopted the clone-by-clone shotgun sequencing strategy to sequence rice. The completion of a framework including a BAC-based physical map, STC database and genome anchoring led to the acceleration of genome sequencing in 1999. In 2000, a BAC-by-BAC partial working draft of the *japonica* genome was released by Monsanto (Barry, 2001) and integrated with the IRGSP sequence data. It was successfully used to reduce the amount of production sequencing required for certain regions of the rice genome and to more precisely anchor the BAC fingerprint map to the rice genome (Chen *et al.,* 2002). In early 2002, a WGS draft of the Nipponbare genome was generated by Syngenta (Goff *et al.,* 2002). It was donated to the IRGSP and used primarily to help fill sequence gaps in the early stages of finishing. In December 2002 the IRGSP completed the most comprehensive draft of the *japonica* genome along with finished chromosomes 1, 4 and 10 (Sasaki *et al.,* 2002; Feng *et al.,* 2002; The Rice Chromosome 10 Sequencing Consortium, 2003). This "high-quality draft" was comprised of approximately 10X sequence coverage of 3448 BACs across the entire genome and effectively released all of the commercial draft sequence data into the public domain. About 65% of the BACs are finished and another 33% of the BACs are in Phase2 status (the contig order is known). The most current sequencing status of the IRGSP, which is updated every night, can be found at: http://rgp.dna.affrc.go.jp/cgi-bin/statusdb/irgsp-status.cgi. The IRGSP is now set to finish the remainder of the genome by the end of 2004.

5.1.2. Automated Rice Genome Annotation

Automated annotation of the rice genome provides a convenient way to obtain a global view of the number and type of genes in the rice genome as well as their genetic position. There are currently three primary resources available for viewing the results of automated rice genome annotation: (i) The Japanese Rice Genome Program (RGP)'s Rice Genome Automated Annotation System (RiceGAAS, Sakata *et al.,* 2002); (ii) The Institute for Genome Research (TIGR); and (iii) Cold Spring Harbor and Cornell University's Gramene (Ware *et al.,* 2002).

Rice GAAS automatically integrates multiple gene prediction and similarity search results, and determines predicted gene models using gene domain prediction algorithms. Predicted gene models are further analyzed using

Table 6. List of analysis programmes used for Rice GAAS, automated annotation system

Programme	Analysis <URL>
HMMER	Protein homology search against motif databases- Pfam, Prosite <http://hammer.wustl.edu>
ProfileScan	Protein homology search against motif databases- Pfam, Prosite <http://hits.isb-sib.ch/cgi-bin/PFSCAN>
MOTIF	Search protein and DNA sequence motifs <http://motif.genome.ad.jp/>
PSORT	Protein localization site prediction <http://psort.nibb.ac.jp/>
SOSUI	Classification and secondary structure prediction of membrane protein <http://sosui.proteome.bio.tuat.ac.jp/sosuiframe0.html>
PLACE-SignalScan	Find *cis*-acting regulatory elements <http://www.dna.affrc.go.jp/htdocs/PLACE/signalscan.html>

protein motif search, localization prediction, protein classification and *cis*-element detection programs and displayed using a web-based graphical viewer. Programs used in RiceGAAS are listed in Table 5 and 6. Currently about 4,000 rice BAC/PACs annotation results using RiceGAAS are available through the RGP rice genome annotation database (http://ricegaas.dna.affrc.go.jp/rgadb/).

The Institute for Genome Research (TIGR) also provides a rice genome automated annotation database (http://www.tigr.org/tigr-scripts/e2k1/irgsp.spl). TIGR Gene Indices and repeat databases (transposons, retrotransposons and MITE) are used to perform similarity searches and FGENESH results are used as default gene models in this system. Finished, phase3 and phase2 rice BAC/PAC sequences are annotated and graphically displayed with their genetic map position. Pseudomolecules (non-overlapping sequences) and their annotation results are also available for all the 12 rice chromosomes.

Rice Genome Browser (http://www.gramene.org/japonica/mapview?chr=1) in Gramene integrates and displays various types of information including Genbank annotation provided from IRGSP, predicted gene structure based on FGENESH, genetic markers from cereals, BAC end sequences, simple

sequence repeats (SSRs), cereal ESTs and rice full-length cDNA sequences. Further information is provided by external databases such as Genbank or MaizeGDB, etc. linking to annotated features. BAC/PAC based annotation is available by chromosome for whole genome and pseudomolecule based annotation is also available for three finished chromosomes (chromosomes 1, 3 and 10).

5.1.3. Grass Comparative Genomics using Rice Genome Sequences

Introduction of RFLP genetic mapping techniques in plants produced new generation of molecular genetic maps and opened the door for genome comparison by cross-hybridization of RFLP markers between related genomes. Since Ahn and Tanksley (1993) reported large blocks of synteny between rice and maize using RFLP markers, many cases of collinearity were reported in grass species such as rice and wheat (Kurata *et al.,* 1994), maize and wheat (Devos *et al.,* 1994), rice, maize, wheat and oat (Van Deynze *et al.,* 1995), and foxtail millet and rice (Devos *et al.,* 1998). Later these results were combined as a "consensus grass map" (Moore *et al.*, 1995; Gale and Devos, 1998) where rice is located in the center of a set of concentric circles, with each circle representing the chromosomes of a single grass genome. Micro-synteny studies with *sh2* and *a1* loci also supported the conservation of gene order and orientation between rice, sorghum and maize (Chen *et al.,* 1997). Although exceptions to these observations were also reported in cases investigating micro-collinearity in the adh1 region of maize, sorghum and rice (Tikhonov *et al.,* 1999; Keller and Feuillet, 2002) and R-gene homologues in rice, barley and foxtail millet (Leister *et al.,* 1998), highly conserved syntenic relationships provide a good opportunity to use rice genome sequences to identify existing orthologous genes and common gene structures between grass genomes. Several databases and web resources have been established to support genome comparisons and orthologous gene studies between grass species (Table 7).

5.2. Maize Genome Sequencing

Maize is the next cereal crop to be sequenced due to its agronomic importance and significance as a model for plant genetic studies. However, the maize genome is 6-7 times larger (about 2,500Mb) than rice of which more than 50% is composed of retrotransposons placed between genes (Bennetzen *et al.,* 1994; Tikhonov *et al.,* 1999; SanMiguel *et al.,* 1998). This complexity makes maize whole genome sequencing difficult and ineffective.

Table 7. Web resources for comparative genome study and maize genome databases

Web resource	URL and curation
Gramene	http://www.gramene.org • Comparative map-rice, maize, sorghum, barley, wheat and oat • Gene ontology and protein database • Rice mutant phenotype database • Rice genome browser
PlantGDB	http://www.plantgdb.org/ • Collection and assembly of plant ESTs and genomic sequences • Annotation and sequence comparison
EGO (Eukaryotic gene orthologs)	http://www.tigr.org/tdb/tgi/ego/ • Collection of eukaryote orthologous gene database using tentative consensus (TC) of assembled ESTs
GrainGenes	http://wheat.pw.usda.gov/index.shtml • Collection of molecular and phenotypic information of wheat, barley, triticale, rye and oat • Rice and wheat genome comparison
MaizeGDB	http://www.maizegdb.org/ • Maize genome resources include genetic maps, ESTs, genes, phenotypes, microarray etc.
ZmDB	http://zmdb.iastate.edu/ • Maize ESTs and genomic sequences • Flanking sequence of *RescueMu* insertional mutation lines and phenotype database
TIGR Maize database	http://www.tigr.org/tdb/tgi/maize/ • High Cot and methyl filtration sequencing, assembly and annotation

Moreover, comparative analysis of grass genomes reveal that conservation of gene order exists but some local rearrangements interrupt synteny at molecular level (Tikhonov *et al.,* 1999; Keller and Feuillet, 2002). These rearrangements often prevent maize gene cloning using rice genome sequence information as a reference. Thus, having a complete maize genome sequence will be extremely beneficial to better understand gene and genome structure of rice and maize, and to understand the evolution of complex grass genomes. Also, identification of non-coding sequences conserved between rice and maize will help to elucidate candidate gene regulatory elements (Inada *et al.,* 2003; Guo and Moose, 2003).

5.2.1. Physical Map and BAC End Sequencing

Public framework projects for maize genome sequencing including FPC maps and BAC end sequences are under development. The maize agarose FPC physical map contains 291,569 *Hin*dIII agarose fingerprints from three BAC libraries (*Hin*dIII, *Eco*RI and *Mbo*I) constructed using maize cv. B73. The map represents about 15X coverage of maize genome and 1,446 of 3,488 contigs are genetically anchored across the 10 maize chromosomes. In addition to this, a HICF (High Information Content Fingerprint, Ding *et al.,* 2001) physical map is being constructed with these identical libraries. Both maps will be integrated to produce a sequence ready physical map of the maize genome that will be useful for efficient positional cloning and genome sequencing. HICF fingerprint uses type II restriction enzyme (*Ear*I) along with 4-basepair cutter (*Taq*I) followed by end label with base specific fluorescent dye. Because this method generates about 120 fingerprint bands per clone, fingerprint contigs can be assembled with much lower cutoff values or less overlap. Current versions of both FPC maps are available at URL http://www.genome.arizona.edu/fpc/maize/.

About 450,000 BAC ends from the three B73 BAC libraries have been generated and deposited in GSS section of Genbank. End sequences will provide basic information about genes and repeats in maize genome and help to select clones for genome sequencing. Maize BAC end sequence data can be retrieved at AGI (http://www.genome.arizona.edu/stc/maize/) and Rutgers University (http://pgir.rutgers.edu/).

5.2.2. Identifying Gene-rich Regions

Maize genome sequencing is focused on determining the gene rich regions in maize as the first targets for sequencing. Two experimental approaches have been initiated for targeting the so called “maize gene space”. One is the methyl filteration method (Rabinowicz *et al.,* 1999) and the other is high Cot cloning (Peterson *et al.,* 2002; Yuan *et al.,* 2003).

Methyl filteration utilizes the bacterial restriction-modification system to restrict methylated cytosine sequences in the genome. Because the primary targets for the methylation in maize genome are transposon, retrotransposon and non-genic regions, bacterial strains having active *mcrAB* genes can “filter” out methylated genomic sequence *in vivo*. Thus genomic clones generated by this method efficiently provide unmethylated sequences or potential genic sequences from the maize genome. The current release

contains 250,000 methyl filtration sequences and library information available through TIGR maize database (Table 7).

High Cot cloning uses DNA renaturation kinetics (Cot analysis) to separate sequences based on copy number using hydroxyapatite chromatography. Randomly sheared genomic DNA is exposed at melting temperature to induce denaturing DNA molecules and the denatured molecules are allowed to renature at a relatively high temperature (at 25^{o}C below Tm). Because high copy sequences have faster reassociation kinetics than low copy sequences, many repetitive elements become double-stranded and low copy sequences remain as single-stranded molecules during this procedure. In order to separate the single-stranded (low copy) from double-stranded (high copy) DNA, the reaction is run over a hydroxyapatite column. Then the collected single-stranded DNA is converted to double-stranded DNA by the addition of klenow DNA polymerase and random primers. Mungbean nuclease is then used to remove the single stranded ends of DNA and the remaining duplex DNA can then be cloned into a common plasmid vector. About 250,000 high Cot sequences and library information is available at TIGR maize database (Table 7).

Both gene enrichment methods generate only short sequence information of non-repetitive regions and thus, a tremendous amount of downstream work is required to localize the information to genetic and physical maps. However, the methods are independent of gene expression, unlike ESTs (Expressed Sequence Tags), and are therefore expected to identify many new genes and low copy sequences in the genome. Initial analysis showed that about 193,000 contigs assembled using 500,000 sequences from methyl filtration and high Cot methods, and about 14.7% and 20.4% of the assembled contigs matched to known genes and ESTs, respectively. About 3,000 contigs are anchored to maize genetic map (www.tigr.org/tdb/tgi/maize/).

5.2.3. Data Depository and Informatics

Proper data management and distribution are of critical importance for the maize genome sequencing project. In addition, the integrated data will be used to understand the genome architecture and determine gene rich regions in genome. Recent progress for the maize genome sequencing project is described at MaizeGDB, ZmDB and TIGR maize databases listed on Table 7. MaizeGDB contains a large collection of information about maize genetic maps, BAC/EST anchoring, QTL and phenotype data. ZmDB provides information related to EST collection, assembly and maize genomic survey sequence. The TIGR maize database is focused on curation and analysis of

methyl filtration and high Cot sequences. These databases provide integrated information about genetic resources, genomic data and preliminary analysis to be used by many researchers interested in the early phase of the maize genome sequencing project and will be utilized to select effectively the gene rich regions for genome sequencing.

5.3. Other Cereals

Prospects for whole genome sequencing of other important cereal genomes such as sorghum, barley and wheat are still in the planning stages. However, critical framework tools are already in place such as high-density molecular genetic maps, deep coverage BAC libraries and physical maps. It is unlikely that the complete genomes of barley and wheat will be attempted, but rather an approach focusing on sequencing gene rich regions similar to the maize genome sequencing project will be used. Most likely methyl filtration, Cot cloning and identification of gene rich BACs are the approaches that will be used to identify and position the majority of gene on these genomes.

5.3.1. cDNA Sequencing

One alternative to whole genome sequencing of large complex genomes is to perform one pass sequencing of random cDNA clones from numerous tissue specific libraries resulting in an Expressed Sequence Tagged (EST) database. This is a convenient and cost effective way to identify at least half of the expressed gene from an organism. The basic concept and strategy for of EST sequencing was developed by Venter and co-workers (Adams *et al.,* 1991) and many EST projects have since followed this lead. Recent improved cDNA technique and advanced DNA sequencing made cloning and sequencing full-length transcripts routine. Mammalian Gene Collection (MGC) project generates and distributes human and mouse full-length expressed transcript sequences and clones to the research community. This resource will experimentally correct or confirm the computational genome annotation and will be utilized to study transcription and transcription processing and gene expression in genome (Strausberg *et al.,* 1999). The Rice Full-Length cDNA Consortium recently reported over 28,000 full-length cDNA sequences mapped to *japonica* and *indica* rice genome sequences. Similarity searches determined that about 76% of the full-length cDNA could

Table 8. List of public cereal expressed sequence tag (EST) projects and TIGR Gene Indices

Common name	Scientific name	[1]dbEST	[2]TIGR Gene Indices		
			[3]TCs	[3]ETs	[3]EST singleton
Rice	*Oryza sativa*	283,935	29,363	17,162	28,479
Maize	*Zea mays*	393,719	29,414	524	26,426
Sorghum	*Sorghum bicolar*	161,813	18,659	164	18,409
Sorghum	*S. propinquum*	21,387	N/A	N/A	N/A
Wheat	*Triticum aestivum*	549,926	44,630	169	79,008
Barley	*Hordeum vulgare* spp. *vulgare*	356,848	21,981	168	27,041
Barley	*H. vulgare* spp. *spontaneum*	24,150	N/A	N/A	N/A
Rye	*Secale cereale*	9,194	1,391	66	3,890

[1]Genbank public EST database as on 5th March, 2004 (dbEST release 030504). [2]TIGR Gene Indices are available at URL <http://www.tigr.org/tdb/tgi/plant.shtml>. [3]TC (Tentative Consensus), ET (mature transcripts) and singleton ESTs are summarized based on current release of OsGI (rice, 16th Jan. 2004), ZmGI (maize, 23rd Dec. 2003), SbGI (sorghum, 22nd Dec. 2003), TaGI (wheat, 25th Dec. 2003), HvGI (barley, 9th Jan. 2004), and RyeGI (rye, 22nd Dec. 2003).

be assigned a tentative protein function and 64% showed homology with Arabidopsis proteins (Kikuchi *et al.,* 2003).

EST assembly (Quackenbush *et al.,* 2001; Christoffels *et al.,* 2001) has been successfully implemented to supplement the weakness of the single pass sequences, which, taken alone, are usually short and unreliable in quality. Also by clustering over-represented expressed sequences, it provides relatively long and accurate expressed gene sequences. The availability of ESTs is essential to the study of gene function and expression using micro array technology and is also extremely useful in the annotation of genomic sequences. To date (November 14, 2003), the number of GenBank EST entries in dbEST is 19,026,423. Of these, 101 organisms have greater than 10,000 EST entries. Currently, there are five major cereal EST projects supported by public funds: rice, maize, sorghum, wheat and barley listed in Table 8. With the rice genome sequence, cereal ESTs will provide tools to study agronomically important genes and their evolutionary relationship amongst the cereals.

6. SUMMARY AND OUTLOOK

Genome sequencing is an essential means to rapidly obtain comprehensive information for a genome. This information is used to catalogue genes and study gene function on a genome wide scale. Additionally, it supports the understanding of genome organization and evolution in related organisms. With these consequences in mind, sequencing of whole genomes and transcripts has been exponential. Much of the expansion in genome sequencing can be attributed to improvements in DNA sequencing technologies and the availability of genome sequencing frameworks. Although there is an on-going debate, two major approaches, clone-by clone sequencing (CBC) and whole genome shotgun sequencing (WGS), are being applied to sequence many genomes. The clone-by-clone approach provides highly accurate sequence across the genome but generates large redundant sequences and progress is relatively slow. Whole genome shotgun sequencing rapidly generates sequences in a cost-effective manner but sequence coverage is poor in large repeat and high G/C areas. More recently, a hybrid method, an integration of CBC and WGS, was used to sequence the mouse genome and resulted in an efficient generation of very accurate genomic sequence.

Despite the highly conserved gene order among major cereals, their genome sizes vary from 750 Mb of sorghum to 16000 Mb of wheat. Therefore it becomes technically and economically more difficult to generate complete genome sequences of the large and complex genomes of maize, barley and wheat. Unlike other cereals, rice has a compact genome (430Mb) and it serves as the principle source of nutrition for over one-third of the world's population. Thus, it is clear that rice is a very attractive model for studying cereal genomes. IRGSP's ambitious objective is underway to provide a complete and publicly available sequence map of rice genome. This comprehensive genome information will provide an unprecedented opportunity to gain a fundamental and basic understanding of the rice genome that can be used as a foundation to study other cereal genomes and plant biology in general.

7. REFERENCES

Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D., Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F., *et al.* The genome sequence of *Drosophila melanogaster*. Science 2000; 287:2185-2195

Adams M.D., Kelley J.M., Gocayne J.D., Dubnick M., Polymeropoulos M.H., Xiao H., Merril C.R., Wu A., Olde B., Moreno R.F., *et al.* Complementary DNA sequencing: expressed sequence tags and human genome project. Science 1991; 252: 1651-1656

Ahn S.N., Tanksley S.D. Comparative linkage maps of the rice and maize genomes. Proc Natl Acad Sci USA 1993, 90: 7980-7984

Anderson S., Bankier A.T., Barrell B.G., de Bruijn B.H., Coulson A.R., Drouin J., Eperon I.C., Nierlich D.P., Roe B.B., Sanger F., *et al*. Sequence and organization of the human mitochondrial genome. Nature 1981; 290: 457-465

Aparicio S., Chapman J., Stupka E., Putnam N., Chia J.M., Dehal P., Christoffels A., Rash S., Hoon S., Smit A., *et al.* Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes. Science 2000; 297: 1301-1310

Barry G.F.. The use of the Monsanto draft rice genome sequence in research. Plant Physiol 2001; 125: 1164-1165.

Batzoglou S., Berger B., Mesirov J., Lander E.S. Sequencing a genome by walking with clone-end sequences: a mathematical analysis. Genome Res 1999; 9: 1163-1174

Batzoglou S., Jaffe D.B., Stanley K., Butler J., Gnerre S., Mauceli E., Berger B., Mesirov J.P., Lander E.S. ARACHNE: a whole-genome shotgun assembler. Genome Res 200212: 177-189

Bennetzen J.L., Schrick K., Springer P.S., Brown W.E., SanMiguel P. Active maize genes are unmodified and flanked by diverse classes of modified, highly repetitive DNA. Genome 1994; 37: 565-576

Burke D.T., Carle G.F., Olson M.V. Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. Science 1987 236: 806-812

Causse M.A., Fulton T.M., Cho Y.G., Ahn S.N., Chungwongse J., Wu K., Xiao J., Yu Z., Ronald P.C., Harrington S.E., *et al.* Saturated molecular map of the rice genome based on an interspecific backcross population. Genetics 1994; 138: 1251-1274

Chen M., SanMiguel P., deOliveira A.C., Woo S.S., Zhang H., Wing R.A., Bennetzen J.L. Microcolinearity in *sh2*-homologous regions of the maize, rice, and sorghum genomes. Proc Natl Acad Sci USA 1997; 94: 3431-3435

Chen M., Presting G., Barbazuk W.B., Goicoechea J.L., Blackmon B., Fang G., Kim H., Frisch D., Yu Y., Sun S., *et al.* An integrated physical and genetic map of the rice genome. Plant Cell 2002; 14: 537- 545

Cheng Z.K., Buell C.R., Wing R.A., Jiang J. Resolution of fluorescence in-situ hybridization mapping on rice mitotic prometaphase chromosomes, meiotic pachytene chromosomes and extended DNA fibers. Chromosome Res 2002; 10: 379-387

Christoffels A., van Gelder A., Greyling G., Miller R., Hide T., Hide W. STACK: Sequence Tag Alignment and Consensus Knowledgebase. Nucl Acids Res 2001; 29: 234-238

Collins J., Hohn B. Cosmids: a type of plasmid gene-cloning vector that is packageable in vitro in bacteriophage lambda heads. Proc Natl Acad Sci USA 1978; 75: 4242-4246

Coulson A., Sulston J., Brenner S., Karn J. Towards a physical map of the genome of the nematode *C. elegans.* Proc Natl Acad Sci USA 1986; 83: 7821-7825

Devos K.M., Chao S., Li Q.Y., Simonetti M.C., Gale M.D. Relationship between chromosome 9 of maize and wheat homeologous group 7 chromosomes. Genetics 1994; 138: 1287-1292

Devos K.M., Wang Z.M., Beales J., Sasaki Y., Gale M.D. Comparative genetic maps of foxtail millet (*Setaria italica*) and rice (*Oryza sativa*). Theor Appl Genet 1998; 96: 63-68

Ding Y., Johnson M.D., Chen W.Q., Wong D., Chen Y.J., Benson S.C., Lam J.Y., Kim Y.M., Shizuya H. Five-color-based high-information-content fingerprinting of bacterial artificial chromosome clones using type IIS restriction endonucleases. Genomics 2001; 74: 142-154

Ewing B., Green P. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res 1998; 8: 186-194

Ewing B., Hillier L., Wendl M.C., Green P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res 1998 8: 175-185

Feng Q., Zhang Y., Hao P., Wang S., Fu G., Huang Y., Li Y., Zhu J., Liu Y., Hu X., *et al.* Sequence and analysis of rice chromosome 4. Nature 2002; 420: 316-320

Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F., Kerlavage A.R., Bult C.J., Tomb J.F., Dougherty B.A., Merrick J.M., *et al.* Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. Science 1995; 269: 496-512

Fransz P.F., Alonso-Blanco C., Liharska T.B., Peeters A.J., Zabel P., de Jong J.H.. High-resolution physical mapping in Arabidopsis thaliana and tomato by fluorescence in situ hybridization to extended DNA fibres. Plant Jour 1996; 9: 421-430

Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A., Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M. *et al.* The minimal gene complement of Mycoplasma genitalium. Science 1995; 270: 397-403

Gale M.D., Devos K.M. Comparative genetics in the grasses. Proc Natl Acad Sci USA 1998; 95: 1971-1974

Gingeras T.R., Sciaky D., Gelinas R.E., Bing-Dong J., Yen C.E., Kelly M.M., Bullock P.A., Parsons B.L., O'Neill K.E., Roberts R.J.. Nucleotide sequences from the adenovirus-2 genome. Jour Biol Chem 1982; 257: 13475-13491

Goff S.A., Ricke D., Lan T.H., Presting G., Wang R.L., Dunn M., Glazebrook J., Sessions A., Oeller P., Varma H., *et al.* A draft sequence of the rice genome (*Oryza sativa* L. ssp *japonica*). Science 2002; 296: 92-100

Goffeau A., Barrell B.G., Bussey H., Davis R.W., Dujon B., Feldmann H., Galibert F., Hoheisel J.D., Jacq C., Johnston M., *et al.* Life with 6000 genes. Science 1996; 274: 562-567

Gordon, D., Abajian C., Green P. Consed: a graphical tool for sequence finishing. Genome Res 1998; 8: 195-202

Gordon D., Desmarais C., Green P. Automated finishing with autofinish. Genome Res 2001; 11: 614-625

Green P. Against a whole-genome shotgun. Genome Res 1997; 7: 410-417

Green P. Documentation for phrap and cross_match (version 0.990319). 1999; http://bozeman.mbt.washington.edu/phrap.docs/phrap.html

Gregory S.G., Sekhon M., Schein J., Zhao S., Osoegawa K., Scott C.E., Evans R.S., Burridge P.W., Cox T.V., Fox C.A., *et al.* A physical map of the mouse genome. Nature 2002; 418: 743-750

Guo H., Moose. S.P. Conserved noncoding sequences among cultivated cereal genomes identify candidate regulatory sequence elements and patterns of promoter evolution. Plant Cell 2003; 15: 1143-1158

Harushima Y., Yano M., Shomura A., Sato M., Shimono T., Kuboki Y., Yamamoto T., Lin S.Y., Antonio B.A., Parco A. *et al.* A high density rice genetic linkage map with 2275 markers using a single F2 population. Genetics 1998; 148: 479-494

Inada D.C., Bashir A., Lee C., Thomas B.C., Ko C., Goff S.A., Freeling M. Conserved noncoding sequences in the grasses. Genome Res 2003; 13: 2030-2041

Ioannou P.A., Amemiya C.T., Garnes J., Kroisel P.M., Shizuya H., Chen C., Batzer M.A., de Jong P.J. A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. Nature Genet 1994; 6: 84-89

Jaffe, D.B., J. Butler, S. Gnerre, E. Mauceli, K. Lindblad-Toh, J.P. Mesirov, M.C. Zody, and E.S. Lander. Whole-genome sequence assembly for mammalian genomes: Arachne 2. Genome Res 2003; 13: 91-96

Keller B., Feuillet C. Colinearity and gene density in grass genomes. Trends Plant Sci 2000; 5: 246-251

Kikuchi S., Satoh K., Nagata T., Kawagashira N., Doi K., Kishimoto N., Yazaki J., Ishikawa M., Yamada H., Ooka H., *et al.* Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice. Science 2003; 301:376-379

Kurata, N., Moore, G., Nagamura, Y., Foote, T., Yano, M., Minobe, Y., and Gale, M. Conservation of genomic structure between rice and wheat. Bio/Technology 1994; 12: 276-278

Lander E.S., Linton L.M., Birren B., Nusbaum C., Zody M.C., Baldwin J., Devon K., Dewar K., Doyle M., FitzHugh W., *et al.* Initial sequencing and analysis of the human genome. Nature 2001; 409:860-921

Lander E.S., Waterman M.S. Genomic mapping by fingerprinting random clones: a mathematical analysis. *Genomics* 1988; 2: 231-239

Leister D., Kurth J., Laurie D.A., Yano M., Sasaki T., Devos K., Graner A., Schulze-Lefert P. Rapid reorganization of resistance gene homologues in cereal genomes. Proc Nat Acad Sci USA 1998; 95: 370-375

Luckey J.A., Drossman H., Kostichka A.J., Mead D.A., Cunha J. D.', Norris T.B., Smith L.M.. 1990. High speed DNA sequencing by capillary electrophoresis. Nucl Acids Res 18: 4417-4421

Mahairas G.G., Wallace J.C., Smith K., Swartzell S., Holzman T., Keller A., Shaker R., Furlong J., Young J., Zhao S., *et al.* Sequence-tagged connectors: a sequence approach to mapping and scanning the human genome. Proc Natl Acad Sci USA 1999; 96: 9739-9744

Mao L., Wood T.C., Yu Y., Budiman M.A., Tomkins J., Woo S., Sasinowski M., Presting G., Frisch D., Goff S., *et al.* Rice transposable elements: A survey of 73,000 sequence-tagged-connectors. Genome Res 2000; 7: 1072-1084

Mardis E., McPherson J., Martienssen R., Wilson R.K., McCombie W.R. What is finished, and why does it matter. Genome Res 2002; 12: 669-671

Marra M., Kucaba T., Sekhon M., Hillier L., Martienssen R., Chinwalla A., Crockett J., Fedele J., Grover H., Gund C., McCombie W.R., *et al.* 1999. A map for sequence analysis of the Arabidopsis thaliana genome. Nature Genet 22: 265-270

McPherson J.D., Marra M., Hillier R.H., Waterston A., Chinwalla J., Wallis M., Sekhon K., Wylie E.R., Mardis R.K., Wilson R., *et al.* A physical map of the human genome. Nature 2001 409: 934-941

Meyers B.C., Dickerman A.W., Michelmore R.W., Sivaramakrishnan S., Sobral B.W., Young N.D. Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. Plant Jour 1999; 20: 317-332

Monna L., Miyao A., Inoue T., Fukuoka S., Yamazaki M., Zhong H.S., Sasaki T., Minobe Y.. Determination of RAPD markers in rice and their conversion into sequence tagged sites (STSs) and STS-specific primers. DNA Res 1994; 1: 139-148

Moore G., Devos K.M., Wang Z., Gale M.D. Grasses line up and form a circle. Curr Biol 1995; 5: 737-739

Mozo T., Dewar K., Dunn P., Ecker J.R., Fischer S., Kloska S., Lehrach H., Marra M., Martienssen R., Meier-Ewert S., Altmann T. A complete BAC-based physical map of the Arabidopsis thaliana genome. Nature Genet 1999 22: 271-275.

Mullikin J.C., Ning Z. The phusion assembler. Genome Res 2003; 13: 81-90

Murray V. Improved double-stranded DNA sequencing using the linear polymerase chain reaction. Nucl Acids Res 1989; 17: 8889

Myers E.W., Sutton G.G., Delcher A.L., Dew I.M., Fasulo D.P., Flanigan M.J., Kravitz S.A., Mobarry C.M., Reinert K.H., Remington K.A., *et al.* A whole-genome assembly of *Drosophila*. Science 2000; 287: 2196-2204

Peterson D.G., Schulze S.R., Sciara E.B., Lee S.A., Bowers J.E., Nagel A., Jiang N., Tibbitts D.C., Wessler S.R., Paterson A.H. Integration of Cot analysis, DNA cloning, and high-throughput sequencing facilitates genome characterization and gene discovery. Genome Res 2002; 12: 795-807

Prober J.M., Trainor G.L., Dam R.J., Hobbs F.W., Robertson C.W., Zagursky R.J., Cocuzza A.J., Jensen M.A., Baumeister K. A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides. Science 1987; 238: 336-341

Quackenbush J., Cho J., Lee D., Liang F., Holt I., Karamycheva S., Parvizi B., Pertea G., Sultana R., White J. The TIGR Gene Indices: analysis of gene transcript sequences in highly sampled eukaryotic species. Nucl Acids Res 2001; 29: 159-164

Rabinowicz P.D., Schutz K., Dedhia N., Yordan C., Parnell L.D., Stein L., McCombie W.R., Martienssen R.A. Differential methylation of genes and retrotransposons facilitates shotgun sequencing of the maize genome. Nature Genet 1999; 23:305-308

Sakata K., Nagamura Y., Numa H., Antonio B.A., Nagasaki H., Idonuma A., Watanabe W., Shimizu Y., Horiuchi I., Matsumoto T., *et al.* RiceGAAS: an automated annotation system and database for rice genome sequence. Nucl Acids Res 2002; 30: 98-102

Sanger F., Coulson A.R., Hong G.F., Hill D.F., Petersen G.B. Nucleotide sequence of bacteriophage lambda DNA. Jour Mol Biol 1982; 162: 729-773

Sanger F., Nicklen S., Coulson A.R. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 1977; 74: 5463-5467

SanMiguel P., Gaut B.S., Tikhonov A., Nakajima Y., Bennetzen J.L. The paleontology of intergene retrotransposons of maize. Nature Genet 1998; 20: 43-45

Sasaki T., Burr B. International Rice Genome Sequencing Project: the effort to completely sequence the rice genome. Curr Opin Plant Biol 2000; 3: 138-141

Sasaki T., Matsumoto T., Yamamoto K., Sakata K., Baba T., Katayose Y., Wu J., Niimura Y., Cheng Z., Nagamura Y., *et al.* The genome sequence and structure of rice chromosome 1. Nature 2002; 420: 312-316

Shizuya H., Birren B., Kim U.J., Mancino V., Slepak T., Tachiiri Y., Simon M. Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. Proc Natl Acad Sci USA 1992; 89: 8794-8797

Siegel A.F., Trask B., Roach J.C., Mahairas G.G., Hood L., van den Engh G. Analysis of sequence-tagged-connector strategies for DNA sequencing. Genome Res 1999; 9: 297-307

Smith L.M., Sanders J.Z., Kaiser R.J., Hughes P., Dodd C., Connell C.R., Heiner C., Kent S.B., Hood L.E.. Fluorescence detection in automated DNA sequence analysis. Nature 1986; 321: 674-679

Smith T.F., Waterman M.S. Identification of common molecular subsequences. Jour Mol Biol 198; 1147: 195-197

Soderlund C.S., Dunham H.A., French L. Contigs built with fingerprints, markers, and FPC V4.7. Genome Res 2000; 10: 1772-1787

Strausberg R.L., Feingold E.A., Klausner R.D., Collins F.S. The mammalian gene collection. Science 1999; 286: 455-457

Swerdlow H., Wu S.L., Harke H., Dovichi N.J.. Capillary gel electrophoresis for DNA sequencing. Laser-induced fluorescence detection with the sheath flow cuvette. Jour Chromatogr 1990; 516: 61-67

TAGI, The Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 2000; 408: 796-815

The *C. elegans* Sequencing Consortium. Genome sequence of the nematode C. elegans: a platform for investigating biology. Science 1998; 282: 2012-2018

The Cold Spring Harbor Laboratory, Washington University Genome Sequencing Center, and PE Biosystems Arabidopsis Sequencing Consortium. 2000. The complete sequence of a heterochromatic island from a higher eukaryote. Cell 2000; 100: 377-386

The Rice Chromosome 10 Sequencing Consortium: Yu Y., Rambo T., Currie J., Saski C., Kim H.R., Collura K., Thompson S., Simmons J., Yang T.-J., Nah G. *et al.* In-depth view of structure, activity, and evolution of rice chromosome 10. Science 2003; 300: 1566-1569

Temnykh S., DeClerck G., Lukashova A., Lipovich L., Cartinhour S., McCouch S. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. Genome Res 2001; 11: 1441-1452

Tikhonov A.P., SanMiguel P.J., Nakajima Y., Gorenstein N.M., Bennetzen J.L., Avramova Z. Colinearity and its exceptions in orthologous adh regions of maize and sorghum. Proc Natl Acad Sci USA 1999; 96: 7409-7414

Van Deynze A.E., Nelson J.C., Yglesias E.S., Harrington S.E., Braga D.P., McCouch S.R., Sorrells M.E. Comparative mapping in grasses. Wheat relationships. Mol Gen Genet 1995; 248: 744-754

Venter J.C., Adams M.D., Myers E.W., Li P.W., Mural R.J., Sutton G.G., Smith H.O., Yandell M., Evans C.A., Holt R.A., *et al.* The sequence of the human genome. Science 2001; 291:1304- 1351

Venter J.C., Adams M.D., Sutton G.G., Kerlavage A.R., Smith H.O., Hunkapiller M. Shotgun sequencing of the human genome. Science 1998; 280: 1540-1542

Venter J.C., Smith H.O., Hood L. A new strategy for genome sequencing. Nature 1996; 381: 364-366

Wang J., Wong G.K., Ni P., Han Y., Huang X., Zhang J., Ye C., Zhang Y., Hu J., Zhang K., *et al.* RePS: a sequence assembler that masks exact repeats identified from the shotgun data. Genome Res 2002; 12: 824-831

Waterston R.H., Lindblad-Toh K., Birney E., Rogers J., Abril J.F., Agarwal P., Agarwala R., Ainscough R., Alexandersson M., An P., *et al.* Initial sequencing and comparative analysis of the mouse genome. Nature 2002; 420: 520-562

Ware D.H., Jaiswal P., Ni J., Yap I.V., Pan X., Clark K.Y., Teytelman L., Schmidt S.C., Zhao W., Chang K., *et al.* Gramene, a tool for grass genomics. Plant Physiol 2002; 130: 1606-1613

Weber J.L., Myers E.W. Human whole-genome shotgun sequencing. Genome Res 1997; 7: 401-409

Woo S.S., Jiang J., Gill B.S., Paterson A.H., Wing R.A. Construction and characterization of a bacterial artificial chromosome library of Sorghum bicolor. Nucl Acids Res 1994; 22: 4922-4931

Yang T.J., Yu Y., Nah G., Atkins M., Lee S., Frisch D.A., Wing R.A. Construction and utility of 10-kb libraries for efficient clone-gap closure for rice genome sequencing. Theor Appl Genet 2003; 107: 652-60

Yoshimura A., Ideta O., Iwata N. Linkage map of phenotype and RFLP markers in rice. *Plant Mol Biol* 1997; 35: 49-60

Yu J., Hu S., Wang J., Wang G., Li S.G., Wong K.S.G., Liu B., Deng Y., Dai L., Zhou Y., Zhang X., *et al.* A draft sequence of the rice genome (*Oryza sativa* L. sspl. *indica*). Science 2002; 296:79-92

Yuan Q.P., Liang F., Hsiao J., Zismann V., Benito M.I., Quackenbush J., Wing R., Buell R. Anchoring of rice BAC clones to the rice genetic map in silico. Nucl Acids Res 2000; 28: 3636-3641

Yuan, Y., SanMiguel P.J., Bennetzen J.L.. 2003. High-Cot sequence analysis of the maize genome. Plant Jour 34: 249-255.

Chapter 14

BIOINFORMATICS AND TRITICEAE GENOMICS: RESOURCES AND FUTURE DEVELOPMENTS

David E. Matthews[1], Gerard R. Lazo[2], Victoria L. Carollo[1] and Olin D. Anderson[2]

[1]USDA-ARS, Cornell University, Ithaca, NY 14853, USA; *[2]USDA-ARS Western Regional Research Center, Albany, CA 94710, USA*
*Corresponding author: dem3@cornell.edu

1. INTRODUCTION

In several chapters of this book, it has been emphasized that bioinformatics is an important component of cereal genomics research. Collection of data in the databases and their retrieval and use for a variety of purposes are now routine steps in cereal genomics. Laboratory scientists increasingly need at least an elementary knowledge of bioinformatics, including knowledge of available databases and of the programmes used to retrieve and manipulate data from these resources.

Although it is frequently stated that due to the size of the Triticeae genomes (for size of cereal genomes, see Gupta and Varshney, Chapter 1) they will not be completely sequenced in the near future, the advance in technologies, the growing evidence of the value of whole genome sequencing, and the overriding importance of wheat as a pillar of the human diet will eventually obviate such considerations. Whatever the near future in Triticeae genomics, it will bring with it a need for appropriate kinds of bioinformatics. For example genome sequencing will require bioinformatics for gene prediction and other genome annotation. Whether or not a complete wheat sequence is soon forthcoming, the similarity of the cereal genomes will make bioinformatics tools for comparative genomics critical for Triticeae research.

P.K. Gupta and R.K. Varshney (eds.), Cereal Genomics, 425-446.

In the initial stages of Triticeae genome research, several activities that were undertaken on priority included the following: (i) sequencing of a large number of cDNA clones to develop ESTs for wheat and barley, (ii) development of a map of thousands of markers relative to wheat chromosome deletion breakpoints (B.S. Gill, Chapter 20), (iii) construction of a physical map of the wheat D genome based on BAC fingerprints, (iv) increasing numbers of full-length genetic maps, and (v) targeted map studies for QTLs (quantitative trait loci), map-based cloning and marker-assisted selection. Availability of the ESTs is making it possible to mine for SSRs (simple sequence repeats) and SNPs (single nucleotide polymorphisms) for mapping and genotyping studies, and to assemble good unigene sets for gene expression experiments. Initial efforts in genome sequencing have focused on regions/loci of particular interest rather than using whole genomes or whole chromosomes for such studies; this has produced increasing numbers of full-length BAC (bacterial artificial chromosome) sequences of 100-300 kb.

During the last decade, the developments in the field of bioinformatics responded fully to the needs of Triticeae genomics research, but have not yet made it possible to integrate the results from different research projects and different areas of research with each other. For instance, the GrainGenes database, http://www.graingenes.org, (Matthews *et al.*, 2003) performs some overall integration of portions of the available data, but for in-depth research it will continue to be necessary to consult many separate information resources in order to obtain all the data that exists. An updated directory of such resources is maintained on the Triticeae Genomics page, http://wheat.pw.usda.gov/ggpages/genomics.shtml. Although, genomics research in all cereals makes use of bioinformatics, its impact is particularly felt in the genomics of the members of the tribe Triticeae, to which wheat (*Triticum aestivum*), barley (*Hordeum vulgare*) and rye (*Secale cereale*) belong. The use of current and planned information resources for only these cereals will be discussed in this chapter. In maize (*Zea mays*, Maydeae) and rice (*Oryza sativa*, Oryzeae), bioinformatics has been used as extensively as in the Triticeae, but these will not be covered in detail in this chapter except for purpose of reference.

2. EXPRESSED SEQUENCE TAGS (ESTs)

During the last few years, one of the major activities in Triticeae genomics research has been the large scale sequencing of cDNAs to produce a large number of ESTs. Since the establishment of the International Triticeae EST Cooperative (http://wheat.pw.usda.gov/genome/) in 1998, when only eight

public *Triticum aestivum* EST sequences were available, the contributions of many laboratories worldwide have produced over 500,000 wheat ESTs and over 300,000 barley ESTs (as of January, 2004). These ESTs each exceed in number the ESTs available in any other plant species, and are being put to use in a variety of ways including the following: (i) development of molecular markers, particularly the SSRs and SNPs; (ii) genetic and physical mapping of ESTs (transcribed sequences), (iii) comparative genomics through the use of individual markers across species, (iv) annotation of genomic sequences, and (iv) analysis of gene expression in time and space under different conditions of biotic and abiotic stresses.

2.1. Clustering and Assemblies

ESTs are ~500 bp long single sequence reads from individual clones chosen at random from a cDNA library, prepared from a specific tissue at a particular developmental stage. These libraries may sometimes also be prepared after subjecting the plants to an experimental treatment (Adams *et al*., 1991) to study the effect of a treatment (including biotic or abiotic stresses) on gene expression. Most of the Triticeae ESTs have been derived from non-subtracted, non-normalized cDNA libraries, so that each of the highly expressed genes is represented by a number of ESTs. An important use of bioinformatics in understanding and using the EST data is to group the ESTs into clusters, each cluster having ESTs derived from the same gene, the different clusters thus having ESTs derived from different genes. Clustering and/or assembly is achieved on the basis of similarities in overlapping sequences of the ESTs with each other.

"Clustering" is a process of building groups of ESTs, each group having ESTs with a defined level of similarity to each other, and optionally combining groups that contain common ESTs. "Assembly", on the other hand involves grouping of ESTs that can be aligned with each other in a satisfactory multiple sequence alignment. The latter approach is more selective but also more difficult computationally. For large numbers of ESTs, generally the assembly operation is not applied to the whole dataset, but to individual clusters prepared through an initial clustering step.

An EST clustering or assembly is always intended to divide the total set of ESTs into a set of unique groups, each group containing ESTs that can be considered equivalent in some way. An ideal situation would be that each group corresponds to a different gene. In practice, however, many assemblies are designed to provide appropriate grouping for the purpose of a specific experiment such as genetic mapping, SNP discovery, or gene expression.

For these purposes, it may be more appropriate to cluster together all homoeologous copies of the gene, sometimes including all paralogues, or even other genes that would cross-hybridize under the conditions of the planned experiment.

2.2. Problems in Grouping ESTs and Possible Errors in ESTs

Whatever be the desired goal of a given assembly, it can be achieved only imperfectly. ESTs are generally short, and a particular gene may have only one or a few representatives among the available ESTs, so that the sample can be too small in these cases to represent all ESTs belonging to a gene of interest. Multiple sequence alignment is a computationally difficult task, and especially error-prone for sequences that overlap only partially, as ESTs do. Moreover, among gene families, different alleles, homoeologues, and paralogues may also differ in actual level of sequence divergence, since different kinds of genes evolve under widely different kinds of selection. Further, genes with internal repeated subsequences wreak havoc on multiple sequence alignment. Given these difficulties, the available software does not guarantee either that dissimilar ESTs will be correctly segregated into separate clusters or that similar ESTs will be correctly included in the same clusters. The adjustable parameters provide only a coarse control over which kind of error predominates.

The complications described above would apply even in the absence of errors in the EST sequences themselves, and a variety of errors are also generally present in a dataset of the size of EST databases. These errors include incorrect base-calls, chimeric ESTs due to ligation of different cDNAs during cloning, and contamination with DNA from diverse other sources. Bioinformatics can be used to deal with the problem of contamination by excluding or trimming ESTs with regions of similarity to undesired sequences such as those belonging to an organelle, a vector, a cloning adapter, ribosomal DNA, or human genome. It may also be necessary to eliminate sequences that are repetitive in nature or bacterial in origin. Of the publicly available assemblies are summarized in Table 1, not all have been screened for all these contamination sources. Therefore, those who are involved in genomics research must be aware of the criteria and limitations of each assembly. In particular, screening for repeat sequences is dependent on knowledge of what the Triticeae repeat sequences are; the TREP database (Wicker *et al.*, 2002), described below, is being helpful for this, but it is not yet complete.

The base-calling software Phred (Ewing *et al*.,1998) also produces a "quality score" for each base. Some EST sequencing laboratories use these Phred scores only to trim off or mask out regions of their sequences with a score less than some threshold. However, some of the assembly programmes such as Phrap (Green, 2003) and CAP3 (Huang and Madan, 1999) can make use of the additional information content of the unmasked sequences in combination with the Phred scores to produce more accurate assemblies. The HarvEST assembly (Table 1) uses this approach whereas most others do not.

2.3. UniGenes (NCBI) and Gene Indices/ Tentative Consensi (TIGR)

The primary distinction between the NCBI UniGenes and the TIGR Gene Indices is that the former involves only a clustering step by sequence similarity, without subsequent assembly of each cluster by multiple sequence alignment. The assembly process frequently splits a cluster into subclusters called "contigs," within which the sequence alignment is closer; it also computes a consensus sequence for each contig. Thus for a given set of ESTs, the NCBI procedure will usually produce fewer UniGene groups than the number of Tentative Consensi (TCs) in the corresponding TIGR Gene Index. Having a consensus sequence is advantageous because it is nearly always longer than the longest EST in the contig, although sometimes it may be a sequence that doesn't actually exist in any of the cDNA clones sampled.

Both the NCBI and TIGR sites offer good information about the possible functions of the genes encoding these ESTs, as inferred from sequence similarity to known genes from other species. Many of the other publicly available assemblies also provide some annotation, but the function assignments based on reciprocal best matches (TIGR's EGO, NCBI's HomoloGene) are expected to be more reliable.

A difficulty encountered during any clustering or assembly exercise is to track the changes that occur during incorporation of new ESTs, which usually causes merging, splitting or rearrangement of some of the old clusters. NCBI and TIGR attempt to preserve the names of UniGenes and TCs between builds, and when this is not possible they cross-reference the old name to the current most closely corresponding new UniGene/TC. The Wheat SNP Development project uses a database that allows querying for contigs in different assemblies having common ESTs, http://www.graingenes.org/cgi-bin/ace/custom/aqlInterface/assemblies.

Table 1. Triticeae EST assemblies and clusterings

Project	Purpose	Species	Website
TIGR Gene Indices	General	Wheat, barley, rye	http://www.tigr.org/tdb/tgi/plant.shtml
NCBI UniGenes	General	Wheat, barley	http://www.ncbi.nlm.nih.gov/UniGene
PlantGDB	General	Wheat, barley, rye	http://www.plantgdb.org
Wheat SNP Development	SNP discovery	Wheat	http://wheat.pw.usda.gov/ITMI/2002/WheatSNP
CerealsDB	SNP discovery	Wheat	http://www.cerealsdb.uk.net/discover.htm
CerealsDB	Gene expression	Wheat	http://www.cerealsdb.uk.net/wheat.htm
HarvEST	Gene expression	Barley, wheat, rye	http://harvest.ucr.edu/
IPK Gatersleben	Gene expression	Barley	http://hordeum.ipk-gatersleben.de/est
NSF Wheat EST	Mapping	Wheat, rye	http://wheat.pw.usda.gov/NSF/data.html

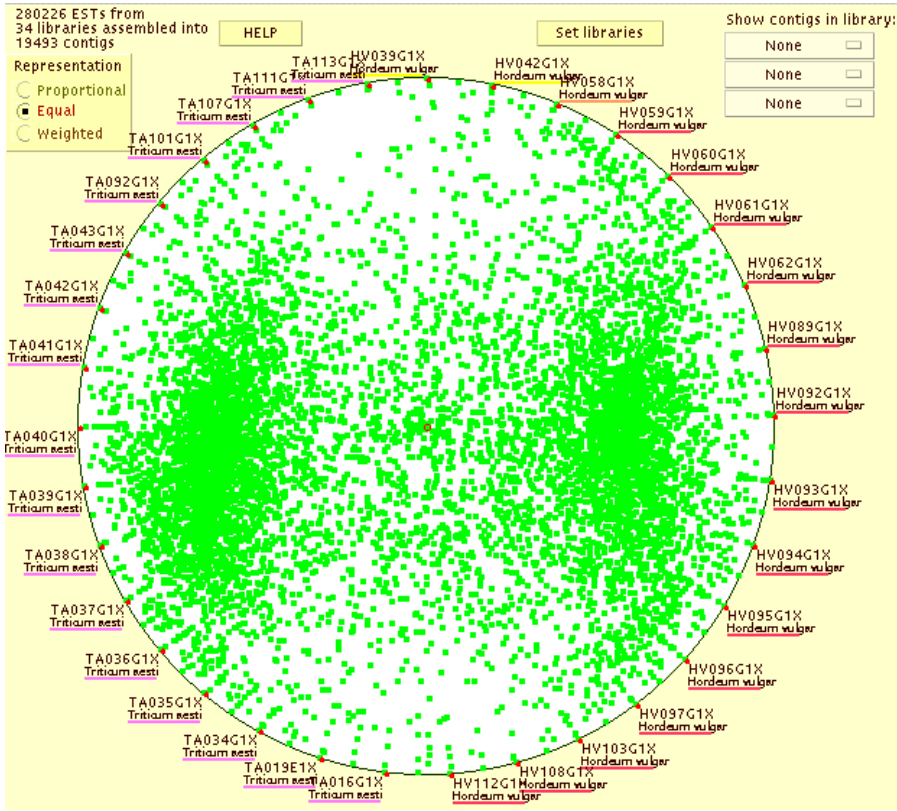


Figure 1. A co-assembly of wheat and barley ESTs.
Phrap assembly was performed on 776,000 ESTs from a total of 264 *Triticum aestivum* (TA) and *Hordeum vulgare* (HV) libraries. Results for the 17 largest libraries of each species (280,000 ESTs) are shown. Each dot represents a contig, and its position represents which libraries contributed to that contig. The 34 libraries are plotted as points on the circumference, TA libraries on the left and HV on the right. Each contig is plotted at the geometric center of the libraries from which its ESTs were derived. Further information about the Contig Constellation Viewer method is available at http://wheat.pw.usda.gov/demos/CCV/ and from G.R. Lazo.

Another approach is to make a BLAST database from the consensus sequences of each assembly, http://wheat.pw.usda.gov/ITMI/2002/WheatSNP/blastWheatSNP.

A question that has been raised is whether there should be a Triticeae UniGene Set, an assembly incorporating all ESTs from Triticeae species, in particular wheat and barley. This is not an unreasonable question given that wheat itself is an allopolyploid containing the genomes of three species that diverged more than two million years ago (Huang *et al.*, 2002). We have asked NCBI and TIGR about this and they responded that it would not be a good idea, because interspecies sequence differences would be too great. An experimental test is illustrated in Figure 1 (Lazo *et al.*, 2003), where by far

the majority of contigs are found either on the left (wheat) or right (barley) side of the graph, indicating that there is little admixture of the two species' ESTs in the same contigs. Further work on this question is warranted, including the related question of how ESTs from *T. aestivum*'s three "internal species" - the A, B and D genomes - behave during assembly.

2.4. Mapping Wheat ESTs

One of the largest research projects derived from the availability of wheat ESTs is the US NSF-sponsored project to map thousands of ESTs on a set of ca. 101 wheat deletion lines, http://wheat.pw.usda.gov/NSF/. To date over 7000 ESTs have been mapped, resulting in over 16,000 mapped loci.

Deletion mapping is performed by hybridizing the cDNA clone corresponding to each EST to a Southern blot of DNA from a panel of wheat deletion stocks, each missing a different terminal portion of a chromosome arm (Qi *et al*., 2003). In this exercise, bioinformatics was used to identify cDNA clones that were sufficiently different from each other, so that they would not cross-hybridize and produce redundant mapping data. Since cross-hybridization does occur among homoeologues of a gene and sometimes among members of a gene family, a very inclusive clustering was needed. The procedure involved preliminary assembly using Phrap, followed by BLAST of the resulting contigs and singletons against each other. A set of clones with less than 90% similarity to each other, based on their 5' and 3' EST sequences, was called "nonduplicated Unigenes" and used for mapping (http://wheat.pw.usda.gov/ NSF/ curator/ wheat_singletons.html).

A database of the mapping results is available at the project website, and can be searched via BLAST from GrainGenes, http://www.graingenes.org. Each EST is mapped to a "bin," corresponding to the interval between adjacent deletion breakpoints. The same set of deletion lines is also being used to map many of the RFLP and SSR markers that have been used for conventional recombination-based mapping, providing anchors to correlate the two kinds of maps.

2.5. EST-Derived Simple-Sequence Repeats (EST-SSRs)

A microsatellite (SSR) is a sequence consisting of short tandem repeats of a simple motif, 2 to 5 bp long. The number of repeats generally varies from 5 to 15, but the exact number for a particular SSR at a locus is highly

polymorphic. For this reason, SSRs are widely used as PCR-based molecular markers for genetic mapping and genotyping. In the past, development of SSR markers using genomic or cDNA clones has been an expensive process using conventional laboratory procedures. However using bioinformatics tools, SSRs can be mined from gene sequences or ESTs at no cost.

Examination of the relatively small number of barley gene sequences in GenBank in 1995 detected SSRs in fifteen of them (Becker and Heun, 1995). Indeed, SSRs can be found at surprisingly high frequencies in EST sequences (Kantety *et al.*, 2002; Varshney *et al.*, 2002). A search of 260,000 ESTs from wheat, barley, rice, maize, and sorghum detected 8500 ESTs that contained SSRs. After clustering and assembly to incorporate other overlapping ESTs and to eliminate redundancy, a total of 5,600 unique SSR-containing EST contig consensus sequences were identified among the five species (Kantety *et al*., 2002). In bread wheat, 10,415 unique SSRs were available when >415,000 ESTs were examined (P.K. Gupta, India, personal communication).

A significant advantage of using ESTs as the source of SSRs is the expectation that the flanking sequences may be relatively more conserved between species, compared to SSRs derived from total genomic DNA. Such conservation would provide useful markers for comparative mapping between species. When the above 5,600 SSR-ESTs from five species were combined and clustered together, 570 clusters containing at least two species were found (Kantety *et al*., 2002). Results of mapping in wheat and rice, and of primer testing in several cereal species, are online at http://wheat.pw.usda.gov/ITMI/EST-SSR/Cornell/, and have been reported in part (Yu *et al*., 2003; Yu *et al*., 2004).

A website for coordination of current research on Triticeae EST-SSRs has been organized by Nils Stein (IPK, Gatersleben, Germany) at http://wheat.pw.usda.gov /ITMI/2002/EST-SSR/. It includes access to the EST-SSRs that have been identified and submitted to the GrainGenes database.

2.6. EST-Derived Single-Nucleotide Polymorphisms (SNPs)

SNPs and short insertion/deletions (InDels) are another kind of polymorphism that is found at high frequency in EST data, and can be used for convenient genotyping of large number of individuals (Kota *et al*., 2001). SNPs can be discovered in the laboratory by using locus-specific primers to amplify DNA from several individuals and sequencing the amplified

products. This approach has been used successfully in barley to identify SNPs in a set of 38 DNA clones that had already been genetically mapped as RFLPs (Kanazin *et al.*, 2002). The results are available online at http://hordeum.oscs.montana.edu/locus/.

Now that such a large number of ESTs are available, SNP discovery can be performed on a larger scale computationally, by first assembling ESTs derived from different individuals, and then scanning the multiple sequence alignment of the contig to find sequence differences that correlate with the source of the EST. The scanning process is frequently done visually, but software has been designed to automate the correlations involved; such software includes AutoSNP (Barker *et al.*, 2003) and SNiPper (Kota *et al.*, 2003).

SNP discovery is also complicated by the polyploid nature of wheat, since a pure line of bread wheat is expected to contain three homoeologous copies of most genes, corresponding to the A, B and D genomes. Thus before allele-specific SNPs can be found by comparing individuals, it is first necessary to identify the sequences characteristic of the three homoeologues. Likewise, to make use of a SNP for genetic mapping it is desirable to design an assay that distinguishes not only the alleles belonging to the same locus, but also the homoeologues.

The Wheat SNP Development project, http://wheat.pw.usda.gov/ITMI/WheatSNP, was organized by Peter Isaac (AgroGene, Moissy Cramayel, France) in February 2002. The project is a cooperative, in which research groups volunteered to design and test SNP assays for a set of twenty contigs. The organizer allocated the contigs to each group in order to avoid duplication of effort. EST assemblies for the project were performed by Daryl Somers (AAFC, Winnipeg, Manitoba). For this purpose the assemblies were intended to include homoeologous genes in the same contig. Thus the parameters were adjusted to be more inclusive ("over-assembled") than would be the case for a general-purpose UniGene assembly.

Another wheat SNP database is available at CerealsDB, http://www.cerealsdb.uk.net/discover.htm. This database includes the results of analysis by the AutoSNP software, which correlates each SNP against the wheat lines in which it is found and provides values for cosegregation and informativeness. The database is searchable with BLAST, as are most of the other publicly available EST assemblies in Table 1.

When developing SNPs from EST data, it is desirable to consider as many relevant sequences as possible. BLAST searching against consensus

sequences of contigs from different assemblies is one useful approach. In addition, the AssembliesDB database, http://www.graingenes.org/cgi-bin/ace/search/assemblies, is structured to allow searching for contigs from different assemblies that correspond by virtue of having ESTs in common. It includes assemblies from both the Wheat SNP Development project and CerealsDB.

2.7. Use of ESTs in Gene Expression Studies

Another major use of ESTs is to determine the relative expression (transcription into mRNA) of genes during plant development or in response to individual experimental treatments. In the pre-EST era, such experiments were often done by extracting mRNA from the tissues to be compared, blotting it onto a filter, and hybridizing it with DNA of an individual cloned gene, the "Northern blot" procedure. A modern high-throughput approach, however, is to do the reverse, using mRNA preparations to probe a microarray of thousands of cloned genes, or of oligonucleotides representing characteristic subsequences of the genes. In Triticeae research, microarray experiments are currently being used for measuring gene expression to a greater extent than other technologies like SAGE (Serial Analysis of Gene Expression) or MPSS (Massively Parallel Signature Sequences). Therefore, we will focus on microarrays.

The first step is to identify a set of non-redundant genes to put on the array. The definition of non-redundancy for this purpose will depend on the experimental protocol to be used. When the array is to contain complete cDNA clones or long oligonucleotides, the potential for cross-hybridization of transcripts from different genes can be addressed by cross-BLASTing the candidates as described above for the "nonduplicated Unigenes" (http:// wheat.pw.usda.gov/ NSF/curator/wheat_singletons.html). The Affymetrix approach of representing each gene with a family of short oligonucleotides designed to maximize specificity (Lipshutz *et al*., 1999) is potentially capable of a higher degree of discrimination, between members of a gene family or perhaps even between homoeologues.

Several efforts are underway to develop microarrays for Triticeae. An Affymetrix array for barley has been produced and is now being used. For this purpose, a set of 21,000 unique gene sequences was generated from a pool of 349,000 ESTs by the HarvEST project, http://harvest.ucr.edu. Assembly was performed under "stringent" conditions, which tend to separate members of multi-gene families. The unigene set is available by downloading the HarvEST software package. It can be searched via BLAST

at PlantGDB, http://www.plantgdb.org, and at GrainGenes, http://www.graingenes.org, database "Barley1 GeneChip exemplars". Results of experiments using this array will be on-line in BarleyBase, http://www.barleybase.org. For wheat, an array of 10,000 cDNA clones is available from the cereals program at the University of Bristol, http://www.cerealsdb.uk.net/wheat.htm. Assembly and construction of the unigene set were based on a set of 26,000 ESTs. The data are available for searching or downloading at the CerealsDB website. The USDA-ARS EST project in California is also developing cDNA arrays of wheat. The current project is an array of 2000 genes expressed in maturing seeds. Progress may be monitored at http://wheat.pw.usda.gov/ microarray/.

A bioinformatic approach for assessing differential gene expression makes use of data based on "electronic Northerns" -- calculating the relative contribution of libraries prepared from different tissues or experimental conditions to the set of ESTs in a given contig. For instance, a contig composed primarily of ESTs from root libraries presumably represents a gene highly expressed in roots. To increase the power of such assessments, comparisons are frequently performed between contrasting sets of libraries, e.g. roots vs. leaves, seeds and spikes.

NCBI's Digital Differential Display (DDD), http://www.ncbi. nlm.nih.gov/UniGene/ddd.cgi?ORG=Ta, uses a statistical test to find NCBI UniGenes in which one set of libraries has a significantly higher representation than the other set. The user can construct his/her own library sets, based on a list of one-line descriptions of all libraries or individually checking the fuller documentation for each library. DDD pages are available for wheat, barley, and several other cereal species.

HarvEST, http://harvest.ucr.edu, is a stand-alone Windows application for the same purpose. It incorporates its own assemblies of Triticeae ESTs and also contains the analysis tools for exploring the correlations between libraries and contigs. HarvEST does not apply a statistical test to determine what data should be shown but rather allows the user to define and refine threshold levels interactively. The advantage of speed that is possible in a stand-alone application makes such user experimentation quite feasible. Accessory data such as functional annotation of each contig is also ready to hand. Recent releases include a viewer for the multiple sequence alignment of each contig, with the useful feature that it allows viewing the quality score for each base-call.

For all kinds of gene expression studies, a well-recognized issue is how to describe the experimental conditions underlying the mRNA preparation

being tested. The MIAME (Minimum Information About a Microarray Experiment) standards (Brazma *et al*., 2001) from the Microarray Gene Expression Data Society, http://www.mged.org/, are valuable in this connection. Beyond the standards for description, it would be desirable to structure and classify the developmental and experimental states, to allow querying for comparable conditions across different studies. In this regard the efforts of the Plant Ontology Consortium (The Plant Ontology™ Consortium, 2002) in developing ontologies for plant anatomy and developmental stages may prove helpful, see http://www.plantontology.org/.

In summary, the Triticeae now occupy an important position as a system for study of plant genomics, in view of the depth of available EST data, and the associated use of appropriate bioinformatics tools. Indeed, in some areas, such as SNP development, the number of available ESTs in wheat and barley make them more suitable than even species whose genomes have been fully sequenced. Like bioinformatics, ESTs are also valuable while sequencing whole genomes (see the following section).

3. GENOME SEQUENCES USING BACS

Although much immediately applicable information can be and is being extracted from the sequences of ESTs alone, this approach can never penetrate all the mysteries of the Triticeae genomes. The lack of detailed information on the higher order structure of these genomes is a critical knowledge gap in our understanding of them and in our ability to apply the most advanced genomic and bioinformatics tools to these crops. Work in this direction has already begun, and involves another set of bioinformatics applications.

3.1. Triticeae BACs (Bacterial Artificial Chromosome Clones)

Sequencing a region of a Triticeae genome usually begins with selection of a clone from a BAC library by hybridization with a cDNA clone or other marker for a region of interest. . Several online resources are available to assist in this step. A list with descriptions of the world's Triticeae BAC libraries is maintained by Jorge Dubcovsky at http://agronomy.ucdavis.edu/Dubcovsky/BAC-library/ITMIbac/ITMIBAC.htm. The barley cv. Morex library (Yu *et al*., 2000) has been probed with over 7500 genetically mapped clones, and the results are available in Andris Kleinhofs's "BACS" database

at http://barleygenomics.wsu.edu/databases/ databases.html; the database can be searched by genetic location (chromosome and bin number) as well as other criteria. Similarly, clones from *Aegilops tauschii* libraries totaling 12.8x genome coverage have been probed with ca. 500 genetically mapped clones, searchable in the WheatDB database at http://wheat.pw. usda.gov/PhysicalMapping/.

Once a BAC has been selected for sequencing, the rest of the work is heavily dependent on bioinformatics. The BAC is sequenced as many random short fragments which are then assembled, usually with PHRAP, and edited with a sequence alignment editor such as consed. To annotate the sequence, it is necessary to identify regions corresponding to interesting features. Software for predicting genes and their exons and introns includes FGENESH (www.softberry.com), GENSCAN, GeneMark, and ORF Finder. Sputnik (espressosoftware.com) can be used to find microsatellites. Inverted and direct repeats within the sequence can be detected visually using DotPlot. Finally, BLAST comparison of the BAC sequence with appropriate sequence databases is invaluable for annotation. BLAST alignment vs. ESTs from the same species and from other grasses can define the ends of genes and the intron/exon boundaries more reliably than motif-based prediction programs like GENSCAN. Likewise, a database of retrotransposons and other repeat elements such as TREP (see below) is useful in identifying and delineating these elements in the BAC sequence, where they may be fragmentary due to multiple nested insertion events.

By the end of 2003, twelve regions of Triticeae genomic DNA longer than 50 kb were sequenced and deposited in GenBank. Eight of these belonged to barley and two each to *Triticum monococcum* and *Aegilops tauschii* (Table 2). Sequences AY013246 and AF459639 are from orthologous regions of barley and *T. monococcum*, permitting comparison with each other as well as with rice.

The picture of the Triticeae genomes at present is similar to that of maize as opposed to rice: a sea of nested retrotransposons in which islands of ca. 1- 5 genes float. Within the islands a strong homology was found between Triticeae and rice at the level of gene order and orientation and intron/exon structure (SanMiguel *et al*., 2002). The distance between islands is small enough, so that two or more of them may be present in a single BAC. However, the BAC clones sequenced to date have a higher overall gene density than the genome as a whole, probably because they were pre-selected as containing at least one gene. For more information, see Chapter 12 by Kulvinder Gill in this book.

Table 2. Large genomic sequences from Triticeae

GenBank accession	Source	Chromosome arm	Gene or marker	Length (kb)	Predicted genes	Reference
Y14573	Barley	4HL	*Mlo*	60	3	Panstruga *et al.* (1998)
AF254799	Barley	2HL	*Rar1*	66	3	Shirasu *et al.* (2000)
AF474071	Barley cv. Morex	6H centromeric	RZ567	103	1	Rostoks *et al.* (2002)
AF474072	Barley cv. Morex	2HL	BCD135	114	4	Rostoks *et al.* (2002)
AF474373	Barley cv. Morex	7HS	pcWX27 (*Wx*)	124	10	Rostoks *et al.* (2002)
AF474982	Barley cv. Morex	Unknown	BCD 1434.2	78	5	Rostoks *et al.* (2002)
AF427791	Barley cv. Morex	1HS	*Mla*	261	32	Wie *et al.* (2002)
AY013246	Barley cv. Morex	5HL	WG644	102	5	Dubcovsky *et al.* (2001)
AF459639	*T. monococcum*	$5A^{m}L$	near *Vrn1*	215	5	SanMiguel *et al.* (2002)
AF326781	*T. monococcum*	$1A^{m}S$	*Lr10*	211	5	Wicker *et al.* (2002)
AF446141	*Ae. tauschii*	1DS	KSUD14, near *Lr21*	106	7	Brooks *et al.* (2002)
AF497474	*Ae. tauschii*	1DL	*Glu-D1*	103	8	Anderson *et al.* (2003)

The total genomic sequence available thus far accounts for less than 0.05% of a diploid Triticeae genome. Much remains to be learned. IGROW (International Genome Research on Wheat) (Gill, 2002) is the project organizing the international effort to sequence the gene-rich regions of the wheat genome.

3.2. Rice - Relevance to Triticeae

The availability of the complete sequence of the rice genome is an important resource for research on the Triticeae genomes. The degree of importance depends, of course, on how similar the genomes are. Early assessments based on comparative mapping of a few hundred RFLP markers showed large, nearly chromosome-sized blocks of conserved marker order (Ahn *et al.*, 1993; Kurata *et al.*, 1994; Moore *et al.,* 1995; Van Deynze *et al.*, 1995a; 1995b). Likewise, a high degree of conservation was observed within gene islands in the SanMiguel study cited above. However, at an intermediate level of resolution, seen by comparative mapping of thousands of wheat ESTs in wheat vs. rice, considerable rearrangement was apparent (Sorrells *et al.*, 2003, see Chapter 19 also). This finding may limit the applicability of rice to Triticeae research for some purposes, such as using the rice map to search for candidate genes corresponding to a QTL in Triticeae. The actual extent of the impact cannot be determined with the data currently at hand. Much more information about the relative organization of the rice and Triticeae genomes will be coming soon with the use of EST-derived SSRs and SNPs for recombination mapping in Triticeae, for higher resolution than the deletion-line mapping used in the Sorrells study.

Graphical interfaces for browsing the rice genome sequence with Triticeae ESTs aligned to it are available at the following sites: (a) Gramene, http://www.gramene.org/perl/contigview; (b) TIGR, http://www.tigr.org/tdb/tgi/ogi/alignTC.html; (c) GrainGenes, http://wheat.pw.usda.gov/cgi-bin/gbrowse?source=japonica. The service at Gramene includes searching for particular Triticeae ESTs with BLAST or by GenBank accession. Using these services, it is now easy to determine exactly where on the rice genome to find homologs of Triticeae ESTs and sequenced genes. Another benefit is that the alignment of ESTs to genomic sequence shows the location of probable introns which are a potential source of polymorphisms for PCR-based genetic markers. In addition, these alignments may be useful to rice researchers for improving the prediction of gene locations in the rice genome sequence, since rice has relatively few ESTs to use for this purpose.

Good information resources about the rice genome itself include Gramene (Ware *et al.,* 2002), the Rice Genome Research Program (RGP), and AGI. Gramene is especially useful for its curated annotation of rice proteins, and for comparative mapping among grass species. RGP's Rice Genome Annotation Database, http://ricegaas.dna.affrc.go.jp/rgadb/, is a good viewer for the predictions of genes and other features. AGI is the best source to consult for the FPC assemblies of BACs into contigs, http://www.genome.arizona.edu/shotgun/rice/status/. Additional information about rice and its applications in Triticeae research can be found in Chapter 18 authored by Sasaki and Antonio in this book.

3.3. Repeat Sequences - Overview and Web Resources

When sequencing any large region of a Triticeae genome, the most prominent feature will be repeat sequences. Indeed, repeats are even a problem in Triticeae ESTs (Echenique *et al.,* 2002). Some of the Triticeae repeat elements have been found before in other monocots and even dicots but many are new. Triticeae repeats are frequently arranged in complex patterns of element fragments due to transposition of new elements into old ones during evolution.

TREP, the Triticeae Repeat Sequence Database, (Wicker *et al*., 2002) was created to fill the need for organized information about Triticeae repeat elements. Beyond just compiling the known repeat sequences, curator Thomas Wicker has classified them and annotated their features such as long terminal repeats (LTRs) and coding regions. The five main groups in the classification are retrotransposons, foldback elements, tandem repeats, other transposons, and unclassified. These groups are classified further to two levels, producing subgroups such as LTR retrotransposons, *copia*, LINEs, MITEs, *Stowaway*, and *Tourist*. There is a complete database incorporating all known repeat sequences and also a nonredundant database containing one or two representatives of each subgroup. Release 4 of TREP contains 166 nonredundant and 809 total elements, 1.2 Mbp of sequence.

The TREP website, http://wheat.pw.usda.gov/ITMI/Repeats, allows browsing or downloading the annotated database, and BLAST searching or downloading the FASTA sequences. It also provides guidance to the community with regard to naming new repeat elements. Jorge Dubcovsky moderates a registry of existing names for repeat element types. Researchers

are encouraged to consult this registry before naming a new element they've discovered, and of course to register the name as soon as possible.

Another resource is the TIGR repeat sequence databases, one for rice and one for cereals which includes rice, maize, wheat, barley and sorghum. The cereals database comprises 2056 sequences, 1.4 Mbp in total. The databases can be searched with BLAST or downloaded as FASTA at http://www.tigr.org/tdb/e2k1/osa1/blastsearch.shtml.

4. PHYSICAL MAPPING

Until a Triticeae genome has been completely sequenced, the next most powerful tool will be a physical map of overlapping BAC clones. Such a map will greatly facilitate many research projects, such as chromosome walking from a known QTL marker in search of a candidate gene controlling the trait. It will also provide much more detailed information about the relation between physical distance and genetic distance across the genome than is possible using deletion-line mapping as the measure of physical distance.

A project is currently underway in the US to produce a physical map of the wheat D genome. The procedure is to compare restriction-enzyme fingerprints from hundreds of thousands of BAC clones and determine the overlaps among them, using the FPC software (Soderlund *et al.,* 2000) for assembly. Results are being posted on the project website as they are generated (http://wheat.pw.usda.gov/PhysicalMapping/). The status of the project at this writing, taken from the WheatDB database at the project website, is that 12349 contigs have been built from 175,000 BACs assembled thus far. The ten largest contigs contain 150 - 300 clones each. Two contigs contain hits from five genetic markers, four contigs contain four, and a total of 600 contigs contain at least one.

5. COMPARATIVE MAPPING

For comparing maps of different species based on common markers, the first place to look is Gramene (Ware *et al.,* 2002). This project has developed and deployed a good software interface called CMap for exploring comparative maps, http://gramene.org/cmap/viewer. The maps available include one each from wheat (Synthetic × Opata), barley (Steptoe × Morex), oat, and sorghum, several from maize, and a dozen from rice. For maize and rice, both genetic

and physical maps are included. The software is freely available and the GrainGenes project is currently installing it for use with its more extensive collection of Triticeae and oat maps.

When the starting point is not a known map region but a particular marker or sequence of interest, there are several other options. From the GrainGenes Database page, http://www.graingenes.org, (a) the "BLAST" link allows searching by DNA sequence in the databases "Mapped wheat ESTs" and "Mapping probes (BCD, CDO etc.)"; the page of BLAST results links to the corresponding map information. (b) the "Quick Queries" link offers a pre-defined query "Nearby Loci", which returns all loci within a specified distance of any desired locus on any of the ca. eighty wheat, barley, rye and oat maps in GrainGenes.

NCBI's Plant Genomes Central also provides a BLAST search against databases of sequences that have been mapped, from wheat, barley, oat, rice and maize. The search results page has a button "Genome View" which displays the map locations of the BLAST hits. This service is at http://www.ncbi.nlm.nih.gov/blast/Genome/PlantBlast.shtml.

6. SUMMARY AND OUTLOOK

Bioinformatics has been an important component of Triticeae genomics research for the last decade. It has played an integral role in the use of EST and genomic sequences for such purposes as (i) development of new molecular markers (EST-SSRs and EST-SNPs), (ii) study of gene expression, (iii) genetic and physical mapping, and (iv) comparative mapping of cereal genomes. No doubt bioinformatics will continue to be an essential research component for the foreseeable future, but different tools will be needed as Triticeae genome research expands into new areas. These research areas will almost certainly include (i) full-length cDNA sequencing, (ii) BAC end sequencing, (iii) leveraging from the D-genome physical map to the other Triticeae genomes, (iv) high resolution genetic mapping, (v) mining for gene-rich regions of the genome for sequencing, (vi) correlation of linked SNPs into haplotypes, and (vii) association and pedigree-based mapping of traits.

As in the past, many bioinformatics tools will be available from previous work on the genomics of other species, especially human, Arabidopsis, rice and maize. Another lesson to be learned from the past is that the results of this research will be scattered over dozens of websites. A special challenge for bioinformatics will be to integrate these results so that all data (e.g., about a particular gene), from any source and any research area, can be accessed

from a single point. This is the role of the genome databases such as GrainGenes and Gramene. Further development of genome databases for Triticeae will be assisted by the progress currently being made by the Generic Model Organism Database project, www.gmod.org, and other leaders in the bioinformatics community.

7. ACKNOWLEDGMENTS

This work was supported by the United States Department of Agriculture, Agricultural Research Service, CRIS #5325-21000-007-00D.

8. REFERENCES

Adams M.D., Kelley J.M., Gocayne J.D., Dubnick M., Polymeropoulos M.H., Xiao H., Merril C.R., Wu A., Olde B., Moreno R.F., *et al*. Complementary DNA sequencing: expressed sequence tags and human genome project. Science 1991; 252: 1651-1656

Ahn S., Anderson J.A., Sorrells M.E., Tanksley S.D. 1993. Homoeologous relationships of rice, wheat and maize chromosomes. Mol Gen Genet 1993; 241: 483-490

Anderson O.D., Rausch C., Moullet O., Lagudah E.S. The wheat D-genome HMW-glutenin locus: BAC sequencing, gene distribution, and retrotransposon clusters. Funct Integr Genom 2003; 3: 56-68

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

Becker J., Heun M. Barley microsatellites: allele variation and mapping. Plant Mol Biol 1995; 27: 835-845

Brazma A., Hingamp P., Quackenbush J., Sherlock G., Spellman P., Stoeckert C., Aach J., Ansorge W., Ball C.A., Causton H.C., *et al*. Minimum information about a microarray experiment (MIAME) -- toward standards for microarray data. Nature Genet 2001; 29: 365-371

Brooks S.A., Huang L., Gill B.S., Fellers J.P. 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 2002; 45: 963-972

Dubcovsky J., Ramakrishna W., SanMiguel P.J., Busso C.S., Yan L., Shiloff B.A., Bennetzen J.L.. Comparative sequence analysis of colinear barley and rice bacterial artificial chromosomes. Plant Physiol 2001; 125: 1342-1353

Echenique V., Stamova B., Wolters P., Lazo G., Carollo V.L., Dubcovsky J. Frequencies of Ty1-*copia* and Ty3-*gypsy* retroelements within the Triticeae EST databases. Theor Appl Genet 2002; 104: 840-844

Ewing B., Hillier L., Wendl M.C., Green P. Base-calling of automated sequencer traces using Phred. I. Accuracy Assessment. Genome Res 1998; 8: 175-185

Gill B.S. Introducing IGROW (International Genome Research on Wheat). Annual Wheat Newsletter 2002; 48: 15-17.
http://wheat.pw.usda.gov/ggpages/awn/48/Textfiles/IGROW.html

Green, P. The Phred/Phrap/Consed system home page. 2003; http://www.phrap.org

Huang, X., Madan, A. CAP3: A DNA sequence assembly program. Genome Res 1999; 9: 868-877

Huang S., Sirikhachornkit A., Su X., Faris J., Gill B., Haselkorn R., Gornicki P. Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the *Triticum/Aegilops* complex and the evolutionary history of polyploid wheat. Proc Natl Acad Sci USA 2002; 99: 8133-8138

Kanazin V., Talbert H,. See D., DeCamp P., Nevo E., Blake T. Discovery and assay of single-nucleotide polymorphisms in barley (*Hordeum vulgare*). Plant Mol Biol 2002; 48: 529-537

Kantety R.V., La Rota M., Matthews D.E., Sorrells M.E. Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum, and wheat. Plant Mol Biol 2002; 48: 501-510

Kurata, N., Moore, G., Nagamura, Y., Foote, T., Yano, M., Minobe, Y., Gale, M. Conservation of genomic structure between rice and wheat. Bio Technol 1994; 12: 276-278

Kota R., Wolf M., Michalek W., Graner A. Application of denaturing high-performance liquid chromatography for mapping of single nucleotide polymorphisms in barley (*Hordeum vulgare* L.). Genome 2001; 44: 523-528

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

Lazo, G.R., Lui, N., You, F., Hummel, D., Chao, S., and Anderson, O.D. Charting contig-component relationships within the Triticeae. In: J.P. Gustafson (ed.) *23rd Stadler Genetics Symposium, Genome Exploitation: Data Mining*, Kluwer Academic/Plenum Publishers, New York, 2004 (in press)

Lipshutz R.J., Fodor S.P., Gingeras T.R., Lockhart D.J. High density synthetic oligonucleotide arrays. Nature Genet 1999; 21: 20-24

Matthews D.E., Carollo V.L., Lazo G.R., Anderson O.D. GrainGenes, the genome database for small-grain crops. Nucl Acid Res 2003; 31: 183-186

Moore G., Devos K.M., Wang Z., Gale M.D. Cereal genome evolution. Grasses, line up and form a circle. Curr Biol 1995; 5: 737-739

Panstruga R., Buschges R., Piffanelli P., Schulze-Lefert P. A contiguous 60 kb genomic stretch from barley reveals molecular evidence for gene islands in a monocot genome. Nucl Acid Res 1998; 26: 1056-1062

Qi L., Echalier B., Friebe B., Gill B.S. Molecular characterization of a set of wheat deletion stocks for use in chromosome bin mapping of ESTs. Funct Integr Genom 2003; 3: 39-55

Rostoks N., Park Y.J., Ramakrishna W., Ma J., Druka A., Shiloff B.A., SanMiguel P.J., Jiang Z., Brueggeman R., Sandhu D., Gill K., Bennetzen J.L., Kleinhofs A. Genomic sequencing reveals gene content, genomic organization, and recombination relationships in barley. Funct Integr Genom 2002; 2: 51-59

SanMiguel P.J., Ramakrishna W., Bennetzen J.L., Busso C.S., Dubcovsky J. Transposable elements, genes and recombination in a 215-kb contig from wheat chromosome $5A^m$. Funct Integr Genom 2002; 2: 70-80

Shirsu K., Schulman A.H., Lahaye T., Schulze-Lefert P. A contiguous 66-kb barley DNA sequence provides evidence for reversible genome expansion. Genome Res 2000; 10: 908-915

Soderlund C., Humphrey S., Dunhum A., French L. Contigs built with fingerprints, markers and FPC V4.7. Genome Res 2000; 10: 1772-1787

Sorrells, M E., La Rota M., Bermudez-Kandianis C.E., Greene R.A., Kantety R., Munkvold J.D., Miftahudin N., Mahmoud A., Ma X., Gustafson P.J., *et al*. Comparative DNA sequence analysis of wheat and rice genomes. Genome Res 2003; 13: 1818-1827

The Plant Ontology™ Consortium. The Plant Ontology™ Consortium and plant ontologies. Comp Funct Genom 2002; 3: 137-142

Van Deynze, A.E., Nelson J.C., O'Donoughue L.S., Ahn S.N., Siripoonwiwat W., Harrington S.E., Yglesias E.S., Braga D.P., McCouch S.R., Sorrells M.E. Comparative mapping in grasses. Oat relationships. Mol Gen Genet 1995a; 249: 349-356

Van Deynze, A.E., Nelson J.C., Yglesias E.S., Harrington S.E., Braga D.P., McCouch S.R., Sorrells M.E. Comparative mapping in grasses. Wheat relationships. Mol Gen Genet 1995b; 248: 744-754

Varshney R.K., Thiel T., Stein N., Langridge P., Graner A. *In silico* analysis on frequency and distribution of microsatellites in ESTs of some cereal species. Cell Mol Biol Lett 2002; 7: 537-546

Ware D.H., Jaiswal P., Ni J., Yap I.V., Pan X., Clark K.Y., Teytelman L., Schmidt S.C., Zhao W., Chang K., Cartinhour S., Stein L.D., McCouch S.R. Gramene, a tool for grass genomics. Plant Physiol 2002; 130: 1606-1613

Wei F., Wing R.A., Wise R.P. Genome dynamics and evolution of the *Mla* (powdery mildew) resistance locus in barley. Plant Cell 2002; 14: 1903-1917

Wicker T., Stein N., Albar L., Feuillet C., Schlagenhauf E., Keller B. Analysis of a contiguous 211 kb sequence in diploid wheat (*Triticum monococcum* L.) reveals multiple mechanisms of genome evolution. Plant Jour 2001; 26: 307-316

Wicker T., Matthews D.E., Keller B. TREP: a database for Triticeae repetitive elements. Trends Plant Sci 2002; 7: 561-562

Yu J.-K., Dake T.M., Singh S., La Rota M., Kantety R.V., Sorrells M.E. Comparative analysis of orthologous SSR loci from EST of the wheat and rice genomes. In: S.R. Heller (ed.) *Abstracts of Plant & Animal Genome XI Conference, 2003*. http://www.intl-pag.org/11/abstracts/W13_W93_XI.html

Yu J.-K., Dake T.M., Singh S., Benscher D., Li W., Gill B., Sorrells M.E. Development and mapping of EST-derived simple sequence repeat (SSR) markers for hexaploid wheat. Genome 2004 (in press)

Yu Y., Tomkins J.P., Waugh R., Frisch D.A., Kudrna D., Kleinhofs A., Brueggeman R.S., Muehlbauer G.J., Wise R.P., Wing R.A. A bacterial artificial chromosome library for barley (*Hordeum vulgare* L.) and the identification of clones containing putative resistance genes. Theor Appl Genet 2000; 101: 1093-1099

Chapter 15

FUNCTIONAL GENOMICS OF SEED DEVELOPMENT IN CEREALS

Andrew S. Milligan, Sergiy Lopato and Peter Langridge*

Australian Centre for Plant Functional Genomics, The University of Adelaide, PMB1, Glen Osmond, SA 5064, Australia.
*Author for correspondence: peter.langridge@adelaide.edu.au

1. INTRODUCTION

The cereal grains represent the most important human food, accounting for the major source of calories for the bulk of the world's population. Over the past century there has been a spectacular increase in yields of all major cereals largely due to the breeding of improved lines and advances in agricultural practices. These advances have followed the application of new technologies. The introduction of computing and mechanization to breeding programmes was critical in the early stages of modern breeding; more recently we have seen the use of hybrids in maize breeding and doubled haploids in wheat and barley breeding. For all the cereals the development and application of molecular markers has represented a major advance. Genetic engineering has also been important in some countries although political resistance has limited its use in other parts of the world. Does genomics represent the next major technology for enhancing grain quality and yield?

The long history of cereal breeding and research has provided considerable data on the genetics of many important traits related to grain quality. However, the highly diverse end-uses of cereal grain have made the analysis of grain quality complex. For example, barley breeding programmes have traditionally focused on a single end-use, malting and brewing. However, more recently efforts have also started to address the animal feed characteristics of barley grain. For wheat, there are very diverse end-uses

P.K. Gupta and R.K. Varshney (eds.), Cereal Genomics, 447-481.

ranging from the many different types of bread and noodles to animal feed and industrial processing for products such as gluten and starch. In the case of maize, it has been estimated that there are over 3,500 different products produced from grain. Developing a molecular understanding of the basis for these different end-uses will influence the future directions for cereal breeding. Genomics and molecular analysis of grain development will also be important in understanding the impact of disease and environmental stresses on grain quality and on studying the relationship between grain development, seedling establishment and early growth.

2. CEREAL GRAIN DEVELOPMENT / ANATOMY

In angiosperms, two major organs develop within the seed after fertilization. Fertilization of the egg cell gives rise to a single-celled zygote that subsequently develops, through the three broad phases of morphogenesis, maturation and dormancy, into a mature, developmentally arrested embryo. At the same time, fertilization by a second sperm of the two fused nuclei of the central cell gives rise to the triploid, primary endosperm nucleus and eventually to an endosperm largely comprising amorphous storage reserves. Cereals, as members of the Poaceae, are albuminous: the endosperm is a prominent part of the mature seed. In the grasses, each single-seeded fruit, or caryopsis, is produced from an individual carpel, which itself contributes seed coats and pericarp tissue to the mature seed (Figure 1a).

2.1 Embryo Development

Early development of the embryo (for a review, see Mordhorst *et al.*, 1997) is driven largely by the polarity inherent in the egg and zygote, being located at the micropylar pole of the embryo sac with the basal portion attached to the embryo sac wall. The first cell division of the zygote is asymmetric, giving rise to a two-celled proembryo. In cereals, the smaller apical cell is dense in cytoplasm and is destined to become the embryo proper, while the larger basal cell, which inherits the large vacuole of the zygote, develops into the suspensor, a conduit for nutrients from the surrounding tissue (Figure 1B). Henceforth, progressive cell divisions, cell expansion and differentiation give rise to the meristems and embryonic organs, providing a prototype for the adult plant. Initially, the embryo relies on nutrition from the endosperm and the surrounding maternal tissue but, during the maturation phase, accumulates its own storage reserves, principally starch, globulins, albumins and lipids (Raghavan, 1997). Embryogenesis terminates with a dormancy period, at

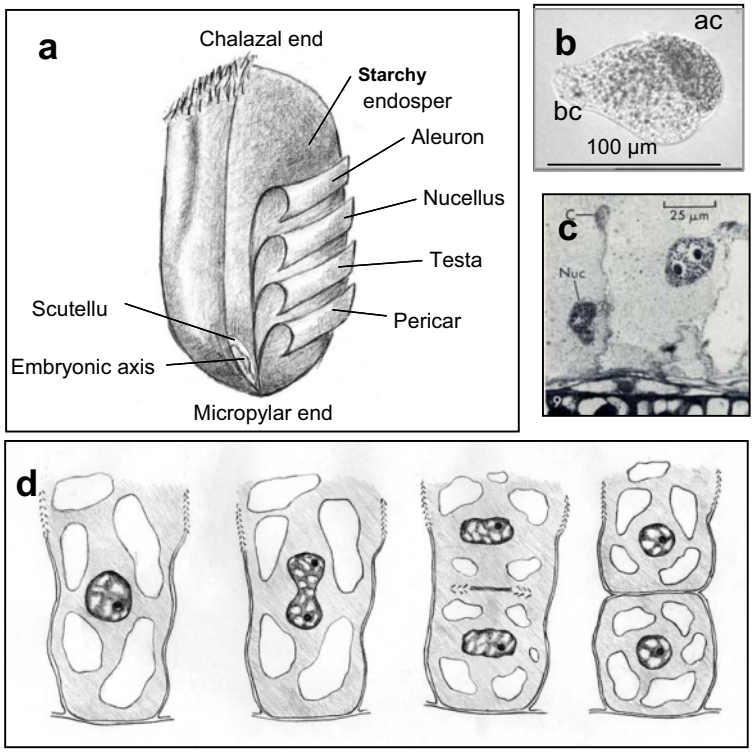


Figure 1**. Cereal grain development. (a) Generalised anatomy of a cereal grain** (adapted from Kent and Evers, 1994); **(b) A wheat two-celled proembryo, 24 hours after pollination** (ac, apical cell; bc, basal cell); **(c) Cross-section of syncytial endosperm showing anticlinal cell wall projection with free ends and the first stage of nuclear division** (reproduced from Mares *et al.*, 1975, with permission); **(d) Diagrammatic representation of the formation of the first cellular layer of the endosperm.**

which time metabolism has ceased and the cells of the embryo and the surrounding seed layers have become dehydrated (Goldberg *et al*., 1994).

Traditionally, cereal breeders have paid little attention to the nutritional quality of the embryo, except in maize where the large embryo contributes significantly to the overall grain oil and essential amino acid composition. The status of the embryo at germination, though, will impact upon early seedling vigour, so breeders may subconsciously have been selecting for this trait for some time. The challenges for genomics now are to identify the 'master regulators' of early embryo development that determine the final embryo shape and size, and to further elucidate the molecular mechanisms that control dormancy and germination. In addition, a greater understanding of the structure and behaviour of the egg cell and early zygote development

will contribute towards the effort to develop novel apomictic systems in cereal crops (see Section 5.1).

2.2. Endosperm Development

Endosperm development proceeds, following fertilization of the central cell, with a series of nuclear divisions giving rise to a multi-nucleate coenocyte (for a review see Olsen, 2001). The multiple nuclei are localized to the periphery of the coenocyte around a large central vacuole. Cell wall synthesis in the cereal endosperm begins at about five days after pollination, depending on environmental conditions and the species, and also the cultivar or line. The growing cell walls extend anticlinally, separating the syncytial nuclei, before periclinal cell divisions and cytokinesis yield a fully cellular layer at the periphery (Figures 1c and 1d). This process repeats until the endosperm is fully cellularized, and the subsequent enlargement of cells is accompanied by the accumulation of starch and protein bodies. Eventually, after cell differentiation, the endosperm comprises starchy endosperm cells, transfer cells that facilitate nutrient uptake from the maternal tissues, and aleurone.

The starchy endosperm is undoubtedly the grain component with the greatest economic value, since it provides the major source of calories to man's diet worldwide. The starchy endosperm cells contain starch granules embedded in a matrix of storage proteins, the profiles of which differ between different cereal species, and which impart the particular physical characteristics which determine the grain end-use, e.g. bread-making, pasta production, brewing, etc. The endosperm of wheat, the most widely grown grain crop in the world, is rich in prolamins, composed of the glutenin and gliadin fractions. Maize and barley also have high prolamin contents, while in oats and rice the globulins (defined classically as the glutelins in rice) account for 70-80% of the storage proteins (Shewry and Halford, 2002). An understanding of the regulation of storage protein, starch and lipid synthesis in the endosperm is of great importance for the engineering of improved grain quality for specific end-uses. In terms of the biology of endosperm development, genomic imprinting, positional cues, cell cycle modifications and programmed cell death all figure prominently as regulatory inputs (Becraft, 2001).

The other cell types of the endosperm, although less obviously contributing to the value of the grain, may also be targets for manipulation. For example, increased yield might be achieved by enhancing the uptake of soluble nutrients through the cells of the transfer layer (Becraft, 2001). Alternatively, seed composition can be manipulated by alterations in the aleurone layer, where the cells already have high concentrations of lipid, protein, phytic acid,

minerals and vitamins. The main interest in the aleurone is from the brewing industry as it is the major source of amylase in the malting process. Other targets include reducing the amount of phytic acid in the aleurone in order to improve the nutritional quality of animal feed (Ertl *et al*., 1998).

2.3. The Nucellus, Testa and Pericarp

The cereal embryo and endosperm develop within a mass of maternal tissue called the nucellus. The nucellus contributes little to the nutritional quality of the grain, with most cells degenerating by a process of programmed cell death (Domínguez *et al*., 2001) so that all that remains in the mature grain is the epidermis with a thin outer cuticle. However, the nucellus provides nutrients to the endosperm (Thorne, 1985) and is therefore an important tissue in grain biology.

The nucellus is surrounded by the testa, usually formed from the inner and outer integuments of the carpel, which is considered to be the true seed coat. In wheat there is a correlation between grain colour conferred by pigmentation in the testa and resistance to pre-harvest sprouting, suggesting a direct influence of the *R* pigmentation genes on coat-imposed seed dormancy (Groos *et al*., 2002).

Finally, the outermost layer of the cereal grain is the pericarp, consisting of dry, empty cells when mature. The presence of simple vascular tissues in the pericarp of wheat and rice (Evers and Millar, 2002) suggests that it, too, has a role in nutrient provision to the endosperm in early stages of grain development.

3. GENOMICS, TRANSCRIPTOMICS, PROTEOMICS AND METABOLOMICS

3.1. Genetics in Cereal Breeding and Gene Discovery

The pool of knowledge on cereal grain genetics is enormous. This is largely due to the significance of cereal grains for human and animal food and the extensive breeding programme worldwide for these crops. A report on public sector plant breeding by Heisey *et al*. (2001) estimated that over 50% of plant breeders in the USA were involved in cereal breeding (equivalent to nearly 800 full-time workers). In Australia, over 70% of funds devoted to plant breeding in the public sector are directed to cereal breeding and this translates

to over US$20 million per annum (Clements *et al*., 1992). The result of this work has been a consistent improvement in the yield and quality of cereal grains and has provided extensive background resources for genomics studies.

3.1.1. Grain Characters and the Genes Involved

Information on the genetic location of key loci influencing the characteristics and development of grain has provided a valuable entry point into the isolation of key genes. Genetic variation in grain characteristics has clearly been important in breeding varieties with improved quality and processing characteristics for all cereals. These include grain shape and size, starch composition, protein composition, colour, texture and a range of other characteristics. Several biological characteristics have also been important to cereal breeders, in particular germination and dormancy. The importance of these characters has meant that reliable genetic data and genetic locations of key controlling traits have been mapped in wheat, barley, maize and rice (for a review of mapping work in wheat, see Langridge *et al*., 2001). The genetic data sets and associated mapping populations provide a powerful resource for the cloning and analysis of genes controlling grain development and the properties of mature grain. Indeed, the extent of genetic information and analysis of grain characteristics has the potential to become a key tool in studying the genomics of grain development. However, these resources have only been poorly exploited to date.

The importance of cereal grains and the relationship between grain composition and quality has meant that considerable effort has focused on relating specific genes to known quality variants. In several cases, the genes controlling known quality loci have been identified through the knowledge of linkage. These include: the puroindolines, responsible for grain texture, in wheat and barley (Rahman *et al*., 1994; Gautier *et al*., 1994; Morris, 2002); the *Waxy* genes (granule-bound starch synthase) in maize (Klösgen *et al*., 1986), barley (Rohde *et al*., 1988), rice (Wang *et al*., 1990) and wheat (Murai *et al*., 1999); β-amylase in barley, which plays a key role in determining malting quality (Yoshigi *et al*., 1995); the *opaque* mutants of maize (Lohmer *et al*., 1991); and the high and low molecular weight glutenins in wheat that are major determinants of dough characteristics (Colot *et al*., 1989; Reddy and Appels, 1993).

The genetic data also provide a useful means for testing the possible significance of candidate genes for grain development. Co-segregation of a

Figure 2. Natural variation for grain morphology in cultivated and wild species of barley. a. Mundah – large plump grain, thick husk, slightly yellow colour; b. BX98A – plump grain with thin wrinkled husk. Represents high extract malting varieties; c. Tadmor – Syrian landrace, black seeded; d. HB104 – Hulless barley; e. Sahara – North African landrace, small grain, blue aleurone; f. *Hordeum spontaneum*, wild barley. Small grain typical of accessions from desert regions; g. *Hordeum spontaneum*, wild barley. Long grain typical of accessions from high rainfall regions.

candidate gene with a known grain related trait provides evidence for a role of the candidate gene.

In addition to genetic variants of grain characteristics found in cultivars and land races (see Figure 2), mutants have been identified or generated for several grain features. These can be broadly classified into the following groups: (i) altered hormone responses to gibberellic or abscisic acid (e.g. Ikeda *et al*., 2001; Yao *et al*., 2002); (ii) altered germination (e.g. Swanston *et al*., 2002); (iii) altered grain composition, such as low phytic acid lines in barley (Larson *et al*., 1998), rice (Larson *et al*., 2000) and maize (Raboy *et al*., 2000; Shi *et al*., 2003), and altered endosperm composition (Domon *et al*., 2002); (iv) altered grain structure in the endosperm (e.g. Patron *et al*., 2002; Zhang *et al*., 2002), embryo (e.g. Elster *et al*., 2000) and aleurone (e.g. Becraft *et al*., 2002).

3.1.2. Information on Mutants

Information on mutants including images of mutant phenotypes can be found at the following websites: barley (http://www.untamo.net/cgi-bin/ace/searches/basic); maize (http://www.maizegdb.org/cgi-bin/imag

ebrowsermutants_by_mutation.cgi); and rice (http://www.gramene.org/rice _mutant/).

3.1.3. Comparative Genetics

Comparative genetics is another powerful tool for gene discovery in cereals, either beginning with a gene sequence from a model organism and searching for sequence similarity among ESTs or protein sequences, or exploiting conserved collinearity to isolate genes by map-based cloning (see Chapter 5 by Paterson and Chapter 11 by Stein and Graner in this book). The most useful models for comparative genetics studies in the cereals, with complete draft sequences of their genomes, are Arabidopsis, for the quality of genome annotations and an extensive collection of mutants, and rice, offering large degrees of synteny with the other cereals. Several studies have demonstrated conservation of gene order and gene content among the cereals (Ahn *et al*., 1993; Moore *et al*., 1995; Devos and Gale, 2000; Smilde *et al*., 2001; Goff *et al*., 2002), meaning that the rice genome data may be extrapolated to wheat, maize, barley and others. For example, recent reports describe the use of syntenic maps as tools in the cloning of the wheat vernalization gene *VRN1* (Yan *et al*., 2003a) and the barley disease resistance gene *ROR2* (Collins *et al*., 2003). The transfer of information and tools provided by comparative studies can be similarly applied to the study and improvement of grain characteristics. Cereal seed development, however, differs substantially from seed development in Arabidopsis, and there may be many genes in temperate cereals, belonging to Triticeae, which have no orthologues in the tropical cereal species, rice.

3.2. ESTs and Gene Discovery

Expressed sequence tags or ESTs, obtained by single-pass sequencing of randomly chosen cDNA clones, present an affordable and relatively fast route to gene discovery. Most of the cereal EST projects coordinated in the public sector have focused on wheat and barley, reflecting the economic importance of these crops coupled with their massive genomes (16,000 Mb and 4,500 Mb respectively), while the relative paucity of EST data for rice is due to resources for this crop being directed towards whole genome sequencing. At the time of writing, there were over 1.5 million cereal EST sequences available in the public databases (see Chapter 14 by Matthews *et al*. in this book). Grain-related ESTs make a large proportion: over 45% of wheat ESTs are derived from such tissues (Table 1). These ESTs originate from cDNA libraries prepared from highly specific cell types, such as egg cells and

Table 1. A list of grain related wheat ESTs

Cultivar	Tissue	No. of ESTs	dbEST Lib ID
Wyuna	Endosperm	1,011	3736
Cheyenne	Endosperm 5-30 DPA	2,139	5449
Wyuna	Endosperm 8-12 DPA	1,047	5450
Wyuna	Endosperm	1,152	5454
Cheyenne	Endosperm 5-30 DPA	2,824	5468
Soleil	Endosperm, 1:1 mix 10:20 DPA	230	5472
Novosibirskaya 67	Pericarp/testa, 3:1 mix 10:20 DPA	571	5473
Chinese Spring	Spike pre-anthesis	13,397	5528
Brevor	Embryo mature, ABA-treated	2,173	5552
Sumai3	Spike, adult plant, *Fusarium*-infected	727	5584
Chinese Spring	Spike 20 DPA, grain 30 & 45 DPA	1,208	5592
Chinese Spring	Spike 5-20 DPA, heat-stressed	1,275	5597
Chinese Spring	Spike 5-15 DPA	3,454	5603
Brevor	Embryo mature	2,954	8825
Frontana	Spike 1 DPA, *Fusarium*-infected	1,586	9788
Sumai3	Spike 0-2 DPA, *Fusarium*-infected	5,500	9984
Chinese Spring	Grain 10 DPA	11,282	10469
Chinese Spring	Spike at flowering date	12,300	10470
Chinese Spring	Spike at heading date	11,285	10474
Chinese Spring	Pistil at heading date	10,349	10475
Chinese Spring	Grain 30 DPA	12,573	10479
Chinese Spring	Spikelet at late flowering	12,030	10480
Chinese Spring	Spikelet at early flowering	11,072	10481
Glenlea	Grain 5 DPA	5,881	10545
Glenlea	Grain 15 DPA	5,496	10546
Wyuna	Endosperm 8, 10 & 12 DPA	4,433	10946
Mercia	Embryo 28 DPA	514	10988
Mercia	Embryo 40 DPA	748	10989
Mercia	Embryo 21 DPA	483	10990
Mercia	Embryo 30 DPA	579	10991
Mercia	Embryo 1 day post-germination	948	10992
Mercia	Embryo 2 days post-germination	1,013	10993
Mercia	Aleurone 2 days post-germination	747	10994
Mercia	Embryo 14 DPA	514	10998
Florida	Ovule 2 DPA	701	11003
Florida	Ovule pre-fertilisation	767	11005
Florida	Egg cell	654	11006
Butte 86	Grain 3-44 DPA	3,649	11095
93FHB37	Spike 1 DPA, *Fusarium*-infected	1,230	11142
FHB148	Spike 1 DPA, *Fusarium*-infected	1,697	11143
Harus	Spike 1 DPA, *Fusarium*-infected	1,724	11144

Table 1. Continued

Unknown	Grain	126	11910
Hard red spring	Grain malted 55 h at 22°C	1,000	12171
Hard red spring	Grain malted 175 h at 4°C	976	12172
Unknown	Grain 3 DPA	4,860	12173
Spring wheat	Grain 7 DPA	5,017	12174
Spring wheat	Grain 14 DPA	5,054	12175
Spring wheat	Grain 21 DPA	971	12176
Spring wheat	Grain 30 DPA	854	12177
Spring wheat	Grain malted 175 h at 4°C, normalised	2,469	12178
Spring wheat	Grain 3, 7, 14 & 21 DPA	331	12180
Spring wheat	Grain 3, 7, 14 & 21 DPA, normalised	314	12181
Hi Line	Endosperm 2-7 DPA	738	12188
Hi Line	Pistil immature	1,955	12190
Hi Line	Endosperm 8-15 DPA	664	12192
Hi Line	Embryo 14 DPA	249	12196
Cheyenne	Endosperm 5-30 DPA	1,216	13732
Cheyenne	Endosperm 5-30 DPA, subtracted	144	13733
Sumai3	Spike 0-2 DPA, *Fusarium*-infected	418	13766
Chinese Spring	Spike 5, 10 & 15 DPA	115	13785
Chinese Spring	Spike pre-anthesis	871	13786
Chinese Spring	Spike 5-20 DPA, heat-stressed	47	13787
Chinese Spring	Spike 20 DPA, grain 30 & 45 DPA	46	13799
Brevor	Embryo mature	148	13822
Recital	Grain 118 degree-days after pollination	73	14011
Recital	Grain 45 degree-days after pollination	85	14012
Recital	Grain 118 degree-days after pollination	9,830	14013
Recital	Grain 174 degree-days after pollination	5,165	14014
Recital	Grain 356 degree-days after pollination	4,249	14015
Recital	Grain 468 degree-days after pollination	5,802	14016
Recital	Grain 550 degree-days after pollination	5,951	14017
Recital	Grain 608 degree-days after pollination	5,261	14018
Recital	Grain 750 degree-days after pollination	4,864	14019
Recital	Grain 45 degree-days after pollination	7,956	14020
Recital	Ovary	5,311	14021

Wheat cDNA libraries prepared from developing grain, components of the grain or spikes, from immediately pre-anthesis to post-germination. Libraries prepared from processed grain are also included. DPA, days post anthesis.

aleurone cells, to more complex tissue types, such as whole spikes. There are a large number of embryo and endosperm libraries, covering many stages of development, and there are additional libraries from tissues which have in some way been treated, including *Fusarium* infection, heat stress and malting.

Library normalization and subtraction have also been carried out in some cases, increasing the likelihood of finding rare clones.

ESTs can be exploited in the search for candidate genes involved in grain development, by comparing the distribution of ESTs among libraries from different tissues (see Section 3.3). Additionally, proteins belonging to particular groups like transcription factors, likely to be key regulators of developmental processes, can be identified by sequence homology to genes of known function (see Section 4.1).

3.3. Gene Expression Profiling For Gene Discovery

3.3.1. Techniques for Gene Expression Profiling

A wide range of techniques based on cDNA subtraction or differential display have been used for comparing gene expression between cell types, plants exposed to or shielded from stresses, or tissues at different developmental stages. These techniques have been useful but they provide only a partial picture of gene expression changes. The extensive EST databases prepared from many different tissues (e.g., Table 1) can be used to estimate gene expression levels by measuring the frequency of appearance of specific sequences, employing computational tools such as Digital Differential Display (http://www.ncbi.nlm.nih.gov/UniGene/info_ddd.shtml) or HarvEST (http://harvest.ucr.edu/) to identify grain-specific genes. An example of the use of ESTs from multiple cDNA libraries to study developmental processes can be seen for wheat in Ogihara *et al.* (2003). This paper described the analysis of 116,232 sequences generated from ten wheat tissues including grain at 10 and 20 days after flowering. However, the use of EST databases to study expression profiles is limited by the availability of cDNA libraries used to develop ESTs, by the depth of EST sequencing. There are also problems in tracking genes that may be represented by several partial EST sequences. Newer techniques allow the estimation of mRNA abundance for large numbers of genes simultaneously. The methods include Serial Analysis of Gene Expression (SAGE), microarrays, macroarrays and Massively Parallel Signature Sequencing (MPSS). Generally, these methods have not been extensively applied to the study of grain development although nearly all have been applied to some aspects of cereal development or stress response.

3.1.1.1. SAGE

SAGE has been successfully applied in barley to the study of pathogen interactions (Thomas *et al.*, 2002) and in rice to look at gene expression

during seedling development (Matsumura *et al.*, 1999). Unfortunately, SAGE does suffer from several problems. In particular, SAGE experiments require large amounts of RNA and can be very expensive if many samples are to be analysed, for example from a developmental series. As with MPSS, the signatures generated can be difficult to assign to particular genes when the technique is applied to wheat, barley, maize or other cereals where a full genome sequence is not available. In the rice example, SAGE was used to produce 10,122 tags derived from 5,921 genes expressed in seedlings. Only 1,367 genes (23.1%) matched the rice cDNA or EST sequences in the DNA database. The majority of highly expressed genes in seedlings belong to the housekeeping genes (Matsumura *et al*., 1999).

3.1.1.2. Microarrays and Macroarrays

Microarrays and macroarrays offer a technique for screening the expression profile of very large numbers of genes simultaneously. Both types of arrays have been used to study grain development in cereals. Macroarrays have the advantage of ease of manufacture and low cost relative to microarrays but macroarrays do not provide the same level of gene or probe density for screening.

The use of a cDNA macroarray to study early caryopsis development in barley has been described (Sreenivasulu *et al*., 2002). The array was small, composed of 711 cDNA fragments prepared from an early caryopsis library and covering about 620 diverse genes. RNA from maternal pericarp and embryo sac tissues was prepared from one to seven days after flowering. The results showed up-regulation of 26 genes in the pericarp and 12 genes in the embryo sac. A larger array of 6,144 cDNA clones was used to screen for genes showing altered expression in rice seedlings treated with the phytohormone abscisic acid (ABA) (Lin *et al.*, 2003). The experiment identified 37 differentially expressed genes of which majority (30) were up-regulated, and sequence analysis showed several that were potentially involved in grain maturation.

Microarrays have been more widely used to profile gene expression (for reviews of the technique see Bouchez and Hofte, 1998; Blohm and Guiseppi-Elie, 2001) and have been applied to the study of grain development to a greater extent than any other mass screening technique. These arrays have been particularly useful in screening changes in gene expression related to disease and abiotic stresses. For example, rice microarrays were used to analyse the expression of genes regulated by the DREB transcription factors in response to drought and salt stresses (Dubouzet *et al.*, 2003). Ozturk *et al.* (2002) also looked at drought and salt stress, using an array of 1,463 cDNA clones from barley. They found 15% of all genes were up- or down-regulated

by drought stress and 5% by salt stress. Genes that showed up-regulation included those encoding ABA and jasmonate response proteins and several late-embryogenesis-abundant (LEA) proteins.

The impact of drought or water deficit on grain development in maize was studied in detail using a 2,500 feature cDNA microarray (Yu and Setter, 2003). The study compared placental tissue with endosperm at nine days after fertilization. In the placenta, 78 genes displayed altered expression in response to water stress and 56 genes changed in the endosperm sample. Only nine genes showing altered expression were common to both placenta and endosperm. Rewatering allowed the discrimination between genes regulated by transient versus continued stress. It was also interesting that the types of genes affected by the stress in the two tissues differed, suggesting different processes were affected by the stress. In the placenta there seemed to be up-regulation of genes related to stress tolerance, such as heat shock proteins and chaperonins, while genes involved in cell growth and division were down-regulated in endosperm, suggesting that the stress led to a delay in endosperm development.

A purely temporal study of gene expression patterns during maize embryogenesis and kernel development was carried out by Lee *et al*. (2002). Using a 1,500 feature glass array, the observed expression profiles of annotated genes were in general agreement with the published data. The major findings were as follows: (i) genes involved in DNA synthesis, protein synthesis and cell division were highly expressed during early embryogenesis; (ii) the most dramatic changes in expression patterns in the embryo were observed between 20 and 25 days after pollination, when reserves such as storage proteins, starch and fatty acids are synthesized and deposited; (iii) the genes involved in glycolysis, and fatty acid and amino acid biosynthesis, were coordinately regulated during embryo development. The identification of co-regulated genes by clustering also allowed the authors to speculate on the function of several unannotated genes; for example, several genes of unknown function displayed expression patterns resembling those of known storage proteins.

Microarrays have also been used to screen for genes showing altered expression in response to mutations. This procedure was first described in Arabidopsis for the mass screening of mutants (Mahalingam and Fedoroff, 2001) but was recently used to study the *xantha* mutants of barley (Zakhrabekova *et al.*, 2002). The barley paper is important as it provides a method for exploiting more than 8,000 mutants that have been collected for barley. The absence of full genome sequences for all cereals except rice has

limited the application of some genomics technologies, so alternatives are particularly valuable.

3.1.1.3. MPSS

An alternative approach to gene profiling, which uses a novel sequencing method to identify gene signatures in a cDNA population, is MPSS (Brenner *et al.*, 2000; http://www.lynxgen.com/). The method generates several million signatures of 17 to 21 bases that can be used to identify the corresponding coding sequence and determine transcript abundance. It offers a greater depth of analysis than many other methods and provides information on absolute amounts of particular transcripts. The method has been applied to transcript profiling in Arabidopsis to examine the expression of TIR-NBS-LRR disease resistance genes (Meyers *et al.*, 2002) and to compare wild type (WT) with an *abscisic acid insensitive 1-1* (*abi1-1*) mutant (Hoth *et al.*, 2002). Data generated in the latter example can be viewed at http://jcs.biologists.org/supplemental.

3.4. Proteome Analysis

The proteome refers to the complement of proteins present in a particular tissue under particular conditions, and there are now several methods for profiling the proteome. The study of the protein composition of cereal grains has been undertaken over an extensive period and has proved an important predictor of grain characteristics, particularly for processing. In wheat, for example, the high molecular weight glutenins are known to be associated with the key dough characteristics of extensibility and strength (see for example Cornish *et al.*, 2001). Protein extracts can be prepared from tissues in a variety of ways depending on the class of proteins to be studied, for example soluble versus membrane-associated. Using two-dimensional (2D) gel electrophoresis based on size and isoelectric points, between 1,000 and 2,000 proteins can be routinely separated (Fig. 3). However, it may be possible to resolve up to 10,000 proteins in some systems. In many cases the protein spots seen after fractionation will represent different isoforms of the same base protein, a single protein with varying degrees of post-translational modification. It is often a complex task to identify individual protein spots, but a range of techniques for this purpose is now available including the following: amino acid composition, isoelectric point, molecular mass, peptide mass spectrometry fingerprinting and NH_2-terminal sequence. This information can then be used to screen protein and nucleotide sequence databases.

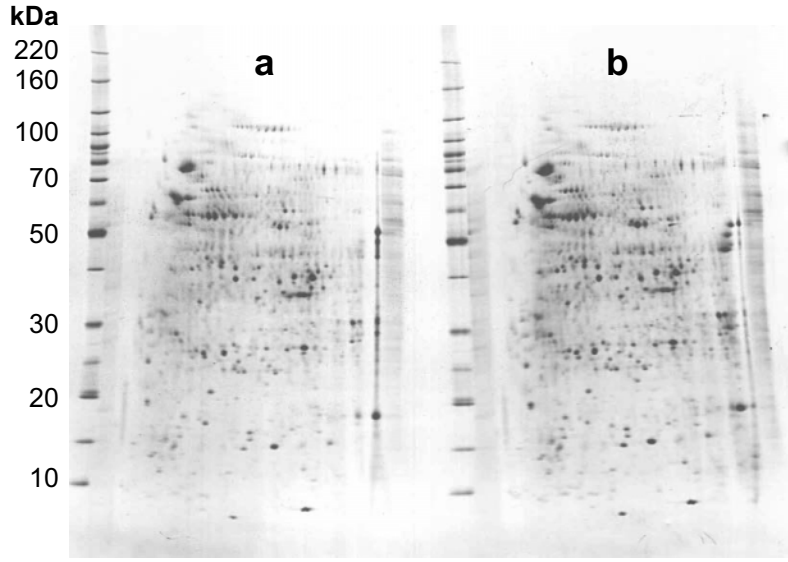


Figure 3. Two-dimensional gel electrophoresis of cytoplasmic proteins from the syncytial endosperm of wheat at 3 (a) and 5 (b) days after pollination. Proteins were identified using peptide mass spectrometry fingerprinting (unpublished data).

Proteomics can be divided into the following three main areas: (i) protein micro-characterization for large-scale identification of proteins and their post-translational modifications; (ii) 'differential display' proteomics for comparison of protein levels (e.g., Figure 3); (iii) studies of protein-protein interactions using techniques such as mass spectrometry or the yeast two-hybrid system. For reviews on proteomics techniques and their applications, see Pandey and Mann (2000) and MacBeath (2002).

Proteome analysis of wheat grain has become an important tool in the assessment of several important quality parameters including protein content, grain hardness, dough strength and extensibility, and starch characteristics (Wrigley *et al*., 2003). In rice, proteomics analysis has revealed 10,589 spots and many could be correlated with specific developmental changes or changes in response to stress (Komatsu *et al.*, 2003). Screening of drought stressed rice leaves showed 42 proteins that changed in abundance and/or position in response to the stress (Salekdeh *et al.*, 2002). Proteomics has also been used to investigate tolerance of maize to *Aspergillus flavus* infection of grain, which can lead to unacceptably high levels of aflatoxins (Chen *et al*., 2002).

Another key activity in proteomics programmes is the investigation of protein-protein interactions that are increasingly recognised as central to cellular function. In work on the early developing wheat grain, we are using the yeast two-hybrid system (see Section 4.3) to progressively build up a picture of multiple interacting proteins that form transcription factor complexes and control the expression of genes that are critical for early endosperm and embryo differentiation and development. This component of proteomics is expected to become increasingly important in the future.

3.5. Metabolite Analysis

Similar to proteome analysis, the metabolites in a particular tissue under defined conditions can be investigated. Metabolite profiles are seen as the ultimate reflection of gene expression at the biochemical level. This implies that metabolites are closer to cellular function than mRNA transcripts or proteins. Proteome analysis will often give an indication of key shifts in important metabolic enzymes that may reflect branch points or regulatory points in biochemical pathways. Through metabolite analysis the flux down a particular pathway can often be predicted. In grain development, activities related to cell division, storage deposition and cell wall synthesis can be critical.

Metabolite profiles are determined from tissue extracts using aqueous or organic solvents. The components can be fractionated by gas or liquid chromatography, and metabolites identified using on-line mass spectrometric analysis and database searching. Standardization and identification of metabolites is often difficult but libraries of mass spectra are now available for many metabolites. Metabolite analysis is becoming increasingly important as information emerges that plant cells have far greater flexibility in biochemical pathways and processes than originally thought.

For studies of grain development, metabolite analysis offers a powerful new tool in understanding development, particularly endosperm development and deposition of storage compounds. Environmental stresses, notably heat, have a major impact on grain filling and composition but the pathways are not well understood. There has also been considerable interest in manipulating grain properties by altering specific components. For example, modifications in amylose/amylopectin ratios impact on many processing characteristics; alterations in protein balance are important in improving the nutritional value of grain; and oil content and composition has become an objective in many maize and oat breeding programmes. Based on experience in other systems, it is probable that these alterations will have more wide ranging effects than

those predicted from the current, largely superficial knowledge of pathways and processes underlying the composition of grains.

4. FUNCTIONAL ANALYSIS

4.1. Computational Assignment of Function

The high quality annotation of the Arabidopsis genome makes it a useful model for assigning function to cereal ESTs on the basis of sequence similarity, as shown by Clarke *et al*. (2003). In a BLASTX comparison of non-redundant ESTs prepared from wheat endosperm with the TIGR Arabidopsis protein data set (http://www.tigr.org/tdb/), 1,456 out of 2,137 ESTs matched 1,130 unique Arabidopsis genes at an e^{-6} cutoff value. Approximately 75% of these Arabidopsis genes had an accompanying annotation, and around 50% had full gene ontology descriptions (Gene Ontology Consortium, http://www.geneontology.org), most being assigned to core cellular machinery. By comparing translated sequences the taxonomic bridge between Arabidopsis and cereals was partially overcome, but comparative data between species as divergent as monocots and dicots requires careful interpretation. Many genes regulating grain development in the cereals will be cereal-specific, and for such genes computational methods can provide, at best, only an approximation of function.

4.2. Reverse Genetics / Transformation

While bioinformatic analysis can provide clues as to the function of a particular gene, experimental approaches must be used to verify this. The most useful approach is to assess loss-of-function and gain-of-function phenotypes through reverse genetics, based on genome-wide mutagenesis or transgenic approaches to target specific loci. Transgenesis also provides the best route to study promoter activity, providing information on the temporal and spatial regulation of gene expression.

Cereal grain development is perhaps one of the processes that has been most manipulated by transformation, because of its importance agronomically and economically. Many of these studies have focused on grain filling in wheat, and, reflecting the position of wheat as the most important food crop in the world, it is the end-use functionality rather than biological functionality which comes under the spotlight. In this respect, transgenic studies in wheat seeds have differed markedly from studies in models like Arabidopsis, which

have been more useful in elucidating the *in planta* biological function of genes.

4.2.1. Wheat HMW-GS Genes and End-Use Functionality

In wheat, the prolamin storage proteins are the major determinants of end-use quality, contributing to both nutritional quality of the grain and to the functionality of dough following processing. These proteins, comprising the monomeric gliadins and the polymeric glutenins, are deposited into protein bodies in the starchy endosperm, beginning at 6-10 days post-anthesis (Shewry *et al*., 2001). Genetic studies first demonstrated that the high molecular weight glutenin subunit (HMW-GS) genes have a quantitative effect on grain quality (Halford *et al*., 1992) and that allelic variants can modify the viscoelasticity of the gluten fraction (Payne, 1987). Since these early findings, transformation has been an indispensable tool in further characterizing the contribution of individual HMW-GS genes to dough quality. Barro *et al*. (1997) introduced two novel HMW-GS genes into a bread wheat background, and were able to show an increase in dough elasticity in T_2 seeds expressing either one or two of the additional subunits. Similarly, overexpression of the HMW subunit 1Dx5 caused a dramatic increase in dough strength, to the extent that the dough was too strong for breadmaking (Rooke *et al*., 1999). The quality of durum wheat can also be altered in this way, as has been shown by overexpression of the subunits 1Ax1 or 1Dx5 (He *et al*., 1999). Expression of the subunit 1Dx5 in tritordeum, a fertile amphiploid between durum wheat and wild barley, conferred breadmaking qualities equivalent to bread wheat even in this poor quality background, while there were only minor changes in dough characteristics effected by co-expression of 1Dx5 and 1Ax1 (Barro *et al*., 2003). Further characterization of the lines generated by Barro *et al*. (1997) has shown that overexpression of 1Dx5 has a much greater effect on glutenin aggregation, and consequently on dough strength, than overexpression of 1Ax1, and this has been attributed to an extra cysteine residue in 1Dx5 that increases intermolecular disulfide cross-links in the gluten (Popineau *et al*., 2001).

4.2.2. Targeted Genetic Studies of Grain Biology

With respect to seed biology in the cereals, most targeted studies have been conducted in rice, using overexpression strategies or the suppression of endogenous genes by antisense technology. For example, the expression of

both sense and antisense constructs of a gibberellin 20-oxidase gene (*rga5*) gave dramatic phenotypes in transgenic rice. Sense expression gave taller plants with longer, more seed-bearing plants while antisense resulted in dwarfing, early flowering and shorter spikes. The up- and down-regulation respectively of *rga5* were confirmed by northern blotting, and were correlated with endogenous GA(1) levels in the plant (Yan *et al*., 2003b). Antisense constructs were also used to suppress the expression of a rice sucrose transporter gene, *OsSUT1*, which normally shows high expression in developing grain (Scofield *et al*., 2002). The resultant phenotype, with reduced grain filling, germination rates and growth of progeny, and with no effect on the source supply of photosynthate, provided compelling evidence that this particular sucrose transporter is involved in transporting sucrose resulting from starch breakdown in germinating seeds.

Recently, several reports have described the use of double-stranded RNA interference (RNAi) technology to investigate gene function in cereal seed biology. The first such studies used transient transformation with RNAi constructs to study the regulation of starch breakdown in barley grain by α-amylase activity. Zentella *et al*. (2002) demonstrated that transient silencing in aleurone cells of GAMYB, a transcription factor, reduced the GA induction of α-amylase, while silencing of SLN1 led to derepression of α-amylase gene expression. Suppression of the transcription factor HvAB15 in a similar fashion had no effect on the hormone regulation of α-amylase gene expression (Casaretto and Ho, 2003).

More recently, RNAi (RNA-interference) technology has been exploited in stable transformation strategies to manipulate cereal seed development. Blocking the mRNA of a gene gives an indication of what that gene does. This technique is achieved by introducing double stranded RNA (dsRNA) matching the gene sequence into cells. The introduction of dsRNA triggers specific RNA degradation. RNAi can be delivered by transient methods (such as particle bombardment or *Agrobacterium* infiltration) or by stable ones (such as the introduction of amplicon transgenes). Segal *et al*. (2003) transformed maize with a double stranded RNAi construct derived from a 22 kDa α-zein storage protein gene. Endosperm-specific expression was achieved by using the 27 kDa γ-zein promoter or the promoter from the *zp22/6* zein gene. The resulting phenotype eliminated 22 kDa zeins without affecting the accumulation of other zein proteins, giving a dominant phenotype with increased lysine content similar to the recessive *opaque* mutations.

The above studies demonstrate that RNAi technology, in transient or stable transformation strategies, is an effective tool for elucidating gene function in

grain development, and it is expected that this technique for generating loss-of-function phenotypes will become more prominent in the future.

4.2.3. Emerging Technologies in Reverse Genetics

Insertional mutagenesis using T-DNA or transposable elements can provide large catalogues of knockout mutants for functional studies, but the technology has been applied systematically in the monocots only to rice and maize (e.g. Rice *Tos17* Insertion Mutant Database, http://tos.nias.affrc.go.jp/~miyao/pub/tos17/; Maize Targeted Mutagenesis Database, http://mtm.cshl.org/). Single-nucleotide mutations induced by traditional mutagens may be more broadly applicable in monocot functional genomics (Henikoff and Comai, 2003), and there now exist robust, high-throughput screening procedures for targeting point mutations, such as the procedure of Targeting Induced Local Lesions in Genomes (TILLING) (McCallum *et al*., 2000). As well as gene knockouts, TILLING can provide allelic series of mutations; this will be especially useful for analysing genes involved in cereal embryogenesis, where knockout mutations often cause lethality and preclude detailed phenotypic analysis.

One alluring prospect for the future is the use of transformation to generate targeted gene knockouts by homologous recombination. The feasibility of this approach has been demonstrated first in the moss *Physcomitrella patens* (Schaefer and Zrÿd, 1997) and more recently in rice (Terada *et al*., 2002). If increases in the efficiency of homologous recombination can be achieved, then this should greatly expedite the investigations on gene function, at least in rice, by reverse genetics.

4.3. Protein Interactions

4.3.1. Protein-Protein Interactions (Known Proteins)

A major step in assigning function to a gene product is to identify protein complexes, pathways and networks. The most widely used technology for this purpose is the yeast two-hybrid system. Bait and prey proteins, fused to the DNA-binding and activation domains, respectively, of a transcription factor, are co-expressed in yeast cells following co-transformation or transformation and mating (Fields and Song, 1989). Recombinant proteins interacting in yeast *in vivo* reconstitute a functional transcription factor, affecting expression of one or more reporter genes. Detection of an interaction is achieved usually by complementation of auxotrophic mutations

and/or by histochemical staining. There are several examples of successful application of this method for the investigation of protein-protein interactions in seeds, providing data which complement genetic, biochemical and biophysical analyses of protein interactions, structure and folding.

4.3.1.1. Validation of interactions between known proteins

Kim *et al*. (2002) used the two-hybrid assay to test directly the interactions between the α-, β-, γ- and δ-zeins in maize endosperm. The differential affinities shown between different zeins, in particular the strong interaction between the 16 kDa γ-zein and the 15 kDa β-zein, suggested that protein body assembly is determined largely by the zein proteins themselves, and not by localization of zein mRNAs which are distributed randomly in the rough endoplasmic reticulum. The yeast two-hybrid assay further identified domains within the 22 kDa α-zein that bind preferentially to zeins of different class, demonstrating the utility of the assay in mapping protein-protein interactions.

Multimeric enzymes are also a suitable target for yeast two-hybrid studies. A direct interaction was shown in yeast between the SH2 and BT2 subunits of the maize tetrameric enzyme ADP-glucose pyrophosphorylase (Greene and Hannah, 1998). Further evidence for their interaction *in planta* was provided by the fact that BT2 subunits remain monomeric in the developing endosperm of *sh2* mutants. Deletions allowed the mapping of motifs required for the interaction (Greene and Hannah, 1998), similarly to the zein study.

The yeast two-hybrid assay, in the context of investigating grain biology, has perhaps been most widely used in studying developmental regulation. In particular, perhaps because the assay can detect interactions in the nucleus only, the technique has proved suitable for identifying the partners of individual transcription factors. For example, the barley endosperm-specific bZIP transcription factor BLZ2 was shown to interact in yeast with another bZIP transcription factor, BLZ1. The latter is known to switch on the expression of genes for seed storage proteins, and the two-hybrid data suggest that it may in fact be a BLZ1/BLZ2 heterodimer that mediates hordein gene transcriptional activation (Oñate *et al*., 1999). Similarly, HvGAMYB, a MYB-type transcription factor from barley that accumulates in the aleurone, starchy endosperm, nucellar projection and embryo, was shown to interact with the C-terminal domain of BPBF (prolamine-box binding factor), a DOF transcriptional activator of the *Hor2* gene (Diaz *et al*., 2002).

4.3.1.2. Screening for novel protein-protein interactions

In addition to assaying potential interactions between two known proteins, the yeast two-hybrid system can be used to identify unknown protein partners.

The gene of interest, in a bait construct, is transformed into yeast cells, which are then mated with cells expressing a library of prey proteins, or a co-transformation is performed. Such a screen was used to identify three interacting partners of the wild oat homologue of the maize VIVIPAROUS1 transcription factor, which has been implicated in the suppression of pre-harvest sprouting (Jones *et al*., 2000). A more systematic approach is facilitated by performing the screen in colony array format, as has been used to study the entire proteome of an organism, e.g., yeast (Uetz *et al*., 2000; Ito *et al*., 2001), or a subset of proteins, as for the proteins involved in vulval development in *Caenorhabditis elegans* (Walhout *et al*., 2000). No systematic approach of this nature has yet been undertaken to study the protein complement of seed protein complexes, the bottleneck being the availability and quality of the prey cDNA libraries. Comprehensive description of protein-protein interactions in the developing grain would contribute considerably to the functional interpretation of data obtained by other methods used in functional genomics.

4.3.2. Protein-DNA Interactions

The yeast one-hybrid system is a method for identifying DNA-binding proteins that recognise a particular stretch of DNA (e.g., a specific binding site on a promoter). The bait in this case is a DNA sequence that is placed upstream of a yeast reporter gene. Library-encoded proteins that bind the target DNA result in activation of the reporter.

There are many examples of application of this method to find specific plant transcriptional factors as well as seed-specific factors. One such protein, the maize endosperm motif binding protein (MEM1), expressed in mid- to late-term endosperm cells, was shown to be capable of activating transcription in the yeast one-hybrid system by binding to the endosperm motif (Heyl *et al*., 2001). Use of the DRE (drought responsive element) *cis*-element of the maize gene *rab17*, expressed during late embryogenesis, as bait in the one-hybrid system identified two novel DRE binding proteins in a maize cDNA library. These proteins, termed DBF1 and DBF2, were then shown to have synergistic effects on *rab17* expression in response to ABA treatment, activating and repressing promoter activity respectively (Kizis and Pagès, 2002). DBF1 and DBF2 were identified as AP2/EREBP transcription factors.

The above studies demonstrate the effectiveness of the one-hybrid system. There are however some limitations. Only relatively short DNA sequences can be used in the screen and long and laborious approaches, using plant

transformation techniques, are needed for the preliminary characterization of specific binding sites.

5. EXAMPLES OF GENOMICS APPLICATION

The importance of cereal grains is highlighted by the huge research effort that has been undertaken to understand and manipulate grain characteristics. Consequently, there are many examples of how genomics and related technologies can and have been used in the study of grain development. The following examples illustrate the ways in which some of the techniques outlined above have been used to address practical problems in plant improvement.

5.1. Apomixis

In apomixis, the embryo develops without the need for the paternal genome normally provided via the pollen. The result of apomixis is that the embryo is a perfect clone, genetically identical to the mother. Although widespread in plants, apomixis does not occur in our cereal crops but would offer large benefits to agriculture by allowing fixation of heterosis. Therefore, growers would have access to the benefits of hybrid vigour without the need for the complex crossing strategies currently needed. This would be particularly important in developing countries where farmers cannot gain the benefits of hybrid seed technologies (Hoisington *et al.*, 1999). A number of different mechanisms or pathways for apomixis have been identified and were recently reviewed (Richards, 2003; Spielman *et al*., 2003).

5.1.1. Strategies for the Development of Apomictic Cereals

In the grasses, several species are known that show apomictic reproduction and these include some close relatives of crop plants. There are also a number of mutations or reproductive variants of cereals that may be valuable in the development of apomixis systems. Spielman *et al.* (2003) describe four approaches taken to develop apomictic crop plants: (i) transfer of apomixis genes or alleles from apomictic wild relatives to crop plants; (ii) identification of mutations that show fertilization-independent seed development; (iii) identification, cloning and transformation into crop species of genes or alleles from more distantly related apomictic plants; (iv) engineering of a novel

apomictic system through a detailed understanding of grain development and the genes involved in the control of grain development.

In cereals, most effort has focused on the first approach and this has led to the extensive screening and analysis of apomixis in wild relatives of the cereals. A first stage in these approaches is to understand the genetics of apomixis in grasses and to develop strategies for transferring the associated gene or alleles. Several AFLP markers have been identified showing linkage to an apomixis locus in *Paspalum simplex* and these have been used to identify the syntenous region in the telomeric region of chromosome 12 of rice (Labombarda *et al.*, 2002; Pupilli *et al*., 2001). Unfortunately, the region appears to be one of recombination suppression and this will make map-based cloning difficult.

Brachiaria has also been studied as a possible source of apomixis genes or alleles. A cross between *B. brizantha* and *B. ruziziensis* allowed the mapping of an "apomixis gene" and the syntenic regions have been identified in maize (chromosome 5) and rice (chromosome 2) (Pessino *et al*., 1998). This appears to be an easier region to work with than that identified in *Paspalum*.

5.1.2. Transfer of Apomixis to Maize

Research into the mapping and transfer of apomixis to maize from *Tripsacum* has been underway for several years. In this case, apomixis appears to be complex due to a relationship between the apomixis locus or loci and genes controlling other aspects of fertility (Blakey *et al*., 2001). The polyploid nature of *Tripsacum* appears to be important in the apomixis process and this has presented added difficulties in the genetic analysis of apomixis and has led to the suggestion that gene dosage may be important (Grimanelli *et al.*, 1998a). However, hybrids between maize and *Tripsacum* showing apomixis have been generated and studied (Kindiger *et al.*, 1996). The results are rather confusing and indicate that transmission does not follow the expected Mendelian segregation (Grimanelli *et al.*, 1998b).

5.1.3. Transfer of Apomixis to Triticeae Cereals

Tripsacum appears to be a good model for apomixis in maize but there are also several good models for the transfer of apomixis to members of the Triticeae (wheat, barley and rye), including *Elymus* species (Murphy and Jones, 1999). Hybrids produced between three species of *Elymus* and wheat,

rye and barley showed many features of apomictic grain development but did not result in transfer of an active apomixis system (Torabinejad and Mueller, 1993). It would appear that transfer of apomixis to wheat may require transfer of a genome segment (Peel *et al.*, 1997). *Thinopyrum distichum* also offers some potential, since hybrids with cereal rye (*Secale cereale*) show some level of apomixis (Marais *et al.*, 1998). A further possibility for developing apomixis in wheat is referred to as the "Salmon System" (Matzk, 1996). This system allows autonomous embryo development and is derived from an initial cross with an *Aegilops* relative of wheat.

5.1.4. Transfer of Apomixis to Rice and Pearl-Millet

Apomixis has been successfully transferred to pearl millet (*Pennisetum glaucum*) from *P. squamulatum* in a BC3 population. However, the lines showed low fertility and they have not progressed to more extensive evaluation (Morgan *et al.*, 1998).

Unfortunately, there are no well described apomictic systems in close relatives of rice although there have been two reports from China on possible apomictic systems (Wu *et al.,* 1991; Zhou *et al*., 1993).

5.1.5. Present Status of Apomixis in Cereals and Future Prospects

The general conclusion of work on apomixis in grasses is that the genetics is more complex than originally thought. Simple transfer via wide crosses has failed to yield useful apomictic cereals. More detailed understanding of the genetic control and molecular biology of apomixis is now seen as the best way forward. Therefore, research has shifted to the analysis of early developmental processes in the cereals with the aim to identify the genes involved in controlling key decisions in developmental changes for both the embryo and endosperm. These studies have been hampered in the past by the small size and relative inaccessibility of early grain tissues, but several technologies promise to overcome these obstacles. The initial events of embryogenesis and endosperm development can now be followed *in vitro*, precisely timed after fertilization (Kranz and Kumlehn, 1999), while global expression analysis in specific cell types is possible using micromanipulation or laser capture in combination with RNA amplification procedures (Kerk *et al*., 2003; Nakazono *et al*., 2003). It is hoped that these studies will help unravel the genetic control of early development and allow the final approach listed above, engineering novel apomictic pathways.

5.2. β-Amylase in Barley

The mobilization of starch reserves in germinating grain is a key determinant of malting quality in barley. This process involves the concerted action of a wide range of enzymes including enzymes required for cell wall degradation, to allow access to the starch granules, and enzymes of starch catabolism. Variation in β-amylase alleles had been known for some time and there was a degree of association between isoforms of β-amylase and malting quality (Allison, 1973). This association was not well understood and was seldom used in breeding programmes as a major criterion for selection. However, with the advent of molecular markers and the associated mapping populations, QTLs for malt quality were placed on many maps (http://wheat.pw.usda.gov/ggpages/maps.shtml#barley).The β-amylase region was consistently shown to be one of several loci related to diastatic power, a key measure of malting quality. This information was used to investigate alternate forms of β-amylase, in particular variants that showed increased thermostability and substrate affinity. Allele discovery programmes revealed several novel alleles of β-amylase in wild barley, *Hordeum spontaneum* (Eglinton *et al*., 1998). The information and sequence data on alleles of β-amylase and the availability of a crystal structure of the enzyme (Mikami *et al*., 1999) allowed the systematic characterization of single amino acid changes to identify the major structural features and to design a variant of the enzyme that combined the most desirable thermostability and substrate affinity properties of the enzyme (Ma *et al*., 2001).

5.3. Pre-Harvest Sprouting in Wheat

Cultivated wheat is susceptible to pre-harvest sprouting (PHS), where the grain germinates while still in the spike, with consequent penalties on grain yield and quality. A full understanding, therefore, of how genetic mechanisms control seed dormancy and the transition to germination would present a strategy for improving wheat with respect to this undesirable trait.

A phenotype similar to PHS exists in the maize *vp1* mutant (McCarty *et al*., 1989) and the Arabidopsis *abi3* mutant (Nambara *et al*., 1992; Nambara *et al*., 1995). The *Vp1* and *Abi3* genes have been cloned (McCarty *et al*., 1991; Giraudat *et al*., 1992) and are believed to be orthologues, based on amino acid sequence conservation and the fact that expression of the *Vp1* gene in Arabidopsis can partially complement the mutant *abi3* phenotype (Suzuki *et al*., 2001). Biochemical studies have shown that *Vp1* is a transcription factor, activating expression of maturation-related genes (McCarty *et al*., 1991)

while repressing the transcription of α-amylase (Hoecker *et al*., 1999). Taking a candidate gene approach, the orthologue of *Vp1*/*Abi3* in wheat (*TaVp1*) has been cloned and shown to map to a chromosomal region in wheat and rice syntenic to the *Vp1* locus in maize (Bailey *et al*., 1999). Three wheat genomic clones, corresponding to a single copy of *TaVp1* on each of the A, B and D genomes, appeared to code full-length transcripts, but in fact the transcripts were found to be highly misspliced (McKibbin *et al*., 2002). It seems likely, therefore, that the resulting loss of TaVP1 activity is responsible for the susceptibility of cultivated wheat to PHS. The finding that *TaVp1* transcript is similarly misspliced in ancestral diploid wheats precludes the use of breeding strategies to introduce a functional copy of *Vp1* into modern bread wheat. Instead, resistance to PHS may be conferred using a transgenic approach, as has been shown by expression of the wild oat (*Avena fatua*) *Vp1* orthologue in transformed wheat (McKibbin *et al*., 2002).

6. SUMMARY AND OUTLOOK

In the past, the success of plant breeding in increasing cereal grain yields has been largely through the adoption of new varieties and changes in agronomic practices. Although the characteristics of cereal grains have changed over the past few decades of breeding and selection, there is still considerable scope for further improvement. For wheat, in particular, we still have only a poor understanding of the molecular basis for many of the quality characteristics that are so critical for the diverse end-uses of this cereal. Molecular markers have been important in defining many of the loci controlling quality but genomics approaches may help define the key criteria and the actual genes.

A detailed understanding of the genetic control of grain quality and related characteristics also opens the path to exploring the wild relatives of our major cereals more comprehensively than has been possible in the past. An example was presented above where the knowledge of the role of β-amylase in determining malting quality allowed an exploration of wild barleys as a source of novel alleles. This led to the identification of more desirable alleles and also a better understanding of the structure and function of the protein. Can we use genomics to further expand the germplasm base for cereals?

The structure of the grain also presents opportunities for improvement of grain quality, if we can define the molecular events involved in grain development. Grain shape is an important aspect of quality for all major cereals. The balance and composition of endosperm, embryo, aleurone and other tissues also impacts on the processing and nutritional quality characteristics of grains. Can these attributes be changed? For example, there

exist large differences among cereal species in the overall composition of the grain. High oil lines of oat and maize are now available but there has been little progress in raising the oil content of rice or wheat. These features vary between cereals and varieties grown under optimal conditions, but we know that stresses, both biotic and abiotic, can have profound effects on the characteristics of grains. Again, genomics approaches offer a means for analysing the molecular events that occur when the plants are exposed to these stresses and may provide genetic or molecular genetic solutions. A further major target is the development of cereal grains for novel end-uses. The scope here is also large through the production of novel products, such as pharmaceuticals, or through the modification of existing components or shifting the balance of compounds. Altered starch composition has been a major target but alterations in cell wall constituents also offer new opportunities.

In addition to the features of the grain that affect nutritional and processing quality, grain characteristics also have a large impact on the early development of the seedling. Even though wheat and barley are very closely related, barley seedlings show greater stress tolerance than wheat. This tolerance is at least partially due to the characteristics of the barley seedling such as long coleoptile, broad leaves and rapid germination. These features are all related to embryo characteristics and embryo size. Alterations in the early architecture of the plant may provide a mechanism for addressing several current production problems such as competition with weeds and more efficient use of surface water.

The progress in developing apomictic cereals has been discussed above. The initial proposal that this could be achieved by crossing the appropriate genes into the cereals from apomictic wild relatives has proved elusive and most research has returned to an analysis of early embryo and endosperm development.

In summary, the opportunities for manipulating grain yield and quality and enhancing early seedling vigour are dependent on a detailed understanding of grain development. Genomics and the related technologies of proteomics and metabolomics provide the tools for elucidating the molecular events leading to grain formation.

7. ACKNOWLEDGEMENTS

The Australian Centre for Plant Functional Genomics receives funding from the Australian Research Council, the Grains Research and Development

Corporation and the South Australian Government. The authors would like to thank Jason Able for critical appraisal of the manuscript, and Natalia Bazanova for providing artwork.

8. REFERENCES

Ahn S., Anderson J.A., Sorrells M.E., Tanksley S.D. Homoeologous relationships of rice, wheat and maize chromosomes. Mol Gen Genet 1993; 241: 483-490

Allison M.J. Genetic studies of the β-amylase isozymes of malt barley. Genetica 1973; 44: 1-15

Bailey P.C., McKibbin R.S., Lenton J.R., Holdsworth M.J., Flintham J.E., Gale M.D. Genetic map locations for orthologous *Vp1* genes in wheat and rice. Theor Appl Genet 1999; 98: 281-284

Barro F., Barceló P., Lazzeri P.A., Shewry P.R., Martín A., Ballesteros J. Functional properties and agronomic performance of transgenic tritordeum expressing high molecular weight glutenin subunit genes 1Ax1 and 1Dx5. Jour Cereal Sci 2003; 37: 65-70

Barro F., Rooke L., Békés F., Gras P., Tatham A.S., Fido R., Lazzeri P.A., Shewry P.R., Barceló P. Transformation of wheat with high molecular weight subunit genes results in improved functional properties. Nature Biotech 1997; 15: 1295-1299

Becraft P.W. Cell fate specification in the cereal endosperm. Semin Cell Dev Biol 2001; 12: 387-394

Becraft P.W., Li K., Dey N., Asuncion-Crabb Y. The maize *dek1* gene functions in embryonic pattern formation and cell fate specification. Development 2002; 129: 5217-5225

Blakey C.A., Goldman S.L., Dewald C.L. Apomixis in *Tripsacum*: comparative mapping of a multigene phenomenon. Genome 2001; 44: 222-230

Blohm D.H., Guiseppi-Elie A. New developments in microarray technology. Curr Opin Biotech 2001; 12: 41-47

Bouchez D., Hofte H. Functional genomics in plants. Plant Physiol 1998; 118: 725-732

Brenner S., Johnson M., Bridgham J., Golda G., Lloyd D.H., Johnson D., Luo S., McCurdy S., Foy M., Ewan M., *et al*. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. Nature Biotech 2000; 18: 630-4

Casaretto J., Ho T.-H.D. The transcription factors HvABI5 and HvVP1 are required for the abscisic acid induction of gene expression in barley aleurone cells. Plant Cell 2003; 15: 271-284

Chen Z.Y., Brown R.L., Damann K.E., Cleveland T.E. Identification of unique or elevated levels of kernel proteins in aflatoxin-resistant maize genotypes through proteome analysis. Phytopathology 2002; 92: 1084-1094

Clarke B., Lambrecht M., Rhee S.Y. Arabidopsis genomic information for interpreting wheat EST sequences. Funct Integr Genom 2003; 3: 33-38

Clements R.J., Rosielle A.A., Hilton R.D. *National review of crop improvement in the Australian grains industry*. A report to the Board of the Grains Research and Development Corporation (GRDC), Canberra, Australia, 1992

Collins N.C., Thordal-Christensen H., Lipka V., Bau S., Kombrink E., Qiu J.-L., Hückelhoven R., Stein M., Freialdenhoven A., Somerville S.C., Schulze-Lefert P. SNARE-protein-mediated disease resistance at the plant cell wall. Nature 2003; 425: 973-977

Colot V., Bartels D., Thompson R., Flavell R. Molecular characterization of an active wheat LMW glutenin gene and its relation to other wheat and barley prolamin genes. Mol Gen Genet 1989; 216: 81-90

Cornish G.B., Skylas D.J., Siriamornpun S., Békés F., Larroque O.R., Wrigley C.W., Wootton M. Grain proteins as markers of genetic traits in wheat. Aust Jour Agric Res 2001; 52: 1161-1171

Devos K.M., Gale M.D. Genome relationships: the grass model in current research. Plant Cell 2000; 12: 637-646

Diaz I., Vicente-Carbajosa J., Abraham Z., Martínez M., Isabel-La Moneda I., Carbonero P. The GAMYB protein from barley interacts with the DOF transcription factor BPBF and activates endosperm-specific genes during seed development. Plant Jour 2002; 29: 453-464

Domínguez F., Moreno J., Cejudo F.J. The nucellus degenerates by a process of programmed cell death during the early stages of wheat grain development. Planta 2001; 213: 352-360

Domon E., Saito A., Takeda K. Comparison of the waxy locus sequence from a non-waxy strain and two waxy mutants of spontaneous and artificial origins in barley. Genes Genet Syst 2002; 77: 351-359

Dubouzet J.G., Sakuma Y., Ito Y., Kasuga M., Dubouzet E.G., Miura S., Seki M., Shinozaki K., Yamaguchi-Shinozaki K. *OsDREB* genes in rice, *Oryza sativa* L., encode transcription activators that function in drought-, high-salt- and cold-responsive gene expression. Plant Jour 2003; 33: 751-763

Eglinton J.K., Langridge P., Evans D.E. Thermostability variation in alleles of barley *beta*-amylase. Jour Cereal Sci 1998; 28: 301-309

Elster R., Bommert P., Sheridan W.F., Werr W. Analysis of four embryo-specific mutants in *Zea mays* reveals that incomplete radial organization of the proembryo interferes with subsequent development. Dev Genes Evol 2000; 210: 300-310

Ertl D.S., Young K.A., Raboy V. Plant genetic approaches to phosphorus management in agricultural production. Jour Environ Qual 1998; 27: 299-304

Evers T., Millar S. Cereal grain structure and development: Some implications for quality. Jour Cereal Sci 2002; 36: 261-284

Fields S., Song O. A novel genetic system to detect protein-protein interactions. Nature 1989; 340: 245-246

Gautier M.-F., Aleman M.-E., Guirao A., Marion D., Joudrier P. *Triticum aestivum* puroindolines, two basic cystine-rich seed proteins: cDNA sequence analysis and developmental gene expression. Plant Mol Biol 1994; 25: 43-57

Giraudat J., Hauge B.M., Valon C., Smalle J., Parcy F., Goodman H.M. Isolation of the Arabidopsis-*Abi3* gene by positional cloning. Plant Cell 1992; 4: 1251-1261

Goff S.A., Ricke D., Lan T.-H., Presting G., Wang R., Dunn M., Glazebrook J., Sessions A., Oeller P., Varma H., *et al*. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science 2002; 296: 92-100

Goldberg R.B., de Paiva G., Yadegari R. Plant embryogenesis: zygote to seed. Science 1994; 266: 605-614

Greene T.W., Hannah L.C. Maize endosperm ADP-glucose pyrophosphorylase SHRUNKEN2 and BRITTLE2 subunit interactions. Plant Cell 1998; 10: 1295-1306

Grimanelli D., Leblanc O., Espinosa E., Perotti E., De Leon D.G., Savidan Y. Mapping diplosporous apomixis in tetraploid *Tripsacum*: one gene or several genes? Heredity 1998a; 80: 33-39

Grimanelli D., Leblanc O., Espinosa E., Perotti E., De Leon D.G., Savidan Y. Non-Mendelian transmission of apomixis in maize-*Tripsacum* hybrids caused by a transmission ratio distortion. Heredity 1998b; 80: 40-47

Groos C., Gay G., Perretant M.-R., Gervais L., Bernard M., Dedryver F., Charmet D. Study of the relationship between pre-harvest sprouting and grain color by quantitative trait loci analysis in a white x red grain bread-wheat cross. Theor Appl Genet 2002; 104: 39-47

Halford N.G., Field J.M., Blair H., Urwin P., Moore K., Robert L., Thompson R., Flavell R.B., Tatham A.S., Shewry P.R. Analysis of HMW glutenin subunits encoded by

chromosome 1A of bread wheat (*Triticum aestivum* L.) indicates quantitative effects on grain quality. Theor Appl Genet 1992; 83: 373-378

He G.Y., Rooke L., Steele S., Békés F., Gras P., Tatham A.S., Fido R., Barcelo P., Shewry P.R., Lazzeri P.A. Transformation of pasta wheat (*Triticum turgidum* L. var. *durum*) with high-molecular-weight glutenin subunit genes and modification of dough functionality. Mol Breed 1999; 5: 377-386

Heisey P.W., Srinivasan, C.S., Thirtle, C. *Public Sector Plant Breeding in a Privatizing World*. Agriculture Information Bulletin No 772; Resource Economics Division, Economic Research Service, U.S. Department of Agriculture, USA, 2001

Heyl A., Muth J., Santandrea G., O'Connell T., Serna A., Thompson R.D. A transcript encoding a nucleic acid-binding protein specifically expressed in maize seeds. Mol Genet Genom 2001; 266: 180-189

Hoecker U., Vasil I.K., McCarty D.R. Signaling from the embryo conditions *Vp1*-mediated repression of α-amylase genes in the aleurone of developing maize seeds. Plant Jour 1999; 19: 371-377

Hoisington D., Khairallah M., Reeves T., Ribaut J.V., Skovmand B., Taba S., Warburton M. Plant genetic resources: what can they contribute toward increased crop productivity? Proc Nat Acad Sci USA 1999; 96: 5937-5943

Hoth S., Morgante M., Sanchez J.-P., Hanafey M.K., Tingey S.V., Chua N.-H. Genome-wide gene expression profiling in *Arabidopsis thaliana* reveals new targets of abscisic acid and largely impaired gene regulation in the *abi1-1* mutant. Jour Cell Sci 2002; 115: 4891-4900

Ikeda A., Ueguchi-Tanaka M., Sonoda Y., Kitano H., Koshioka M., Futsuhara Y., Matsuoka M., Yamaguchi J. *slender* rice, a constitutive gibberellin response mutant, is caused by a null mutation of the *SLR1* gene, an ortholog of the height-regulating gene *GAI*/*RGA*/*RHT*/*D8*. Plant Cell 2001; 13: 999-1010

Ito T., Chiba T., Ozawa R., Yoshida M., Hattori M., Sakaki Y. A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc Nat Acad Sci USA 2001; 98: 4569-4574

Jones H.D., Kurup S., Peters N.C., Holdsworth M.J. Identification and analysis of proteins that interact with the *Avena fatua* homologue of the maize transcription factor VIVIPAROUS 1. Plant Jour 2000; 21: 133-142

Kent N.L., Evers A.D. *Kent's Technology of Cereals*, Pergamon Press Ltd., Oxford, UK, 1994.

Kerk N.M., Ceserani T., Tausta S.L., Sussex I.M., Nelson T.M. Laser capture microdissection of cells from plant tissues. Plant Physiology 2003; 132: 27-35

Kim C.S., Woo Y.-M., Clore A.M., Burnett R.J., Carneiro N.P., Larkins B.A. Zein protein interactions, rather than the asymmetric distribution of zein mRNAs on endoplasmic reticulum membranes, influence protein body formation in maize endosperm. Plant Cell 2002; 14: 655-672

Kindiger B., Bai D., Sokolov V. Assignment of a gene(s) conferring apomixis in *Tripsacum* to a chromosome arm: cytological and molecular evidence. Genome 1996; 39: 1133-1141

Kizis D., Pagès M. Maize DRE-binding proteins DBF1 and DBF2 are involved in *rab17* regulation through the drought-responsive element in an ABA-dependent pathway. Plant Jour 2002; 30: 679-689

Klösgen R.B., Gierl A., Schwarz-Sommer Z., Saedler H. Molecular analysis of the *waxy* locus of *Zea mays*. Mol Gen Genet 1986; 203: 237-244

Komatsu S., Konishi H., Shen S., Yang G. Rice proteomics: a step toward functional analysis of the rice genome. Mol Cell Proteom 2003; 2: 2-10

Kranz E., Kumlehn J. Angiosperm fertilisation, embryo and endosperm development in vitro. Plant Science 1999; 142: 183-197

Labombarda P., Busti A., Caceres M.E., Pupilli F., Arcioni S. An AFLP marker tightly linked to apomixis reveals hemizygosity in a portion of the apomixis-controlling locus in *Paspalum simplex*. Genome 2002; 45: 513-519

Langridge P., Lagudah E.S., Holton T.A., Appels R., Sharp P.J., Chalmers K.J. Trends in genetic and genome analyses in wheat: a review. Aust Jour Agric Res 2001; 52: 1043-1077

Larson S.R., Rutger J.N., Young K.A., Raboy V. Isolation and genetic mapping of a non-lethal rice (*Oryza sativa* L.) *low phytic acid 1* mutation. Crop Sci 2000; 40: 1397-1405

Larson S.R., Young K.A., Cook A., Blake T.K., Raboy V. Linkage mapping of two mutations that reduce phytic acid content of barley grain. Theor Appl Genet 1998; 97: 141-146

Lin F., Xu S.L., Ni W.M., Chu Z.Q., Xu Z.H., Xue H.W. Identification of ABA-responsive genes in rice shoots via cDNA macroarray. Cell Res 2003; 13: 59-68

Lohmer S., Maddaloni M., Motto M., Di Fonzo N., Hartings H., Salamini F., Thompson R.D. The maize regulatory locus *Opaque-2* encodes a DNA-binding protein which activates the transcription of the *b-32* gene. EMBO Jour 1991; 10: 617-624

Ma Y.F., Evans D.E., Logue S.J., Langridge P. Mutations of barley β-amylase that improve substrate-binding affinity and thermostability. Mol Genet Genom 2001; 266: 345-352

MacBeath G. Protein microarrays and proteomics. Nature Genet 2002; 32: 526-532

Mahalingam R., Fedoroff N. Screening insertion libraries for mutations in many genes simultaneously using DNA microarrays. Proc Nat Acad Sci USA 2001; 98: 7420-7425

Marais G.F., Marais A.S., Badenhorst P.C. Hybridization of *Thinopyrum distichum* and *Secale cereale*. Plant Breed 1998; 117: 107-111

Mares D.J., Norstog K., Stone B.A. Early stages in the development of wheat endosperm. I. The change from free nuclear to cellular endosperm. Aust Jour Bot 1975; 23: 311-326

Matsumura H., Nirasawa S., Terauchi R. Technical advance: transcript profiling in rice (*Oryza sativa* L.) seedlings using serial analysis of gene expression (SAGE). Plant Jour 1999; 20: 719-726

Matzk F. The 'Salmon system' of wheat - a suitable model for apomixis research. Hereditas 1996; 125: 299-301

McCarty D.R., Carson C.B., Stinard P.S., Robertson D.S. Molecular analysis of *Viviparous-1* - an abscisic acid insensitive mutant of maize. Plant Cell 1989; 1: 523-532

McCarty D.R., Hattori T., Carson C.B., Vasil V., Lazar M., Vasil I.K. The *Viviparous-1* developmental gene of maize encodes a novel transcriptional activator. Cell 1991; 66: 895-905

McKibbin R.S., Wilkinson M.D., Bailey P.C., Flintham J.E., Andrew L.M., Lazzeri P.A., Gale M.D., Lenton J.R., Holdsworth M.J. Transcripts of *Vp-1* homeologues are misspliced in modern wheat and ancestral species. Proc Nat Acad Sci USA 2002; 99: 10203-10208

Meyers B.C., Morgante M., Michelmore R.W. TIR-X and TIR-NBS proteins: two new families related to disease resistance TIR-NBS-LRR proteins encoded in Arabidopsis and other plant genomes. Plant Jour 2002; 32: 77-92

Mikami B., Yoon H.J., Yoshigi N. The crystal structure of the sevenfold mutant of barley β-amylase with increased thermostability at 2.5 Å resolution. Jour Mol Biol 1999; 285: 1235-1243

Moore G., Devos K.M., Wang Z., Gale M.D. Cereal genome evolution - grasses, line up and form a circle. Curr Biol 1995; 5: 737-739

Mordhorst A.P., Toonen M.A.J., de Vries S.C. Plant embryogenesis. Crit Rev Plant Sci 1997; 16: 535-576

Morgan R., Ozias-Akins P., Hanna W.W. Seed set in an apomictic BC3 pearl millet. Int Jour Plant Sci 1998; 159: 89-97

Morris C.F. Puroindolines: the molecular genetic basis of wheat grain hardness. Plant Mol Biol 2002; 48: 633-647

Murai J., Taira T., Ohta D. Isolation and characterization of the three *Waxy* genes encoding the granule-bound starch synthase in hexaploid wheat. Gene 1999; 234: 71-79

Murphy M.A., Jones C.E. Observations on the genus *Elymus* (Poaceae: Triticeae) in Australia. Aust Syst Bot 1999; 12: 593-604

Nakazono M., Qiu F., Borsuk L.A., Schnable P.S. Laser-capture microdissection, a tool for the global analysis of gene expression in specific plant cell types: Identification of genes expressed differentially in epidermal cells or vascular tissues of maize. Plant Cell 2003; 15: 583-596

Nambara E., Keith K., McCourt P., Naito S. A regulatory role for the *Abi3* gene in the establishment of embryo maturation in *Arabidopsis thaliana*. Dev Genet 1995; 121: 629-636

Nambara E., Naito S., McCourt P. A mutant of Arabidopsis which is defective in seed development and storage protein accumulation is a new *Abi3* allele. Plant Jour 1992; 2: 435-441

Ogihara Y., Mochida K., Nemoto Y., Murai K., Yamazaki Y., Shin-I T., Kohara Y. Correlated clustering and virtual display of gene expression patterns in the wheat life cycle by large-scale statistical analyses of expressed sequence tags. Plant Jour 2003; 33: 1001-1011

Olsen O.A. Endosperm development: cellularization and cell fate specification. Ann Rev Plant Physiol Plant Mol Biol 2001; 52: 233-267

Oñate L., Vicente-Carbajosa J., Lara P., Díaz I., Carbonero P. Barley BLZ2, a seed-specific bZIP protein that interacts with BLZ1 *in vivo* and activates transcription from the GCN4-like motif of B-hordein promoters in barley endosperm. Jour Biol Chem 1999; 274: 9175-9182

Ozturk Z.N., Talame V., Deyholos M., Michalowski C.B., Galbraith D.W., Gozukirmizi N., Tuberosa R., Bohnert H.J. Monitoring large-scale changes in transcript abundance in drought- and salt-stressed barley. Plant Mol Biol 2002; 48: 551-573

Pandey A., Mann M. Proteomics to study genes and genomes. Nature 2000; 405: 837-846

Patron N.J., Smith A.M., Fahy B.F., Hylton C.M., Naldrett M.J., Rossnagel B.G., Denyer K. The altered pattern of amylose accumulation in the endosperm of low-amylose barley cultivars is attributable to a single mutant allele of granule-bound starch synthase I with a deletion in the 5 '-non-coding region. Plant Physiol 2002; 130: 190-198

Payne P.I. Genetics of wheat storage proteins and the effect of allelic variation on breadmaking quality. Ann Rev Plant Physiol 1987; 38: 141-153

Peel M.D., Carman J.G., Liu Z.W., Wang R.R.-C. Meiotic anomalies in hybrids between wheat and apomictic *Elymus rectisetus* (Nees in Lehm.) A. Löve & Connor. Crop Sci 1997; 37: 717-723

Pessino S.C., Evans C., Ortiz J.P.A., Armstead I., do Valle C.B., Hayward M.D. A genetic map of the apospory-region in *Brachiaria* hybrids: identification of two markers closely associated with the trait. Hereditas 1998; 128: 153-158

Popineau Y., Deshayes G., Lefebvre J., Fido R., Tatham A.S., Shewry P.R. Prolamin aggregation, gluten viscoelasticity, and mixing properties of transgenic wheat lines expressing 1Ax and 1Dx high molecular weight glutenin subunit transgenes. Jour Agric Food Chem 2001; 49: 395-401

Pupilli F., Labombarda P., Caceres M.E., Quarín C.L., Arcioni S. The chromosome segment related to apomixis in *Paspalum simplex* is homoeologous to the telomeric region of the long arm of rice chromosome 12. Mol Breed 2001; 8: 53-61

Raboy V., Gerbasi P.F., Young K.A., Stoneberg S.D., Pickett S.G., Bauman A.T., Murthy P.P.N., Sheridan W.F., Ertl D.S. Origin and seed phenotype of maize *low phytic acid 1-1* and *low phytic acid 2-1*. Plant Physiol 2000; 124: 355-368

Raghavan V. *Molecular Embryology of Flowering Plants*, Cambridge University Press, Cambridge, UK, 1997

Rahman S., Jolly C.J., Skerritt J.H., Wallosheck A. Cloning of a wheat 15-kDa grain softness protein (GSP) – GSP is a mixture of puroindoline-like polypeptides. Eur Jour Biochem 1994; 223: 917-925

Reddy P., Appels R. Analysis of a genomic DNA segment carrying the wheat high-molecular-weight (HMW) glutenin Bx17 subunit and its use as an RFLP marker. Theor Appl Genet 1993; 85: 616-624

Richards A.J. Apomixis in flowering plants: an overview. Philos T Roy Soc B 2003; 358: 1085-1093

Rohde W., Becker D., Salamini F. Structural analysis of the *waxy* locus from *Hordeum vulgare*. Nucl Acid Res 1988; 16: 7185-7186

Rooke L., Barro F., Tatham A.S., Fido R., Steele S., Békés F., Gras P., Martin A., Lazzeri P.A., Shewry P.R., Barcelo P. Altered functional properties of tritordeum by transformation with HMW glutenin subunit genes. Theor Appl Genet 1999; 99: 851-858

Salekdeh G.H., Siopongco J., Wade L.J., Ghareyazie B., Bennett J. A proteomic approach to analyzing drought- and salt-responsiveness in rice. Field Crop Res 2002; 76: 199-219

Schaefer D.G., Zrÿd J.P. Efficient gene targeting in the moss *Physcomitrella patens*. Plant Jour 1997; 11: 1195-1206

Scofield G.N., Hirose T., Gaudron J.A., Upadhyaya N.M., Ohsugi R., Furbank R.T. Antisense suppression of the rice sucrose transporter gene, *OsSUT1*, leads to impaired grain filling and germination but does not affect photosynthesis. Funct Plant Biol 2002; 29: 815-826

Shewry P.R., Halford N.G. Cereal seed storage proteins: structures, properties and role in grain utilization. Jour Exp Bot 2002; 53: 947-958

Shewry P.R., Tatham A.S., Fido R., Jones H., Barcelo P., Lazzeri P.A. Improving the end use properties of wheat by manipulating the grain protein composition. Euphytica 2001; 119: 45-48

Shi J., Wang H., Wu Y., Hazebroek J., Meeley R.B., Ertl D.S. The maize low-phytic acid mutant *1pa2* is caused by mutation in an inositol phosphate kinase gene. Plant Physiol 2003; 131: 507-515

Smilde W.D., Haluškova J., Sasaki T., Graner A. New evidence for the synteny of rice chromosome 1 and barley chromosome 3H from rice expressed sequence tags. Genome 2001; 44: 361-367

Spielman M., Vinkenoog R., Scott R.J. Genetic mechanisms of apomixis. Philos T Roy Soc B 2003; 358: 1095-1103

Sreenivasulu N., Altschmied L., Panitz R., Hähnel U., Michalek W., Weschke W., Wobus U. Identification of genes specifically expressed in maternal and filial tissues of barley caryopses: a cDNA array analysis. Mol Gen Genom 2002; 266: 758-767

Suzuki M., Kao C.-Y., Cocciolone S., McCarty D.R. Maize VP1 complements Arabidopsis *abi3* and confers a novel ABA/auxin interaction in roots. Plant Jour 2001; 28: 409-418

Swanston J.S., Sopena A., Moralejo M.A., Molina-Cano J.-L. Germination and malting properties of mutants derived from malting barley cv. Triumph. Cereal Chem 2002; 79: 392-396

Terada R., Urawa H., Inagaki Y., Tsugane K., Iida S. Efficient gene targeting by homologous recombination in rice. Nature Biotech 2002; 20: 1030-1034

Thomas S.W., Glaring M.A., Rasmussen S.W., Kinane J.T., Oliver R.P. Transcript profiling in the barley mildew pathogen *Blumeria graminis* by serial analysis of gene expression (SAGE). Mol Plant Microbe Interact 2002; 15: 847-856

Thorne J.H. Phloem unloading of C and N assimilates in developing seeds. Ann Rev Plant Physiol 1985; 36: 317-343

Torabinejad J., Mueller R.J. Genome analysis of intergeneric hybrids of apomictic and sexual Australian *Elymus* species with wheat, barley and rye: implication for the transfer of apomixis to cereals. Theor Appl Genet 1993; 86: 288-294

Uetz P., Giot L., Cagney G., Mansfield T.A., Judson R.S., Knight J.R., Lockshon D., Narayan V., Srinivasan M., Pochart P., *et al.* A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. Nature 2000; 403: 623-627

Walhout A.J., Boulton S.J., Vidal M. Yeast two-hybrid systems and protein interaction mapping projects for yeast and worm. Yeast 2000; 17: 88-94

Wang Z.-Y., Wu Z.-L., Xing Y.-Y., Zheng F.-G., Guo X.-L., Zhang W.-G., Hong M.-M. Nucleotide sequence of rice waxy gene. Nucl Acid Res 1990; 18: 5898

Wrigley C.W., Batey I.L., Skylas D.J., Sharp P.J. Grain quality assessment using proteomics and genomics. Food Aust 2003; 55: 143-146

Wu B.J., Xie M.T., Chen Y.P., Jiang H., Guo X.X. Cytologic and embryologic studies on apomicts discovered from male-sterility line C1001 in rice (*Oryza sativa* L). Sci China Ser B-Chem Life Sci Earth Sci 1991; 34: 823-831

Yan L., Loukoianov A., Tranquilli G., Helguera M., Fahima T., Dubcovsky J. Positional cloning of the wheat vernalization gene *VRN1*. Proc Nat Acad Sci USA 2003a; 100: 6263-6268

Yan Y., An C., Li L., Gu J., Wang L., Mi S., Chen Z. The sense and antisense expression of gibberellin 20-oxidase gene (*rga5*) in rice and its effects on GA(1) level and agronomic traits. Chin Sci Bull 2003b; 48: 443-448

Yao S.-G., Taketa S., Ichii M. A novel short-root gene that affects specifically early root development in rice (Oryza sativa L.). Plant Sci 2002; 163: 207-215

Yoshigi N., Okada Y., Sahara H., Tamaki T. A structural gene encoding β-amylase of barley. Biosci Biotech Biochem 1995; 59: 1991-1993

Yu L.X., Setter T.L. Comparative transcriptional profiling of placenta and endosperm in developing maize kernels in response to water deficit. Plant Physiol 2003; 131: 568-582

Zakhrabekova S., Kannangara C.G., von Wettstein D., Hansson M. A microarray approach for identifying mutated genes. Plant Physiol Biochem 2002; 40: 189-197

Zentella R., Yamauchi D., Ho T.-H.D. Molecular dissection of the gibberellin/abscisic acid signaling pathways by transiently expressed RNA interference in barley aleurone cells. Plant Cell 2002; 14: 2289-2301

Zhang Z.Y., Salamini F., Thompson R. Fine mapping of the defective endosperm maize mutant *rgf1* using different DNA pooling strategies and three classes of molecular markers. Maydica 2002; 47: 277-286

Zhou K.D., Wang X.D., Luo M., Gao K.M., Zhou S.L., Yan Z.B., Li P., Chen F., Zhou G.M. Initial study on Sichuan Apomixis Rice (Sar-1) Sci China Ser B-Chem 1993; 36: 420-429

Chapter 16

FUNCTIONAL GENOMICS FOR TOLERANCE TO ABIOTIC STRESS IN CEREALS

Nese Sreenivasulu[1,*], Rajeev K. Varshney[1], Polavarpu B. Kavi Kishor[2] and Winfriede Weschke[1]

[1]*Institute of Plant Genetics and Crop Plant Research (IPK), Corrensstrasse 3, D- 06466 Gatersleben, Germany*; [2]*Department of Genetics, Osmania University, Hyderabad 500 007, India*

[*]Author for correspondence: srinivas@ipk-gatersleben.de

1. INTRODUCTION

The world food grain production needs to be doubled by the year 2050 to meet the ever growing demands of the population (Tilman *et al.*, 2002). This goal needs to be achieved despite decreased arable land, dwindling water resources, and the environmental constraints such as drought, water logging, excess heat, frost, salinity, metal toxicity and nutrient imbalances, which cause major losses in cereal grain production. Drought, salinity and cold stress alone are known to cause nearly 35% of cereal crop losses throughout the world (Quarrie *et al.,* 1999). The effectiveness of traditional breeding approaches to deal with the problem is limited due to complex nature of stress tolerance traits and due to incompatibility barriers encountered during transfer of genes from wild species to cultivated ones. Therefore, newer strategies need to be used for developing crop plants that are tolerant to abiotic stresses. Such strategies will include molecular breeding and genetic engineering based on our fast increasing knowledge in genetics, genomics and molecular physiology.

Abiotic stress conditions cause changes in plant metabolism, involving generation of reactive oxygen species (ROS), membrane disorganization, inhibition of photosynthesis and altered nutrient acquisition (Bray, 1993; Ingram and Bartels, 1996; Hasegawa *et al.*, 2000). These changes in turn lead to alterations in development, growth and productivity. Severe stress

P.K. Gupta and R.K. Varshney (eds.), Cereal Genomics, 483-514.

even threatens plant survival. However, tolerant plants can adjust themselves in a number of ways by changing their phenology, morphology, anatomy and physiology. At the molecular level, protection can be achieved by diverse mechanisms. Accumulation of osmoprotectants, water channel activities, production of chaperones, superoxide radical scavenging mechanisms, exclusion or compartmentation of ions by efficient transporter and symporter systems are some of the factors that determine tolerance against salinity, drought and cold (for reviews see Ingram and Bartels, 1996; Ishitani *et al.,* 1997; Thomashow, 1999; Hasegawa *et al.*, 2000; Zhu, 2001; Apse and Blumwald, 2002; Iba, 2002; Shinozaki *et al.,* 2003). Each response involved in stress tolerance is regulated and coordinated by multiple genes, such that the alterations in gene expression profiles of stress-responsive genes are integral parts of stress resistance mechanisms.

The genomic tools and methods that have become available recently provide new opportunities to characterize the gene networks involved and to gain a more holistic view of abiotic stress responses. Expressed sequence tags (ESTs) from abiotic stress-treated libraries of various crop plants, complete genome sequence information for rice and Arabidopsis and the development of new bioinformatics tools allow us to identify the key stress-responsive gene-pools. Furthermore, use of multi-parallel techniques such as expression profiling by microarrays, random and targeted mutagenesis, complementation and promoter-trapping strategies provide important clues for functional characterization of stress responsive genes and stress tolerance mechanisms (Bohnert *et al.*, 2001). Recent genomic studies show considerable overlap of plant responses to cold, drought and salinity stresses (Knight and Knight, 2001; Kreps *et al.*, 2002; Chen *et al.*, 2002; Seki *et al.,* 2002b; Abe *et al.,* 2003) underlining the complexity and provide opportunities to engineer new stress-resistant crop varieties. However, to successfully deal with this complexity, genomics, genetics, physiology and breeding disciplines need to join together to manipulate the genome with precision for abiotic stress tolerance (for reviews see Cushman and Bohnert, 2000; Bohnert *et al.*, 2001). The role of different disciplines and a broad outline of experimental strategies in crop improvement for stress tolerance are indicated in Fig. 1. At first, sources of genetic variation have to be identified and used in strategies to develop new cultivars with greater yield potential and stability over seasons and ecogeographic locations. With the advent of newly developed genomics, two major approaches could be used in exploiting the gene-pool for imparting abiotic stress tolerance: *firstly*, identification and introduction of genes imparting stress-tolerance into crops of interest, and *secondly*, development and identification of molecular markers associated with genes or QTLs (quantitative trait loci) conferring tolerance to stress in germplasm collections and their use in marker-assisted

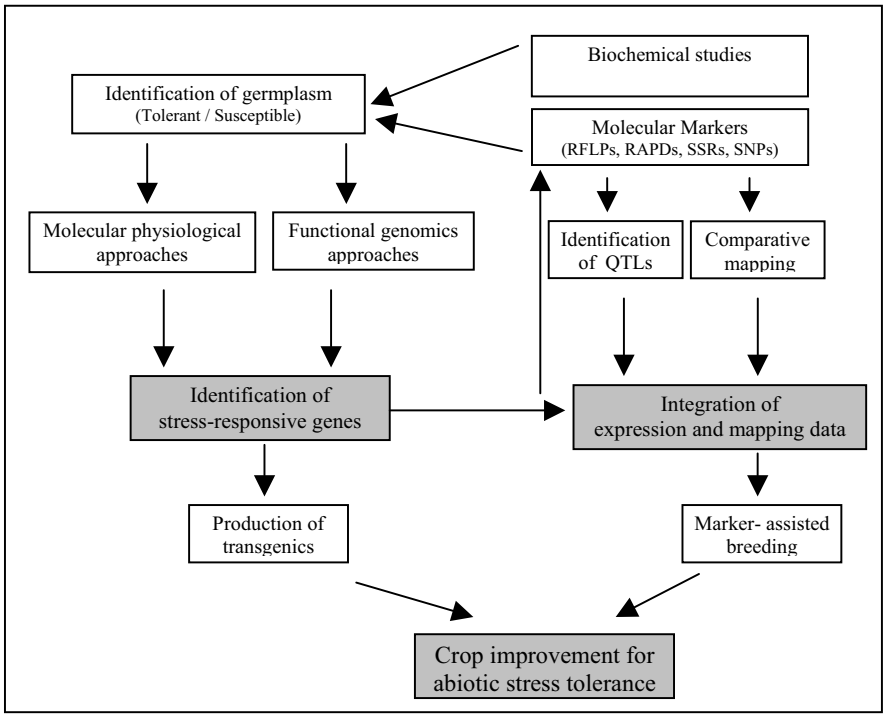


Figure 1. Integrative functional genomic approaches for abiotic stress tolerance.

breeding programs. A complete overview on prime abiotic stress tolerance QTLs is given in detail by Tuberosa and Salvi in Chapter 9 of this book. In the following sections, we review the results of functional genomic approaches to analyze and manipulate abiotic stress tolerance.

2. DISCOVERY OF STRESS-RESPONSIVE GENES FROM CEREALS BY FUNCTIONAL GENOMICS APPROACHES

Functional genomics approaches for abiotic stress tolerance include discovery of novel genes, determination of expression levels of genes induced in response to abiotic stress, studies to understand the functional roles of abiotic stress-responsive genes and generation of stress tolerant

transgenic plants. The strategies of functional genomics approaches to identify abiotic stress-associated mechanisms are discussed in this section.

2.1. Tracing Genes Responsible for Abiotic Stress Tolerance through ESTs

In cereals, large numbers of ESTs have been generated which have great potential to provide functional genomics information (refer web pages http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html; http://www.tigr.org/tdb/tgi/plant.shtml) including that on abiotic stress tolerance. Projects based on this approach were started in rice, barley, wheat, maize and sorghum (http://stress-genomics.org/stress.fls/dbase/dbase.html). Based on the search results (gene index of TIGR database; dated 20.09.03) a total of approximately 13,022 abiotic-stress related ESTs were reported from *Hordeum vulgare*, 13,058 from *Oryza sativa*, 2,641 from *Secale cereale*, 17,189 from *Sorghum bicolor*, 20,846 from *Triticum aestivum* and 5,695 from *Zea mays*. However, the number of ESTs generated so far solely from stress-treated libraries is low, as compared to total ESTs. Therefore, there is a need to enforce sequencing programmes from stress-tolerant genotypes of cereals (treated with different abiotic stresses) covering a wider range of tissue types and developmental stages. We surveyed all the publicly available EST collections from cereal species (barley, maize, rice, and wheat) and identified drought- and salt stress-responsive genes (Table 1). Since we used an EST dataset from non-normalized libraries, the EST clustering results provide information on relative expression levels of stress-responsive genes belonging to different pathways. The ESTs are selected from non-normalized cDNA libraries of cereals and subjected for clustering analysis by software StackPACK v2.1.1. As a result of clustering, homologous sequences were grouped together to identify abundantly expressed gene sets. Among them, genes associated with stress-relevant pathways were found commonly expressed in drought- and salt-treated cereal plants (Table 1). Recently, Reddy *et al.* (2002) used normalized cDNA libraries from drought-stressed seedlings of rice to select novel stress-responsive genes. They identified genes metallothionein-like proteins, glyceraldehyde-3-phosphate dehydrogenase, aldolase, rd22, glycine-rich protein, glutathione-S-transferase, catalase, LEA and HSP, and several transcription factors (DREB, MYB, MYC, AP2, Zinc finger protein) as well as kinases (mitogen activated protein kinases, calcium-dependent protein kinase) that were abundantly expressed upon drought stress.

Table 1. ESTs that are abundant in cDNA libraries from drought, salt-treated cereals

Functional catalogue	Putative gene identification	Drought	Salinity
		No.ESTs /cluster	No.ESTs /cluster
Aquaporins	Tonoplast intrinsic protein, (gamma tip)	***	**
	Plasma membrane intrinsic protein 1	**	***
	Plasma membrane intrinsic protein 2	***	**
Antioxidants	Glycine-rich RNA-binding protein	******	*****
	Glutathione s-transferase	****	***
	Glutathione peroxidase (PHGPX), chloroplast	**	**
	Phenylalanine ammonia-lyase	****	****
	L-ascorbate peroxidase, cytosolic	**	**
Growth regulators	Indole-3-acetic acid induced protein arg 2	**	***
	Abscisic acid induced protein	*	**
Osmoprotectants	Proline-rich protein	**	***
	Glyceraldehyde-3-phosphate dehydrogenase	**	**
	Mannitol dehydrogenase	**	*
Protein destination	Metallothionein-like protein 1	******	******
	Ubiquitin	***	*****
	Cysteine proteinase	***	****
	Cysteine protease inhibitor	****	*****
Photosynthetic	Chlorophyll a-b binding protein 3c	******	******
	Rubisco small subunit	******	*****
	Rubisco small subunit c	******	*****
	Photosystem I reaction center subunit psak	*****	*****
	Rubisco activase a	****	****
	Rubisco large subunit	*****	****
Stress responsive genes	Non-specific lipid-transfer protein	*****	****
	Glycine-rich-protein	****	***
	Osmotin-like protein	**	*
	Thaumatin-like protein	**	*
	Late embryogenesis abundant protein 3	**	*
	Late embryogenesis abundant protein 2	**	*
	Late embryogenesis abundant protein 1	*	*
	Heat shock protein 81	**	**
	Heat shock protein 70	**	*
	Heat shock protein 17	**	**

EST mining approach: The consensus sequences of abundantly expressed transcripts were subjected to similarity search (BLASTX) against public protein database (SWISSPORT) for functional annotation using the arbitrary criteria expectation value less than 1.0 x 10^{-15}. The genes (represented by sequence clusters) with their putative function assigned to the functional catalogues are listed in the table. Number of ESTs present in each sequence cluster reflects the relative expression level of the corresponding gene (*2 to 5 ESTs; **6 to 10; ***11 to 15; ****16 to 25; *****26 to 50; ****** 51 and above) in different cereals (barley, rice, wheat and maize).

2.1.1. Transcription Profiling using ESTs from Abiotic Stress Related cDNA Libraries

The cDNA macro/microarray technology for transcript profiling has been established, based on EST programmes, in cereal species such as barley (Ozturk *et al.,* 2002; Sreenivasulu *et al.*, 2002, 2004a), maize (Lee *et al.*, 2002) and rice (Kawasaki *et al.*, 2001) after generating non-redundant unigene sets. Transcript profiling based on micro/macro-arrays was carried out in cereal crops (Kawasaki *et al.*, 2001; Ozturk *et al.,* 2002; Sreenivasulu *et al.,* 2004b) as well as in Arabidopsis (Seki *et al.*, 2001, 2002a, 2002b) to analyse gene expression in response to a variety of stresses. Array technology provides a powerful tool (i) to compare the relative expression levels between tolerant and sensitive cultivars within the same species under stress and control conditions and (ii) to identify stress-specific transcriptional responses as well as cross-talks between different stress responses.

2.1.1.1. Expression levels in tolerant and sensitive cultivars

In rice, microarrays based on 1,728 stress-regulated transcripts (obtained from seedlings of the salt-tolerant rice variety Pokkali) were used for large-scale gene expression profiling in salt-tolerant Pokkali as well as in the salt-sensitive variety IR29 during 15-minutes to 7-day time intervals under control as well as high salinity treatments (Kawasaki *et al.,* 2001). In the tolerant cultivar, changes in transcript levels were observed first as early as 15 min after salt stress. Upregulated genes could be assigned to signaling pathways (calcium-dependent protein kinases, nucleoside diphosphate kinase), cell division processes (40S ribosomal proteins and elongation factor-1α, glycine rich proteins), protease inhibitors and hormonal induced genes (see Table 2). As a corollary, these genes were generally downregulated in the sensitive cultivars. During long-term salt stress (24 h and 7 days), tolerant rice plants showed upregulation of antioxidant transcripts (glutathione-S transferase, ascorbate peroxidase), aquaporins (water channel protein I and IV), protease inhibitors (subtilisin inhibitor, trypsin inhibitor), hormonal induced and some unknown genes. Similar transcript profiling studies were carried out in 3-week old barley seedlings, where salt as well as drought-responsive genes were identified (Ozturk *et al.* , 2002). The upregulated gene set in salt-stressed leaf and root tissues encodes antioxidants and osmoprotectants and in addition contains genes for protein destination, and regulatory and stress-response processes (see Table 2). In drought-stressed barley leaves, transcripts encoding proteins of

jasmonate biosynthesis (allene oxide synthases) and several jasmonate-induced proteins were upregulated along with amino acid metabolism genes, osmoprotectants, protein destination and stress responsive genes (see Table 2). Since extensive synteny exists between different grass genomes (Gale and Devos, 1998), cDNA arrays developed from one grass species can be used for the analysis of other species. Therefore, we explored the possibility to use barley cDNA arrays for examination of gene expression patterns in tolerant and sensitive seedlings of foxtail millet (*Setaria italica* L.) exposed to 250 mM NaCl. The upregulated 14 transcripts in the salt-tolerant line includes protease inhibitors, antioxidative enzymes and some unknown genes, which are similar to salt-responsive genes already identified in rice and barley (Sreenivasulu *et al.,* 2004b).

Scientists at Pioneer Hi-Bred performed cDNA microarray analysis in a maize breeding population that showed improved tolerance to water stress during ear growth (c.f. Bruce *et al.,* 2002). They reported that synthesis of water channel aquaporins and β-glucosidase transcripts was down-regulated during stress, whereas during the recovery period these transcripts were up-regulated. Their results indicated that a family of cell cycle genes exhibit three different gene expression patterns: (a) increasing mRNA levels during drought stress; (b) decreasing mRNA levels during stress followed by a subsequent increase during recovery; (c) increasing mRNA levels only during recovery. These results suggest specific functions of different members of the cell cycle gene family. Recently, transcript profiling was performed for placenta and endosperm of maize kernels grown under water deficit (Yu and Setter, 2003). Only eight out of 70 genes upregulated under water stress in the placenta were also upregulated in the endosperm. The related proteins have expected roles in stabilization of proteins as well as membrane structure during stress (for instance, 70 kD heat shock protein, DNAJ and lipid transfer protein), show aquaporin function (plasma membrane intrinsic protein) or are involved in trehalose synthesis (trehalose-6-phosphate synthase) expected to stabilize macromolecule structures during stress (Garg *et al*., 2002). Microarrays based on 11,000 unique, full-length cDNA sequences (outcome of the Rice Genome Research Program) were used to study responses of rice seedlings to UV-B and gamma irradiation (Kikuchi *et al.*, 2002). Although both types of irradiation induce similar physiological effects, very few genes were induced in parallel, including those for polygalacturonase inhibitor (PGIP), major intrinsic protein, beta tubulin, eukaryotic initiation factor, lipid transfer protein and metallotionein-like protein.

Table 2. Genes upregulated by abiotic stress – an index from microarray analysis

Functional class	Genes
I. Barley seedlings (3 week-old) exposed to dehydration: 6 & 10 h Ozturk *et al.* (2002)	
Amino acid metabolism	Arginine decarboxylase 2, Arginine decarboxylase SPE2 Asparagine synthetase, Tryptophan synthase beta chain 1
Jasmonate biosynthesis	Allene oxide synthase
jasmonate induced proteins	Lipoxygenase 2 (methyl jasmonate-inducible) Jasmonate-induced protein (jip) 60 kD, jip 23kD, jip 1 and jip 6
Osmoprotectants	Delta-1-pyrroline-5-carboxylate synthetase
Protein destination	Metallothionein-like protein type 2
Stress responsive genes	Dehydrin 9, Late embryogenesis abundant protein 14-A
II. Barley seedlings (3 week-old) exposed to 150 mM NaCl: 24 h Ozturk *et al.* (2002)	
Antioxidants	Glutathione-S-transferase (auxin-induced)
Jasmonate biosynthesis	Allene oxide synthase
Osmoprotectants	Proline rich protein, Delta-1-pyrroline-5-carboxylate synthetase
Photosynthetic	Photosystem II 10 K protein
Protein destination	Metallothionein-like protein type 2, Aspartic proteinase
Regulatory	Transcription factor POU3A, Acidic ribosomal protein 60S Replicase associated polyprotein
Stress responsive genes	Heat shock protein DNAJ, Lipid transfer protein cw18, Late embryogenesis abundant like protein
Unknown	6 unknown genes
III. Rice seedlings exposed to 150 mM NaCl: 15 min, 1h, 3h and 6h Kawasaki *et al.* (2001)	
Hormonal induced	Gda-1 (gibberellic acid-induced gene) Asr1 (ABA and stress-induced protein) Osr40c1 (ABA and salt-induced protein)
Protein destination	Subtilisin-chymotrypsin inhibitor 2, Trypsin inhibtor 1
Regulatory	Calcium-dependent protein kinase, Nucleoside diphosphate kinase Calmodulin, Protein phosphatase 2C homologue, Elongation factor 1 40S ribosomal protein S4, 40S ribosomal protein S7
Stress responsive genes	Glycine/serine-rich protein (grp) 1, grp 2
Unknown	5 unknown genes
IV. Rice seedlings exposed to 150 mM NaCl: 24h and 7 days Kawasaki *et al.* (2001)	
Antioxidants	Glutathione-S-transferase, Ascorbate peroxidase, cyt
Aquaporins	Water channel protein I, Water channel protein IV
Hormonal induced	Gda-1 (gibberellic acid-induced gene) Osr40c1 (ABA and salt-induced protein), Osr40g2
Protein destination	Trypsin inhibtor 1, Metallothionein-like protein
Unknown	3 unknown genes

Table 2. Continued

V. Foxtail millet seedlings exposed to 250 mM NaCl: 7 days Sreenivasulu *et al.* (2004b)	
Antioxidants	Glutathione peroxidase, L-ascorbate peroxidase, cyt, Catalase
Protein destination	Trypsin inhibitor, Subtilisin-chymotrypsin inhibitor
Regulatory	Kruppel-like transcription factor, Argonaute protein, Cyclophilin
Unknown	1 unknown gene
VI. Maize developing kernels exposed to drought stress Yu and Setter (2003)	
Aquaporins	Plasma membrane intrinsic protein
Protein destination	20S proteasome beta subunit PBD2
Regulatory	Calcium-dependent protein kinase, TATA binding protein, Small nuclear ribonucleoprotein, Histone H2A, Cyclophilin
Stress responsive genes	Heat shock protein 70 kDa, Lipid transfer protein
Unknown	4 unknown genes

Significantly upregulated transcripts in barley, rice, maize and millet were considered (2.5 fold deviation from the control plant expression values, includes repeat experiments).

2.1.1.2. Stress specific responses

Since only few transcriptional profiling studies were conducted in cereals, we here include studies on Arabidopsis in order to gain deeper insights into functional genomic aspects of multiple stress interactions. Using 1300 full-length clones (Seki *et al.,* 2001) and 7,000 full-length clone inserts (Seki *et al.,* 2002a, 2002b) multistress interactions of abiotic stress treatments were studied to identify genes of potential interest to salt, drought and cold responses. By using 1,300 full-length clones, Seki *et al.* (2001) identified a set of only 44 genes, which are induced either by drought or cold stress response. Among them, 12 were identified as stress-inducible target genes of the DREB1 transcription factor family. By using 7,000 full-length inserts, 299 drought-inducible genes, 213 high-salinity-stress-inducible genes, 54 cold-inducible genes and 245 ABA-inducible genes were identified (Seki *et al.,* 2002a, 2002b). Multistress interactions of abiotic stress treatments were studied by Kreps *et al.* (2002) using a larger array containing oligonucleotides for about 8,100 Arabidopsis genes, to identify genes of potential interest to salt, drought and cold responses. They identified changes in gene expressions (more than 2-folds over control) for 2,409 out of 8,100 genes as part of cold, drought and salt responses. Above differences in the lists of stress-inducible genes found by using the full-length cDNA array or the oligonucleotide gene chip array might be due to the presence of different sets of genes on the respective arrays (only 1919 genes are common between both arrays) and different plant growth conditions as well as stress treatments used for experiments. Shinozaki *et al.*

(2003) analysed the complex cascades of gene expression in drought and cold stress responses and made an attempt to demonstrate the regulatory network of gene expression in drought and cold stress responses. Recently, Chen *et al.* (2002) identified approximately 21 transcription factors preferentially induced by abiotic stress conditions such as salinity-, osmotic-, cold- and jasmonic acid treatment. These transcription factors include DRE/CRT binding factors (shown to be activated by cold stress by Liu *et al.*, 1998), CCA1 and Athb-8 (shown to be regulated by hormones by Baima *et al.*, 2001), Myb proteins, bZIP/HD-ZIPs and AP2/EREBP domain transcription factors.

Comparative analysis of the response to abiotic stresses among diverse tolerant species can lead to the identification of evolutionarily conserved and unique stress defense mechanisms. By applying clustering algorithms to large-scale gene expression data of abiotic stress responses, stress-regulons, i.e. sets of genes regulated in a similar fashion, can be identified. This approach also enables the identification of new promoter elements/transcription factor binding sites in co-expressed gene sets and further helps to explore regulatory networks controlling abiotic stress responses (Aarts *et al.*, 2003). However, mining information will not reveal the complete functions of stress-regulated genes. Other approaches are necessary as, for instance, activation tagging. In Arabidopsis (ecotype C24) 43,000 T-DNA insertion lines were generated (Weigel *et al.*, 2000; http://stress-genomics.org/stress.fls/tools/mutants.html), of which about 30,000 lines were screened for stress-related gene regulation mutants (Xiong *et al.*, 1999); details of these results are available on web (http://stress-genomics.org/stress.fls/tools/mutants/arabid/T_DNA_ mutants/table1.html).

2.2. Functional Aspects of Abiotic Stress Tolerance Mechanisms Identified Through Molecular-Physiological Studies and Transgenics

In silico mining and transcription profiling led to the discovery of a larger number of genes involved in abiotic stress responses (Tables 1, 2 and 3). These genes can be used in functional studies, preferentially by transgenic approaches (Table 4 and Grover *et al.*, 2003). The results of these studies will be discussed in the following this section.

Table 3. Genes encoding enzymes/proteins associated with abiotic stress response in cereals

Gene category	Gene	Species	Cellular response	Reference
Antioxidants				
Superoxide dismutase	*FeSOD*	Maize	Cold stress	Van Breusegem *et al.* (1999)
	cyt Cu/ZnSOD	Rice	Drought, Heat stress	Sakamoto *et al.* (1995)
	chl Cu/ZnSOD	Rice	Abiotic stress	Kaminaka *et al.* (1997)
	MnSOD	Millet	Salt stress	Sreenivasulu *et al.* (2000)
Catalase	*CAT*	Maize	Cold stress	Prasad *et al.* (1994)
Ascorbate peroxidase	*APX*	Maize	Cold stress	Prasad *et al.* (1994)
		Millet	Salt stress	Sreenivasulu *et al.* (2000)
Osmolyte compounds				
Proline	*P5CS*	Wheat	Salt stress	Sawahel and Hassan (2002)
Glycine betaine	*Chlcod*	Rice	Salt, Cold stress	Sakamoto *et al.* (1998)
	bet A	Rice	Salt, Cold stress	Takabe *et al.* (1998)
	BADH	Sorghum	Osmotic stress	Wood *et al.* (1996)
Mannitol	*ADC*	Rice	Drought stress	Capell *et al.* (1998)
	ADC	Rice	Salt stress	Roy and Wu (2001)
Regulatory genes				
bZIP	*OSBZ8*	Rice	ABA	Nakagawa *et al.* (1996)
	OsZIP-1a	Rice	ABA	Nantel and Quatrano (1996)
	EmBP1	Wheat	ABA	Hobo *et al.* (1999)
	TRAB1	Wheat	ABA	Choi *et al.* (2000)
Stress-responsive genes				
LEA proteins	*HVA1*	Rice	Salt, Drought stress	Xu *et al.* (1996)
	HVA1	Barley	Abiotic stress	Hong et (1992)
	HVA1	Barley	Cold stress	Sutton *et al.* (1992)
	HVA1	Wheat	Freezing tolerance	Sivamani *et al.* (2000)
	DHN1-DHN12	Barley	Salt, Drought stress	Choi *et al.* (1999)
	DHN	Wheat	Drought stress	Labhilili *et al.* (1995)
COR or BLT genes	*COR14b*	Barley	Cold stress	Cattivelli and Bartels (1990)
	BLT4, BLT14	Barley	Cold stress	Pearce *et al.* (1998)
	BLT63	Barley	Cold stress	Dunn *et al.* (1993)
	BLT801	Barley	Cold stress	Dunn *et al.* (1996)
Thaumatin-like protein	*TLP-D34*	Rice	Osmotic stress	Datta *et al.* (1999)
Low temperature induced protein	*LIP5, LIP9, LIP19*	Rice	Cold stress	Aguan *et al.* (1991)
RAB genes	*RAB16A*	Rice	Drought stress	Mundy *et al.* (1990)
	RAB17	Wheat	Drought stress	Vilardell *et al.* (1990)

Table 3. Continued

RAB genes	*RAB28*	Maize	Drought stress	Pla *et al.* (1993)
WCS genes	*WCS120*	Wheat	Cold stress	Oullet *et al.* (1998)
	WCS19	Wheat	Cold stress	Chauvin *et al.* (1993)
Heat shock protein	*HSP90*	Rice	Heat stress	Pareek *et al.* (1995)
	HSP104	Rice	Heat stress	Singla and Grover (1994)
	HSP16.9	Rice	Heat stress	Tzeng *et al.* (1992)
Transporters				
Na^+-K^+-symporter	*OsHKT1*	Rice	Salt stress	Horie *et al.* (2001)
	OsHKT2	Rice	Salt stress	Horie *et al.* (2001)
Na^+-H^+-dependent K^+ transporter	*EcHKT2*	Barley	Salt stress	Rubio *et al.* (1999)

2.2.1. Genetic Engineering for Osmolyte Biosynthesis in Cereals during Stress

Many monocotyledonous plants including cereals evolved different mechanisms for balancing osmotic strength of cells under salt/water stress conditions. Cereals like wheat, sorghum, maize and pearl millet can avoid dehydration by synthesizing different organic osmolytes that are compatible with cellular functions and can help as osmotic balancing agents, if accumulated in large quantities. A majority of the compounds can function as osmoprotectants. Almost all cereals accumulate proline albeit to a lesser extent relative to other osmo-tolerant plants; some cereals (wheat, maize, sorghum, barley) accumulate glycine betaine in response to salt and drought stresses. It seems that cereals do not accumulate sugars such as trehalose and sugar alcohols like ononitol, pinitol, etc. during exposure to abiotic stresses. Genes associated with the accumulation of various osmoprotectants have been the target for genetic engineering studies for more than a decade to develop genotypes tolerant to salt and water stresses. In most of the cases, introduction of a single gene into a plant (mostly dicots) resulted in only a moderate increase in tolerance with a modest accumulation of osmoprotectants. In the following we describe molecular physiology and genetic engineering work related to the synthesis of osmoprotectants such as proline, glycine betaine and sugar alcohols in cereal crops. Transgenic cereals that accumulate various compatible solutes and could sustain moderate abiotic stress treatments are listed in Table 4.

2.2.1.1. Proline

Proline accumulates in plants exposed to many abiotic stresses, due to upregulation of the gene for pyrroline 5-carboxylate synthetase (P5CS), a key enzyme that converts glutamate to Δ^1-pyrroline-5-carboxylic acid (P5C)

Table 4. Transgenic cereal plants developed for abiotic stress tolerance

Gene category/Gene	Species	Cellular response	Reference
Antioxidants			
Mn-superoxide dismutase	Rice	Salt tolerance	Tanaka *et al.* (1999)
Mn superoxide dismutase	Rye grass	Winter hardiness	McKersie (1999)
Fe- superoxide dismutase	Maize	Cold tolerance	Van Bruesegem *et al.* (1999)
Catalase	Rice	Chilling tolerance	Matsumura et al. (2002)
Osmolyte compounds			
Pyrroline carboxylate synthase *(p5cs)*	Rice	Drought, Salt tolerance	Igarashi *et al.* (1997) Zhu *et al.* (1998)
p5cs	Rice	Oxidative, Osmotic tolerance	Hong *et al.* (2000)
p5cs	Wheat	Salt tolerance	Sawahel and Hassan (2002)
Choline dehydrogenase	Rice	Drought, Salt tolerance	Takabe *et al.* (1998)
Choline oxidase	Rice	Cold, Salt tolerance	Sakamoto and Murata (1998) Mohanty *et al.* (2002)
Trehalose-6-P-synthase Trehalose-6-P-phosphatase	Rice	Salt, Drought, Cold tolerance	Garg *et al.* (2002) Jang *et al.* (2003)
Mannitol dehydrogenase	Wheat	Drought, Salt tolerance	Abebe *et al.* (2003)
Glycerol-3-phosphate acyltransferase	Rice	Cold tolerance	Yokoi *et al.* (1998)
waxy gene	Rice	Cold tolerance	Hirano and Sano (1998)
Glutamine synthetase	Rice	Salt, Cold tolerance	Hoshida *et al.* (2000)
Arginine decarboxylase	Rice	Drought tolerance	Capell *et al.* (1998)
Regulatory genes			
Calcium depdendent protein kinase	Rice	Salt, Drought, Cold tolerance	Saijo *et al.* (2000)
DREB1A	Wheat	Drought tolerance	Pellegrineschi *et al.* (2002)
Stress-responsive genes			
Late embryogenesis protein group 3 *(HVA1)*	Oat	Drought tolerance	Maqbool *et al.* (2002)
HVA1	Wheat	Drought tolerance	Sivamani *et al.* (2000)
HVA1	Rice	Drought, Salt tolerance	Xu *et al.* (1996) Rohila *et al.* (2002)
Thaumatin-like protein	Rice	Osmotic adjustment	Datta *et al.* (1999)
Heat shcok protein 101	Rice	High temperature tolerance	Katiyar-Agarwal *et al.* (2003)
Ferritin	Rice	Enhanced iron storage	Deak *et al.* (1999)
Pyruvate decarboxylase1	Rice	Submergence tolerance	Quimio *et al.* (2000)
Alcohol dehydrogenase	Rice	Flooding tolerance	Minhas and Grover (1999)
Transporters/symporter			
Potassium transporter *(HKT1)*	Wheat	Salt tolerance	Laurie *et al.* (2002)
Na+/H+ antiporter	Rice	Salt tolerance	Ohta *et al.* (2002)

in the proline biosynthetic pathway. Proline is formed from P5C by P5C reductase (P5CR) both in prokaryotes and eukaryotes. Initially, Kishor *et al.* (1995) overexpressed a mungbean P5CS gene in transgenic tobacco and reported accumulation of proline up to 18-fold over control plants resulting in enhanced biomass production under salt stress. Similarly, a P5CS gene isolated from rice was transferred back into rice (Igarashi *et al.,*1997), where over expression resulted in enhanced root biomass and flower development under water and salt stress conditions. In another set of experiments, a P5CS gene from *Vigna aconitifolia* was introduced into wheat plants using *Agrobacterium*-mediated gene transfer (Sawahel and Hassan, 2002). Transgenic analyses proved the expression of the transferred gene, and salinity tests indicated increased salt tolerance (Table 4) supporting the notion that proline acts as an osmoprotectant in transgenic wheat plants also. Proline is synthesized not only from glutamate but also from arginine/ornithine. Ornithine is transaminated to glutamic semi-aldehyde (GSA) by ornithine δ-aminotransferase (δ-OAT), which subsequently gets converted to proline via P5C (Delauney *et al.*, 1993). However, this gene has not been transferred yet into cereals though it was introduced and conferred salt stress tolerance in other plants (Madan *et al.,* 1995; Roosens *et al.*, 1998). A proline transporter (ProT) cDNA was isolated from *Oryza sativa* cv. Akibare (Igarashi *et al.*, 2000) and was shown to specifically transport L-proline in a transport assay. Although mRNA levels of ProT2 were observed throughout the plant, its transcript levels were found to be strongly induced by water or salt stress (Hare and Cress, 1997), suggesting an increase of proline transport during osmotic stress conditions.

Proline has also been shown to reduce enzyme denaturation caused by abiotic stress treatments such as salt, water, heavy metal, and UV radiation. (Iyer and Caplan, 1998). Under stress, it is mainly synthesized in chloroplasts and protects photosystem II against photodamage. Intermediates in proline biosynthesis and catabolism, such as glutamine and P5C also increase the expression of several osmotically regulated genes in rice, including *salT* and *dhn4* (Iyer and Caplan, 1998).

2.2.1.2. Glycine betaine

Betaines have been found to stabilize the quaternary structure of proteins and membranes. They also protect photosystem II from salt induced inactivation (Papageorgiou and Murata, 1995). Glycine betaine (GB) is a dipolar, electrically neutral molecule. It is synthesized from serine, which gets converted to choline *via* a series of steps that are not characterized properly. Unlike bacteria, plants possess choline monooxygenase (CMO), a ferridoxin dependent soluble Rieske-type protein, which oxidizes choline to

betaine aldehyde. CMO is a stress inducible iron-sulphur enzyme localized in the chloroplast stroma (Russell *et al.*, 1998). Betaine aldehyde dehydrogenase (BADH) is a soluble NAD^+ dependent enzyme that converts betaine aldehyde to glycine betaine. A positive correlation was found between the accumulation of betaines and tolerance to salt and cold, respectively, in maize and barley (Kishitani *et al.*, 1994). The gene coding for BADH is upregulated under high salt or drought conditions in wheat plants (Guo *et al.*, 2000).

Pathways for production of glycine betaine vary between organisms. In some bacteria choline gets converted to betaine directly by choline oxidase (*codA*), but the gene encoding the enzyme was not found in higher plants. Genetically engineered rice with the ability to synthesize GB was established by introducing the *codA* gene from the soil bacterium *Arthrobacter globiformis*. Levels of GB were high in two types of transgenic plants in which *codA* was targeted either to the chloroplasts (ChlCOD) or the cytosol (CytCOD). Inactivation of photosynthesis, used as a measure of cellular damage, indicated that ChlCOD plants were more tolerant than CytCOD plants to photoinhibition under salt and low-temperature stress. These results indicate that the subcellular compartmentalization of GB biosynthesis is a critical element in the enhancement of tolerance to stress in the engineered plants (Sakamoto and Murata 1998). Rice plants that produced bacterial choline dehydrogenase (CDH) targeted to the mitochondria were also generated (Takabe *et al.*, 1998). These transgenics accumulated GB at levels similar to those in transgenic rice that produced COD and showed enhanced tolerance.

2.2.1.3. Sugars and sugar alcohols

Sugars play an essential role as osmolytes and function also in signal transduction during development and under stress (Smeekens, 2000). Based on molecular genetic approaches, a link between hexose-sugar sensing and ABA signal transduction was found in Arabidopsis (Smeekens, 2000), and it was shown that an ABA-dependent signal transduction pathway is involved in the induction of stress genes (Zhu, 2002). This complex situation suggests that genes involved in carbohydrate metabolism and those involved in ABA biosynthesis can also be used to engineer abiotic stress tolerance.

Accumulation of a variety of polyhydroxylated sugar alcohols (polyols) such as trehalose, sorbitol, mannitol, ononitol, pinitol, etc. was reported in organisms osmotically stressed by drought and salinity (Csonka and Hanson, 1991) but not in cereals. While mannitol is synthesized from fructose 6-phosphate, other sugar alcohols like sorbitol, ononitol and pinitol are synthesized from glucose 6-phosphate. Sorbitol accumulates under

drought and salinity stress conditions and plays an important role in abiotic stress tolerance. The enzyme aldose-6-phosphate reductase involved in sorbitol synthesis was identified in barley. This enzyme was shown to be transcriptionally regulated under osmotic stress conditions (Bartels and Nelson, 1994) and hence could be an important candidate for overexpression. Mannitol also provides enhanced tolerance in response to high salinity or water stress. Recently, Abebe *et al.* (2003) demonstrated that ectopic expression of the *E. coli* gene for mannitol-1-phosphate dehydrogenase (mtlD) involved in mannitol biosynthesis improves tolerance to drought and salinity stress in wheat.

Trehalose as a compatible solute, might be involved in the stabilization of biological structures under abiotic stress conditions. Trehalose accumulation is reported in *Escherichia coli* but not in plants. Trehalose biosynthesis is controlled by the *ots*A and *ots*B loci in *E. coli*, which encodes trehalose 6-phosphate synthase (*ots*A) and trehalose 6-phosphate phosphatase (*ots*B). *Ots*A catalyzes the formation of trehalose 6-phosphate from UDP-glucose and glucose 6-phosphate. Further, *ots*B catalyzes the formation of trehalose from trehalose 6-phosphate (Kaasen *et al.,* 1994). Garg et al. (2002) reported the overexpression of *E. coli* trehalose biosynthetic genes (*ots*A and *ots*B) in Pusa Basmati rice as a fusion gene by using tissue-specific and stress-dependent promoters. In this study, comparison to control plants, several transgenic rice lines accumulated increased amounts of trehalose and exhibited sustainable plant growth under salt, droght and low-temperature stress conditions. Also, the transgenic plants in this study exhibited improved photosystem II function.

2.2.2. Genetic Engineering of Detoxification Pathways for Abiotic Stress Tolerance

All cereal crops that grow under a variety of adverse environmental conditions are prone to oxidative damage. Therefore, they have to deal with the highly reactive nature of oxygen derivatives such as superoxide radicals, hydrogen peroxide, hydroxyl and lipid radicals. Higher plants possess an array of antioxidant molecules (vitamin C, vitamin E, carotenoids, flavonoids) and antioxidant enzymes such as superoxide dismutase (SOD), catalase, ascorbate peroxidase (APX), glutathione peroxidase (GPX), glutathione S-transferase (GST) and glutathione cycle enzymes. It is evident that radical-induced damage (oxidative damage) is typically found in stress situations such as heat, cold, ultraviolet light, drought, salinity and heavy metals. A variety of physiological studies showed correlations between levels of antioxidants and stress tolerance among diverse cereal varieties

and biotypes (see, for instance, Sreenivasulu *et al.*, 2000; Prasad *et al.*, 1994). Comparisons between heat and drought-tolerant maize inbreeds displayed a correlation between antioxidant defense enzymes and heat as well as drought stress (Malan *et al.*, 1990). Similarly, comparative studies between salt tolerant and salt sensitive millet (*Setaria italica*) for general peroxidases, ascorbate peroxidase and superoxide dismutases revealed higher expression of these antioxidative transcripts/enzymes in the salt-tolerant cultivar during salt exposure (Sreenivasulu *et al.*, 1999; 2000). Since the antioxidant enzymes are important in protection against a variety of environmental stresses, production of transgenic cereals with genes encoding modified antioxidant enzymes is very promising. When rice was transformed with the yeast mitochondrial MnSOD gene, the transgenics displayed resistance to H_2O_2. Also, these transgenics were more resistant to salt stress than the control plants (Tanaka *et al.*, 1999). A *Nicotiana plumbaginifolia* MnSOD gene was used to generate transgenic maize plants overproducing MnSOD. To target this mitochondrial enzyme into chloroplasts, the MnSOD-coding sequence was fused to a sequence encoding a chloroplast transit peptide from a pea ribulose-1,5-bisphosphate carboxylase/oxygenase gene and engineered behind the CaMV 35S promoter. Transgenic MnSOD activity contributed to 20% of the total SOD activity and had clear effects on foliar tolerance to chilling and oxidative stresses. The results suggested that overproduction of MnSOD in the chloroplasts increased the antioxidant capacity of the maize leaves (Van Breusegem *et al.*, 1999). Similarly, overexpression of MnSOD as well as Cu/ZnSOD conferred freezing and drought tolerance in alfalfa (McKersie *et al.*, 1999). Among hydrogen peroxide scavenging enzymes, wheat catalase gene was overexpressed in rice and the transgenic rice plants exhibited reduction of hydrogen peroxide levels under chilling stress (Matsumura *et al.,* 2002). Likewise, transgenic overexpression of GST and GPX in tobacco led to accumulation of higher levels of glutathione and ascorbate relative to wild type seedlings, which in turn resulted in reduced oxidative damage and a higher degree of salt tolerance (Roxas *et al.*, 2000).

2.2.3. Stress Responsive Genes from Cereals and Their Effect on Stress Tolerance in Transgenic Plants

Stress-related genes were also isolated from cereals, which could be broadly classified into Late Embryogenesis Abundant (*LEA*) genes**,** Dehydrin genes (*DHN*), Cold Responsive genes (*COR*), Early Light Inducible Protein genes (*ELIPs*), etc. *LEA* genes are induced in vegetative tissues during dehydration, salinity, cold, ABA treatments and also in seeds during the desiccation phase (Dure, 1993). They are grouped into three classes (1, 2

and 3), and many of them were cloned from Triticeae species also. Among the *LEA* class 2 proteins, dehydrins are characterized by lysine rich amino acid sequences at the C-terminus. Dehydrins (*DHN*; *LEA D11*) are water-soluble lipid-associating proteins that are exclusively expressed during dehydration conditions, and are thought to play a role in freezing and drought tolerance in plants (Close, 1997; Ismail *et al.,* 1999). Choi *et al.* (1999) identified 11 unique *DHN* genes and estimated a total of 13 *DHN* genes in the barley genome. In addition, *DHN* genes were characterized in a wheat drought-tolerant cultivar (Labhilili *et al.,* 1995). The *LEA* class 3 gene, *HVA1* was isolated from barley and transferred to rice (Japonica). The transgenics exhibited enhanced accumulation of the *HVA1* protein, increased tolerance to water deficit and salt stress, higher growth rates, delayed stress-related damage symptoms as well as faster and improved recovery after stress removal (Xu *et al.*, 1996). Transgenic wheat plants containing the constitutively expressed *HVA1* gene also resulted in improved growth characteristics under water-deficit conditions. As compared to the control, the transgenics produced more biomass and showed higher water use efficiency (Sivamani *et al.*, 2000).

Cold responsive genes such as *COR* or *BLT* form a small gene family shown to be involved in cold and frost tolerance (Grossi *et al.,* 1998; Cattivelli *et al.,* 2002). Cattivelli and Bartels (1990) isolated the cold induced chloroplast localized *COR14b* gene from barley. Constitutive expression of *COR15a* gene of *Arabidopsis thaliana* results in a significant increase in the survival of isolated protoplasts frozen at –7 ^{o}C (Steponkus *et al.,* 1998). The *BLT* genes found to be induced under low temperaurtes encode *BLT4* (non-specific lipid transfer protein), *BLT63* (elongation factor 1α), *BLT801* (RNA binding protein) and *BLT14* (Dunn *et al.,* 1993; Dunn *et al.,* 1996; Pearce *et al.,* 1998). High temperatures cause high membrane fluidity and plants adapted to high temperatures contain a high proportion of saturated fatty acids in the membranes. Exposure to high temperatures causes synthesis of heat shock proteins (*HSP*) that play a defensive role by stabilizing proteins and membrane structures. Recently it was reported that transfer of the *HSP*101 gene from *Arabidopsis thaliana* mediates enhanced tolerance to high temperature stress in rice (Katiyar-Agarwal *et al.*, 2003).

2.2.4. Engineering Ion Transport and Homeostasis Genes

Salt stress causes both osmotic and ionic effects. Different factors including Na^+/H^+ antiport are known to be involved in maintaining ion homeostasis in plants exposed to salt stress (Zhang *et al.*, 2001). Arabidopsis salt overly sensitive (*sos*) mutant 1 was shown to encode a plasma membrane Na^+/H^+

antiporter with sequence similarity to plasma membrane Na^+/H^+ antiporters from bacteria and fungi (Shi *et al.*, 2000). A vacuolar Na^+/H^+ antiporter gene (AtNHX1) from Arabidopsis was transferred into *Brassica napus* (Zhang *et al.*, 2001) and *Lycopersicon esculentum* (Zhang and Blumwald, 2001). The transgenic Brassica and tomato plants were able to grow, flower, and produce seeds in the presence of 200 mM NaCl. Another gene that encodes a vacuolar Na^+/H^+ antiporter was isolated from *Atriplex gmelini* (AgNHX1) and transferred to *O. sativa.* The transgenic rice plants survived up to 300 mM NaCl for 3 days and conferred significant improvement in salt stress tolerance (Ohta *et al.*, 2002). According to Horie *et al.* (2001) plant growth under salt stress conditions requires the maintenance of a high cytosolic K^+/Na^+ concentration ratio. Therefore, relevant ion transporters are the likely candidates to be tested in transgenic plants. Rus *et al.* (2001) found that a high affinity potassium transporter (HKT1) from *A. thaliana* functions as a selective Na^+ transporter and also mediates K^+ transport. A HKT gene was introduced into wheat in sense and antinsense orientation and the transgenic lines showed enhanced growth in the presence of 200 mM NaCl. Na^+: K^+ ratios were reduced in salt-stressed transgenic tissue when compared to control (Laurie *et al.*, 2002).

The regulation of ion homeostasis under salt stress has been extensively studied by using salt overly sensitive (SOS) mutants of Arabidopsis (for review see Zhu, 2003; Gong *et al*., 2004). There is substantial evidence that SOS pathway involving several SOS genes plays a key role in regulation of ion transporter expression. For instance, one of these genes, SOS3 encodes a novel EF-hand Ca^{2+} sensor (Liu and Zhu, 1998) and their associated SOS2 gene (Ser/ Thr protein kinase) interacts physically. The SOS3-SOS2 complex mediates expression of Na^+/ H^+ antiporter (SOS1) gene. The transporter in turn, maintain low Na^+ and high K^+ levels in the cytoplasm during salt stress. Ion homeostasis during salt stress is also dependent on signaling via the calcium- and calmodulin-dependent protein phosphatase calcineurin (Liu and Zhu, 1998; Pardo *et al.*, 1998). A truncated form of the catalytic subunit, and the regulatory subunit of yeast calmodulin-dependent protein phosphatase calcineurin (CaN) were coexpressed in transgenic tobacco plants to activate the phosphatase *in vivo*. Like in yeast, transgenic tobacco plants expressing activated CaN exhibited substantial NaCl tolerance by regulating the calmodulin-dependent CaN signal pathway (Pardo *et al.,* 1998). A rice gene encoding Ca^{2+}-dependent protein kinase (CDPK) was overexpressed in transgenic rice plants and shown to enhance induction of stress-responsive genes in response to salt and drought stress (Saijo *et al.*, 2000). The authors concluded that CDPK is a positive regulator commonly involved in tolerance to both salt and drought stress in transgenic rice plants overexpressing CDPK.

2.2.5. Engineering of Regulatory Genes

Transcription factors that control gene expression under stress conditions play an important role during stress adaptation. Here, we describe mainly two major families, ERF and bZIP proteins. The ERF (ethylene-responsive-element-binding factor) family is a large group of transcription factors containing a C-repeat dehydration-responsive element (DRE), which is unique to plant systems. DRE elements play an important role in the regulation of gene expression in response to various stresses (Yamaguchi-Shinozaki and Shinozaki, 1994). It was found that the transcription factor DREB1A specifically interacts with the DRE motif and induces the expression of stress tolerance genes. Overexpression of the DREB1A gene under control of the stress inducible rd29A promoter resulted in a better growth of the transgenic plants in comparison to those transformed with a CaMV35S promoter-DREB1A construct (Kasuga *et al.*, 1999). This work indicates the importance of stress-inducible promoters for generation of transgenic plants. When a DREB1 gene was introduced into wheat, transformants survived a short but intensive water stress (Pellegrineschi *et al.*, 2002).

Another large family of transcription factors in plants are the bZIPs, among which one subclass ABRE/ABF (ABA-responsive-element-binding protein/ ABRE binding factor) is a well-studied example, which is linked to stress signaling, including salt, drought and UV light stresses. Different abiotic stresses and ABA induce ABRE/ABF expression and ABA triggers ABRE phosphorylation. This phosphorylation is necessary to induce downstream genes, which could occur on the casein kinase II phosphorylation sites. Therefore, ABA and different abiotic stresses induce both transcriptional and post-translational regulation of several bZIP transcription factors (Jakoby *et al.*, 2002). In rice, a cDNA encoded bZIP protein (OSBZ8) was shown to bind G-box-like elements including ABREs (Nakagawa *et al.,* 1996). Constitutive overexpression of ABRE binding factors (ABF3 or ABF4) led to altered expression of ABA/stress-regulated genes and in turn reduced transpiration and enhanced drought tolerance (Kang *et al.*, 2002).

2.2.6. The Future of Developing Stress-Tolerant Transgenic Cereals

So far, relatively few transgenic cereals (mostly rice and wheat) have been developed, each containing usually only one stress response gene (Table 4). However, this may not be enough to serve the purpose since many genes and components control stress tolerance. Furthermore, plants have to be tolerant and at the same time have to produce high yields (Pental, 2003).

Thus, only introduction of multiple genes into a single plant might yield higher tolerance without negative effects on other important agronomic parameters. Further, the expression levels of the transgenes have to be increased and expression must be controlled preferentially by using stress-induced promoters. Moreover, no scientific reports have been published yet on extensive field tests, which can give clear results about tolerance levels. Positive transgenic lines tested in this way, could be used for further development of stress tolerant varieties through breeding.

2.3. AB-QTL Analysis and Genetical Genomics

Dense molecular marker maps are now available for a number of cereals like wheat (reviewed in Gupta *et al.*, 1999), barley (reviewed in Varshney *et al.*, 2004; Forster *et al.,* 1997), maize (http://www.agron.missouri. edu/maps.html), rice (http://rgp.dna.affrc.go.jp/ Publicdata.html; Kikuchi *et al.*, 2003) and sorghum (http://sorghumgenome. tamu.edu). In all major crops including cereals, these molecular genetic maps and the available molecular markers were extensively used for identification of genes or QTLs for a variety of traits. In particular, the molecular markers linked with QTLs that confer tolerance to abiotic stresses have a great potential for their use in marker-assisted selection (MAS) in breeding programmes aimed at crop improvement. This aspect has been discussed in detail in Chapter 9 by Tuberosa and Salvi, and in Chapter 10 by Koebner.

Advanced-backcross QTL analysis (ABQA) for simultaneous discovery and transfer of QTLs from a wild species to a crop variety, proposed earlier by Tanksley and Nelson (1996), may also be useful for the development of tolerance to abiotic stresses in cereals. In this approach, a wild species is backcrossed to a superior cultivar, and during backcrosses, the transfer of desirable gene/QTL is monitored by employing molecular markers. The segregating BC2F2 or BC2F3 population is then used not only for recording data on the trait of interest, but also for genotyping it using polymorphic molecular markers. This data is then used for QTL analysis, leading to simultaneous discovery of QTLs, while transferring these QTLs by conventional backcrossing. However, for transfer of tolerance to abiotic stresses, this ABQA approach has yet to be utilized in cereals, although for other traits like yield and yield components it has already been successfully used in tomato (Tanksley *et al.*, 1996), rice (Xiao *et al.*, 1998; Moncada *et al.,* 2001), wheat (Huang *et al.*, 2003) and barley (Pillen *et al.*, 2003; Talamè *et al.,* 2003).

Recently, a new approach, called 'genetical genomics' has also been proposed, where QTL mapping is combined with expression profiling of individual genes in a segregating (mapping) population (Jansen and Nap, 2001). In this approach, total mRNA or cDNA of the organ/tissue from each individual of a mapping population is hybridized onto a microarray carrying a high number of cDNA fragments representing the species/tissue of interest and quantitative data are recorded reflecting the level of expression of each gene on the filter. Under the presumption, that every gene showing transcriptional regulation is mapped within the genome of the species of interest, the expression data can be subjected to QTL analysis, thus making it possible to identify the so-called 'ExpressQTLs' (eQTLs). The recently developed software tool 'Expressionview' for combined visualization of gene expression data and QTL mapping (Fischer *et al.*, 2003) will be very useful in this connection. Based on segregating populations, eQTL analysis identifies gene products influencing the quantitative trait (level of mRNA expression) in *cis* (mapping of the regulatled gene within the QTL) or *trans* (the gene is located outside the QTL). The latter gene product (second order effect) is of specific interest because more than one QTL can be connected to such a *trans*-acting factor (genes acting on the transcription of other genes) (Schadt *et al.*, 2003). The mapping of eQTLs allows multifactorial dissection of the expression profile of a given mRNA/cDNA, protein or metabolite into its underlying genetic components, and also allows locating these components on the genetic map (see Jansen and Nap, 2001; Jansen, 2003). Eventually, for each gene or gene product analyzed in the segregating population (by using expression profiling methodology), eQTL analysis will underline the regions of the genome influencing its expression. For crops like rice, where sequence of the whole genome is available, the annotation of those genomic regions will be helpful for the identification of the genes and their regulatory sequences involved in the expression of an individual trait.

Recently, in mouse, humans and maize, 'genetical genomics' approach has been used for a genome-wide study of the genetics of variation in expression of individual genes/QTLs for specific triats (Schadt *et al.* 2003). For instance, eQTLs were identified that influenced expression of about 10% of genes, differentially expressed in two typical inbred lines of maize- a stiff salk syntethic type and a Lancaster type. Gene-gene interactions similar to epistatsis were also noticed, and the interacting eQTLs were sometimes found on different chromosomes. This approach provides a powerful source to implicate genes as being involved either in the same or related transduction pathways involved in the expression of individual genes. Although the 'genetical genomics' approach is still in its infancy, efforts are underway in this direction in some plant species like tomato (Bai

et al. 2003), eucalyptus (Kirst *et al.* 2003) and barley (Potokina *et al.* 2003). We believe that availability of large EST collections for genome-wide expression profiling (see section 2.1) and analytical tools for molecular marker analysis in different cereals will accelerate the use of this approach in cereals for different agronomic traits including abiotic stress tolerance.

3. SUMMARY AND OUTLOOK

In the last two decades, biochemical pathways that are involved in conferring tolerance against abiotic stresses were studied in great detail, and genes involved in different steps of these biosynthetic pathways were isolated and characterized. Based on these studies, it is now known that accumulation of osmolytes, scavenging of reactive oxygen species, higher expression of chaperones and a control over sodium uptake might bring about at least partial tolerance against drought, salinity and cold. Nevertheless, we are still far from having a complete understanding of the molecular basis and regulatory mechanisms involved in conferring abiotic stress tolerance/susceptibility. A dissection of the complexity of tolerance against salinity, drought and temperature stress in (tolerant) crop plants will be possible in future through a variety of approaches. These approaches include whole genomic sequencing, high-throughput transcript profiling and discovery of gene functions. This will facilitate identification of candidate genes conferring tolerance, development of transgenic crops with higher tolerance, and selection of markers for marker-assisted selection/breeding. New approaches like 'genetical genomics' offer great promise to identify genes or genomic regions (QTLs) that are involved in conferring tolerance to abiotic stresses.

4. ACKNOWLEDGEMENTS

We are gratefully indebted to Prof. U. Wobus, Director, IPK, Gatersleben (Germany) and Prof. P.K. Gupta, Ch Charan Singh University, Meerut (India) for helpful comments and improving the quality of manuscript. This work was supported by IPK internal funds.

5. REFERENCES

Aarts M.G.M., Fiers M.W.E.J. What drives plant stress genes. Trends Plant Sci 2003; 8: 99-102

Abe H., Urao T., Ito T., Seki M., Shinozaki K., Yamaguchi-Shinozaki K. Arabidopsis AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signaling. Plant Cell 2003; 15: 63-78

Abebe T., Guenzi A.C., Martin B, Cushman J.C. Tolerance of mannitol-accumulating transgenic wheat to water stress and salinity. Plant Physiol 2003; 131: 1748-1755

Aguan K., Sugawara K., Suzuki Kusano T. Isolation of genes for low temperature-induced proteins in rice by a simple subtractive method. Plant Cell Physiol 1991; 32: 1258-1289

Apse M.P., Blumwald E. Engineering salt tolerance in plants. Curr Opin Biotech 2002; 13: 146-150

Bai Y. Huang C., van der Hulst R., Meijer-Dekens F., Li C., Bonnema G., Lindhout P. Qualitative and quantitative resistance to the tomato powdery mildew *Odium lycopersici* studied by genetical genomics. In: *Plant & Animal Genomes XI Conference*, January 11-15, 2003, San Diego, USA, Abstract no P524

Baima S., Possenti M., Matteucci A., Wisman E., Altamura M.M., Ruberti I., Morelli G. The Arabidopsis ATHB-8 HD-zip protein acts as a differentiation-promoting transcription factor of the vascular meristems. Plant Physiol 2001; 126: 643-655

Bartels D., Nelson D. Approaches to improve stress tolerance-using molecular genetics. Plant Cell Environ 1994; 17: 659-667

Bohnert H.J., Ayoubi P., Borchert C., Bressan R.A., Burnap R.L., Cushman J.C., Cushman M.A., Deyholos M., Fischer R., Galbraith D.W., *et al.*, A genomics approach towards salt stress tolerance. Plant Physiol Biochem 2001; 39: 295-311

Bray E.A. Molecular response to water deficit. Plant Physiol 1993; 103: 1035-1040

Bruce W.B., Edmeades G.O., Barker T.C. Molecular and physiological approaches to maize improvement for drought tolerance. Jour Exp Bot 2002; 53: 13-25

Capell T., Escobar C., Liu H., Burtin D., Lepri O., Cristou P. Overexpression of oat arginine decarboxylase cDNA in transgenic rice (*Oryza sativa* L.) affects normal development pattern *in vitro* and results in putrescine accumulation in transgenic plants. Theor Appl Genet 1998; 97: 246-254

Cattivelli L., Baldi P., Crosatti C., Di Fonzo N., Faccioli P., Grossi M., Mastrangelo A.M., Pecchioni N., Stanca A.M. Chromosome regions and stress-related sequences involved in resistance to abiotic stress in Triticeae. Plant Mol Biol 2002; 48: 649-665

Cattivelli L., Bartels D. Molecular cloning and characterization of cold-regulated genes in barley. Plant Physiol 1990; 93: 1504-1510

Chauvin L.P., Houde M., Sarahan F. A leaf-specific gene stimulated by light during wheat acclimation to low temperature. Plant Mol Biol 1993; 23: 255-265

Chen W.Q., Provart N.J., Glazebrook J., Katagiri F., Chang H.S., Eulgem T., Mauch F., Luan S., Zou G.Z., Whitham S.A., *et al.*, Expression profile matrix of Arabidopsis transcription factor genes suggests their putative functions in response to environmental stresses. Plant Cell 2002; 14: 559-574

Choi H.I., Hong J.H., Ha J.O., Kang J.Y., Kim S.Y. ABFs, a family of ABA-responsive element binding factors. Jour Biol Chem 2000; 275: 1723-1730

Choi D.W., Zhu B., Close T.J. The barley (*Hordeum vulgare L.*) dehydrin multigene family: sequences, allele types, chromosome assignments, and expression characteristics of 11 Dhn genes of cv Dicktoo. Theor Appl Genet 1999; 98: 1234-1247

Close T.J. Dehydrins: A commonality in the response of plants to dehydration and low temperature. Physiol Plant 1997; 100: 291-296

Csonka L.N., Hanson A.D. Prokaryotic osmoregulation: genetics and physiology. Annu Rev Microbiol 1991; 45: 569-606

Cushman J.C., Bohnert H.J. Genomic approaches to plant stress tolerance. Curr Opin Plant Biol 2000; 3: 117-124

Datta K., Velazhahan R., Oliva N., Ona I., Mew T., Khush G.S., Muthukrishnan S., Datta S.K. Over-expression of the cloned rice thaumatin-like protein (PR-5) gene in

transgenic rice plants enhances environmental friendly resistance to *Rhizoctonia solani* causing sheath blight disease. Theor Appl Genet 1999; 98: 1138-1145

Deak M., Hovarth G.V., Davletova S., Torok K., Sass L., Vass I., Barna B., Kiraly Z., Dudits D. Plants ectopically expressing the iron-binding protein ferritin, are tolerant to oxidative damages and pathogens. Nature Biotech 1999; 17: 192-196

Delauney A.J., Hu C.A.A., Kishor P.B.K., Verma D.P.S. Cloning of ornithine delta-aminotransferase cDNA from *Vigna-aconitifolia* by transcomplementation in Escherichia-coli and regulation of proline biosynthesis. Jour Biol Chem 1993; 268: 18673-18678

Dunn M.A., Brown K., Lightowers R.L., Hughes M.A. A low-temperature-responsive gene from barley encodes a protein with single stranded nucleic acid binding activity which is phosphorylated in vitro. Plant Mol Biol 1996; 30: 947-959

Dunn M.A., Morris A., Jack P.L., Hughes MA. A low-temperature-responsive translation elongation factor 1α from barley (*Hordeum vulgare L*). Plant Mol Biol 1993; 23: 221-225

Dure L. III. The LEA proteins of higher plants. In: Verma D.P.S. (ed*.) Control of Plant Gene Expression*, CRC Press, Boca Raton, USA, 1993; pp 325-335

Fischer G., Ibrahim S.M., Brockmann G.A., Pahnke J., Bartocci E., Thiesen H.-J., Serrano-Fernandez P., Möller S. Expressionview: visualization of quantitative trait loci and gene-expression data in Ensembl. Genome Biol 2003; 4: R477

Forster B.P., Russell J.R., Ellis R.P., Handley L.L., Robinson D., Hackett C.A., Nevo E., Waugh R., Gordon D.C., Keith R., *et al.*, Locating genotypes and genes for abiotic stress tolerance in barley: a strategy using maps, markers and the wild species. New Phytol 1997; 137: 141-147

Garg A.K., Kim J.-K., Owens T.G., Ranwala A.P., Choi Y.D., Kochian L.V., Wu R.J. Trehalose accumulation in rice plants confers high tolerance levels to different abiotic stresses. Proc Natl Acad Sci USA 2002; 99: 15898-15903

Gale M.D., Devos K.M. Comparative genetics in the grasses. Proc Natl Acad Sci USA 1998; 95: 1971-1974

Gong D., Guo Y., Schumaker K.S., Zhu J.K. The SOS3 family of calcium sensors and SOS2 family of protein kinases in Arabidopsis. Plant Physiol 2004; 134: 919-926

Grossi M., Giorni E., Rizza F., Stanca A.M, Cattivelli L. Wild and cultivated barleys show differences in the expression pattern of a cold-regulated gene family under different light and temperature conditions. Plant Mol Biol 1998; 38: 1061-1069

Grover A., Aggarwal P.K., Kapoor A., Katiyar-Agarwal S., Agarwal M., Chandramouli A. Addressing abiotic stresses in agriculture through transgenic technology. Curr Sci 2003; 84: 355-367

Guo B.H., Zhang Y.M., Li H.J., Du L.Q., Li Y.X., Zhang J.S., Chen S.Y., Zhu Z.Q. Transformation of wheat with a gene encoding for the betaine aldehyde dehydrogenase (BADH). Acta Bot Sin 2000; 42: 279-283

Gupta P.K., Varshney R.K., Sharma P.C., Ramesh B. Molecular markers and their application in wheat breeding. Plant Breed 1999; 118: 369-390

Hare P.D., Cress W.A. Metabolic implications of stress-induced proline accumulation in plants. Plant Growth Reg 1997; 21: 79-102

Hasegawa P.M., Bressan R.A., Zhu J.K., Bohnert H.J. Plant cellular and molecular responses to high salinity. Ann Rev Plant Physiol Plant Mol Biol 2000; 51: 463-499

Hirano H.Y., Sano Y. Enhancement of *Wx* gene expression and the accumulation of amylose in response to cool temperatures during seed development in rice. Plant Cell Physiol 1998; 39: 807-812

Hobo T., Asada M., Kowyama Y., Hattori T. ACGT-containing abscisic acid response element (ABRE) and coupling element 3 (CE3) are functionally equivalent. Plant Jour 1999; 19: 679-689

Hong B., Barg R., Ho T.H.D. Developmental and organ-specific expression of an ABA- and stress- induced protein in barley. Plant Mol Biol 1992; 18: 663-674

Hong Z.L., Lakkineni K., Zhang Z.M., Verma D.P.S. Removal of feedback inhibition of Delta(1)-pyrroline-5-carboxylate synthetase results in increased proline accumulation and protection of plants from osmotic stress. Plant Physiol 2000; 122: 1129-1136

Horie T., Yoshida K., Nakayama H., Yamada K., Oiki S., Shinmyo A. Two types of HKT transporters with different properties of Na+ and K+ transport in *Oryza sativa*. Plant Jour 2001; 27: 129-138

Hoshida H., Tanaka Y., Hibino T., Hayashi Y., Tanaka A., Takabe T. Enhanced tolerance to salt stress in transgenic rice that overexpresses chloroplast glutamine synthetase. Plant Mol Biol 2000; 43: 103-111

Huang X.Q., Coster H., Ganal M.W., Roder M.S. Advanced backcross QTL analysis for the identification of quantitative trait loci alleles from wild relatives of wheat (*Triticum aestivum* L.). Theor Appl Genet 2003; 106: 1379-1389

Iba K. Acclimative response to temperature stress in higher plants: Approaches of gene engineering for temperature tolerance. Annu Rev Plant Biol 2002; 53: 225-245

Igarashi Y., Yoshiba Y., Sanada Y., Yamaguchi-Shinozaki K., Wada K., Shinozaki K. Characterization of the gene for p5cs and correlation between the expression of the gene and salt tolerance in *Oryza sativa* L. Plant Mol Biol 1997; 33: 857-865

Igarashi Y., Yoshiba Y., Takeshita T., Nomura S., Otomo J., Yamaguchi-Shinozaki K., Shinozaki K. Molecular cloning and characterization of a cDNA encoding proline transporter in rice. Plant Cell Physiol 2000; 41: 750-756

Ingram J., Bartels D. The molecular basis of dehydration tolerance in plants. Annu Rev Plant Physiol Plant Mol Biol 1996; 47: 377-403

Ishitani M., Xiong L., Stevenson B., Zhu J.K. Genetic analysis of osmotic and cold stress signal transduction in Arabidopsis: interaction and convergence of abscisic acid-dependent and abscisic acid-independent pathways. Plant Cell 1997; 9: 1935-1949

Ismail A.M., Hall A.E., Close T.J. Allelic variation of a dehydrin gene cosegregates with chilling tolerance during seedling emergence. Proc Nat Acad Sci USA 1999; 96: 13566-13570

Iyer S., Caplan A. Products of proline catabolism can induce osmotically regulated genes in rice. Plant Physiol 1998; 116: 203-211

Jakoby M., Weisshaar B., Droge-Laser W., Vicente-Carbajosa J., Tiedemann J., Kroj T., Parcy F. bZIP transcription factors in Arabidopsis. Trends Plant Sci 2002; 7: 106-111

Jang I.C., Oh S.J., Seo J.S., Choi W.B., Song S.I., Kim C.H., Kim Y.S., Seo H.S., Do Choi Y., Nahm B.H., Kim J.K. Expression of a bifunctional fusion of the Escherichia coli genes for trehalose-6-phosphate synthase and trehalose-6-phosphate phosphatase in transgenic rice plants increases trehalose accumulation and abiotic stress tolerance without stunting growth. Plant Physiol 2003; 131: 516-524

Jansen R.C. Studying complex biological systems using multifactorial perturbation. Nature Rev Genet 2003; 4:145-151

Jansen R.C., Nap J.-P. Genetical genomics: the added value from segregation. Trends Genet 2001, 17:388-391

Kaasen I., McDougall J., Strom A.R. Analysis of the OTSBA operon for osmoregulatory trehalose synthesis in *Escherichia coli* and homology of the OTSA and OTSB proteins to the yeast trehalose-6-phosphate synthase phosphatase complex. Gene 1994; 145: 9-15

Kaminaka H., Morita S., Yokoi H., Masumura T., Tanaka K. Molecular cloning and characterization of a cDNA for plastidic copper/zinc-superoxide dismutase in rice (*Oryza sativa* L). Plant Cell Physiol 1997; 38: 65-69

Kang J.Y., Choi H.I., Im M.Y., Kim S.Y. Arabidopsis basic leucine zipper proteins that mediate stress-responsive abscisic acid signaling. Plant Cell 2002; 14: 343-357

Kasuga M., Liu Q., Miuras S., Yamaguchi-Shinozaki K., Shinozaki K. Improving plant drought, salt and freezing tolerance by gene transfer of a single stress-inducible transcription factor. Nature Biotech 1999; 17: 287-291

Katiyar-Agarwal S., Agarwal M., Grover A. Heat-tolerant basmati rice engineered by over-expression of hsp101. Plant Mol Biol 2003; 51: 677-686

Kawasaki S., Borchert C., Deyholos M., Wang H., Brazille S., Kawai K., Galbraith D., Bohnert H.J. Gene expression profiles during the initial phase of salt stress in rice. Plant Cell 2001; 13: 889-905

Kikuchi S., Satoh K., Nagata T., Kawagashira N., Doi K., Kishimoto N., Yazaki J., Ishikawa M., Yamada H., Ooka H., *et al.*, Collection, mapping, and annotation of over 28,000 cDNA clones from *japonica* rice. Science. 2003; 301: 376-379

Kikuchi S., Yazaki J., Kishimoto N., Ishikawa M., Kojima K., Takahiro N., Shimbo K., Fujii F., Ohata T., Shimatani Z., *et al.*, Rice functional genomics via cDNA microarray: Systems for the microarray analysis and the expression profiles of stress responsible gene expression. JIRCAS Report 2002; 93-98

Kirst M., Myburg A.A., Sederoff R.R. Genetical genomics of *Eucalyptus*: combining expression profiling and genetic segregation analysis. In: *Plant & Animal Genomes XI Conference*, January 11-15, 2003, San Diego, USA, Abstract no W151

Kishitani S., Watanabe K., Yasuda S., Arakawa K., Takabe T. Accumulation of glycine betaine during cold acclimation and freezing tolerance in leaves of winter and spring barley plants. Plant Cell Environ 1994; 17: 89-95

Kishor P.B., Hong Z., Miao G.H., Hu C.A., Verma D.P.S. Overexpression of D1-pyrroline-5-carboxylase synthetase increases proline production and confers osmotolerance in transgenic plants. Plant Physiol 1995; 108: 1387-1394

Knight H., Knight M.R. Abiotic stress signalling pathways: specificity and cross-talk. Trends Plant Sci 2001, 6: 262-267

Kreps J.A., Wu Y.J., Chang H.S., Zhu T., Wang X., Harper J.F. Transcriptome changes for Arabidopsis in response to salt, osmotic, and cold stress. Plant Physiol 2002; 130: 2129-2141

Labhilili M., Joudrier P., Gautier M.F. Characterization of cDNA encoding *Triticum durum* dehydrins and their expression patterns in cultivars that differ in drought tolerance. Plant Sci 1995; 112: 219-230

Laurie S., Feeney K.A., Maathuis F.J.M., Heard .PJ., Brown S.J., Leigh R.A. A role for HKT1 in sodium uptake by wheat roots. Plant Jour 2002; 32: 139-149

Lee J., Williams M.E., Tingey S.V., Rafalski J.A. DNA array profiling of gene expression changes during maize embryo development. Funct Integr Genom 2002; 2: 13-27

Liu Q., Kasuga M., Sakuma Y., Abe H., Miura S., Yamaguchi-Shinozaki K., Shinozaki K. Two transcription factors, DREB1 and DREB2, with an EREBP/AP2 DNA binding domain separate two cellular signal transduction pathways in drought- and low-temperature-responsive gene expression, respectively, in Arabidopsis. Plant Cell 1998; 10: 1391-1406

Liu J., Zhu J.K. A calcium sensor homolog required for plant salt tolerance. Science 1998; 280: 1943-1945

Madan S, Nainawatee HS, Jain RK, Chowdhury J.B. Proline and proline metabolising enzymes in in vitro selected NaCl-tolerant *Brassica-juncea L*. under salt stress. Ann Bot 1995; 76: 51-57

Malan C., Greyling M.M., Gressel J. Correlation between Cu/Zn superoxide dismutase and glutathione reductase, and environmental and xenobiotic stress tolerance in maize inbreds. Plant Sci 1990; 69: 157-166

Maqbool S.B., Zhong H., El-Maghraby Y., Ahmad A., Chai B., Wang W., Sabzikar R., Sticklen M.B. Competence of oat (*Avena sativa* L.) shoot apical meristems for integrative transformation, inherited expression, and osmotic tolerance of transgenic lines containing *hva1* Theor Appl Genet 2002; 105: 201-208

Matsumura T., Tabayashi N., Kamagata Y., Souma C., Saruyama H. Wheat catalase expressed in transgenic rice can improve tolerance against low temperature stress. Physiol Plant 2002; 116: 317-327

McKersie B.D., Bowley S.R., Jones K.S. Winter survival of transgenic alfalfa overexpressing superoxide dismustase. Plant Physiol 1999; 119: 839-847

Minhas D., Grover A. Towards developing transgenic rice plants tolerant to flooding stress. Proc Ind Natl Sci Acad 1999; B65: 33-50

Mohanty A., Kathuria H., Ferjani A., Sakamoto A., Mohanty P., Murata N., Tyagi A.K. Transgenics of an elite indica rice variety Pusa Basmati 1 harbouring the codA gene are highly tolerant to salt stress. Theor Appl Genet 2002; 106: 51-57

Moncada P., Martinez C., Borrero J., Gauch H.J., Guimaraes E., Tohme J., McCouch S. Quantitative trait loci for yield and yield components in an *Oryza sativa* × *Oryza rufipogon* BC_2F_2 population evaluated in an upland environment. Theor Appl Genet 2001; 102: 41-52

Mundy J., Yamaguchi-Shinozaki K., Chua N.H. Nuclear proteins bind conserved elements in the abscisic acid responsive promoter of a rice RAB gene. Proc Natl Acad Sci USA 1990; 87: 1406-1410

Nakagawa H., Ohmiya K., Hattori T. A rice bZIP protein, designated OSBZ8, is rapidly induced by abscisic acid. Plant Jour 1996; 9: 212-227

Nantel A., Quatrano R.S. Characterization of three rice basic/leucine zipper factors, including two inhibitors of EmBP-1 DNA binding activity. Jour Biol Chem 1996; 271: 31296-31305

Ohta M., Hayashi Y., Nakashima A., Tsunetomi N., Hamada A., Tanaka A., Nakamura T., Hayakawa T. Salt tolerance of rice is conffered by introduction of a Na+/H+ antiporter gene from Atriplex gmelini. Plant Cell Physiol 2002; 43: S56-S56

Oullet F., Vazquez-Tello A., Sarhan F. The wheat *wcs120* promoter is cold inducible in both monocotyledons and dicotyledoos species. FEBS Lett 1998; 423: 324-328

Ozturk Z.N., Talame V., Deyholos M., Michalowski C.B., Galbraith D.W., Gozukirmizi N., Tuberosa R., Bohnert H.J. Monitoring large-scale changes in transcript abundance in drought- and salt-stressed barley. Plant Mol Biol 2002; 48: 551-573

Papageorgiou G.C., Murata N. The unusually strong stabilizing effects of glycine betaine on the structure and function of the oxygen-evolving photosystem II complex. Photosynth Res 1995; 44: 243-252

Pardo J.M., Reddy M.P., Yang S., Maggio A., Huh G.H., Matsumoto T., Coca M.A., Paino-D' Urazo M., Koiwa H. Stress signaling through Ca^+ calmodulin-dependent protein phosphatase calcineurin mediates salt adaptation in plants. Proc Natl Acad Sci USA 1998; 95: 9681-9686

Pareek A., Singla S.L., Grover A. Immunological evidence for accumulation of two high-molecular weight (104 and 90 kDa) HSPs in response to different stresses in rice and in response to high temperature stress in diverse plant genera. Plant Mol Biol 1995; 29: 293-301

Pearce R.S., Houlston C.E., Atherton K.M., Rixon J.E., Harrison P., Hughes M.A., Dunn M.A. Localization of expression of three cold-induced genes, *blt101*, *blt4.9* and *blt14*, in different tissues of the crown and developing leaves of cold-acclimated cultivated barley. Plant Physiol 1998; 117: 787-795

Pellegrineshi A., Ribaut J.M., Trethowan R., Yamaguchi-Shinozaki K., Hoisington D. Progress in the genetic engineering of wheat for water limited conditions. JIRCAS Report 2002; pp 55-60

Pental D. Transgenics for productive and sustainable agriculture: Some considerations for the development of a policy framework. Curr Sci 2003; 84: 413-424

Pillen K., Zacharias A., Léon J. Advanced backcross QTL analysis in barley (*Hordeum vulgare* L.) Theor Appl Genet 2003; 1253-1259

Pla M., Goday A., Vilardell J., Guiltinan M.J., Marcotte W.R., Niogret M.F., Quatrano R.S., Pages M. The cis-regulatory elements CCACGTGG is involved in ABA and water stress responses of the maize gene *rab28*. Plant Mol Biol 1993; 21: 259-266

Potokina E., Zhang H., Prasad M., Stein N., Graner A**.** Malting quality in barley: a case study for genetical genomics. In: *11th Molecular Markers Symposium of GPZ*, September 16-17, 2003, Gatersleben, Germany, Abstract no 150

Prasad T.K., Anderson M.D., Martin B.A., Stewart C.R. Evidence for chilling-induced oxidative stress in maize seedlings and a regulatory role for hydrogen-peroxide. Plant Cell 1994; 6: 65-74

Quarrie S.A., Stojanovic J., Pekie S. Improwing drought resistance in small-grained cereals: A case study, Progress and prospects. Plant Growth Regul 1999; 29: 1-21

Quimio C.A., Torrizo L.B., Stter T.L., Ellis M., Grover A., Abrigo E.M., Oliva N.P., Ella E.S, Carpena A.L, Ito O., *et al.* Enhancement of submergence tolerance in transgenic rice overproducing pyruvate decarboxylase. Jour Plant Physiol 2000; 156: 516-521

Reddy A.R., Ramakrishna W., Sekhar A.C., Ithal N., Babu P.R., Bonaldo M.F., Soares M.B., Bennetzen J.L. Novel genes are enriched in normalized cDNA libraries from drought-stressed seedlings of rice (*Oryza sativa* L. subsp. *indica* cv. Nagina 22). Genome 2002; 45: 204-211

Rohila J.S., Jain R.K., Wu R. Genetic improvement of Basmati rice for salt and drought tolerance by regulated expression of a barley *Hva1* cDNA. Plant Sci 2002; 163: 525-532

Roosens N.H.C.J., Thu T.T., Iskandar H.M., Jacobs M. Isolation of the ornithine δ-aminotransferase cDNA and effect of salt stress on its expression in *Arabidopsis thaliana*. Plant Physiol 1998; 117:263-271

Roxas V.P., Lodhi S.A., Garrett D.K., Mahn J.R., Allen R.D. Stress tolerance in transgenic tobacco seedlings that overexpress glutathione S-transferace/glutathione peroxidae. Plant Cell Physiol 2000; 41: 1229-1234

Roy M., Wu R. Arginine decarboxylase transgene expression and analysis of environmental stress tolerance in transgenic rice. Plant Sci 2001; 160: 869-875

Rubio F., Schwarz M., Gassmann W., Schroeder J.I. Genetic selection of mutations in the high affinity K^+ transporter HKT1 that define functions of a loop site for reduced Na^+ permeability and increased Na^+ tolerance. Jour Biol Chem 1999; 274: 6839-6847

Rus A., Yokai S., Sharkhuu A., Reddy M., Lee B. Matasumoto T.K., Kolwa H., Zhu J.K., Bressan R.A., Hasegawa P.M. AtHKT1 is a salt tolerance determinant that controls Na^+ entry into plant roots. Proc Nat Acad Sci USA 2001; 98: 14150-14155

Russell B.L., Rathinasabapathi B., Hanson A.D. Osmotic stress induces expression of choline monooxygenase in sugar beet and amaranth. Plant Physiol 1998; 116: 859-865

Saijo Y., Hata S., Kyozuka J., Shimamoto K., Izui K. Over-expression of a single Ca^{2+}-dependent protein kinase confers both cold and salt/drought tolerance. Jour Exp Bot 2000; 51: 81-88

Sakamoto A., Murata N. Metabolic engineering of rice leading to biosynthesis of glycine betaine and tolerance to salt and cold. Plant Mol Biol 1998; 38: 1011-1019

Sakamoto A., Okumura T., Kaminaka H., Sumi K., Tanaka K. Structure and differential response to abscisic-acid of 2 promoters for the cytosolic copper zinc-superoxide dismutase genes, SodCC1 and SodCC2, in rice protoplasts. FEBS Lett 1995; 358: 62-66

Sawahel W.A., Hassan A.H. Generation of transgenic wheat plants producing high levels of osmoprotectant proline. Biotech Lett 2002; 24: 721-725

Schadt E.E., Monks S.A., Drake T.A., Lusis A.J., Che N., Colinayo V., Ruff T.G., Milligan S.B., Lamb J.R., Cavet G., *et al.* Genetics of gene expression surveyed in maize, mouse and man. Nature 2003; 422: 297-301

Seki M., Ishida J., Narusaka M., Fujita M., Nanjo T., Umezawa T., Kamiya A., Nakajima M., Enju A., Sakurai T., *et al.,* Monitoring the expression pattern of around 7,000

Arabidopsis genes under ABA treatments using a full-length cDNA microarray. Funct Integr Genom 2002a; 2: 282-291

Seki M., Narusaka M., Abe H., Kasuga M., Yamaguchi-Shinozaki K., Carninci P., Hayashizaki Y., Shinozaki K. Monitoring the expression pattern of 1300 Arabidopsis genes under drought and cold stresses by using a full-length cDNA microarray. Plant Cell 2001; 13: 61-72

Seki M., Narusaka M., Ishida J., Nanjo T., Fujita M., Oono Y., Kamiya A., Nakajima M., Enju A., Sakurai T., *et al.,* Monitoring the expression profiles of 7000 Arabidopsis genes under drought, cold and high-salinity stresses using a full-length cDNA microarray. Plant Jour 2002b; 31: 279-292

Shi H.Z., Ishitani M., Kim C.S., Zhu J.K. The *Arabidopsis thaliana* salt tolerance gene SOS1 encodes a putative Na+/H+ antiporter. Proc Nat Acad Sci USA 2000; 97: 6896-6901

Shinozaki K., Yamaguchi-Shinozaki K., Seki M. Regulatory network of gene expression in the drought and cold stress responses. Curr Opin Plant Biol 2003; 6: 410-417

Singla S.L., Grover A. Detection and quantitation of a rapidly accumulating and predominant 104 kD heat shock polypeptide in rice. Plant Sci 1994; 97:23-30

Sivamani E., Bahieldin I.A., Wraith J.M., Al-Niemi T., Dyer W.E., Ho T.D., Wu R. Improved biomass productivity and water use efficiency under water deficit conditions in transgenic wheat constitutively expressing the barley *HVA1* gene. Plant Sci 2000; 155: 1-9

Smeekens S. Sugar-induced signal transduction in plants. Ann Rev Plant Physiol Plant Mol Biol 2000; 51: 49-81 2000

Sreenivasulu N., Altschmied L., Panitz R., Hahnel U., Michalek W., Weschke W., Wobus U. Identification of genes specifically expressed in maternal and filial tissues of barley caryopses: A cDNA array analysis. Mol Gen Genom 2002; 266: 758-767

Sreenivasulu N., Altschmied L., Radchuk V., Gubatz S., Wobus U., Weschke W. Transcript profiles and deduced changes of metabolic pathways in maternal and filial tissues of developing barley grains. Plant Jour 2004a; 37: 539-553

Sreenivasulu N., Grimm B., Wobus U., Weschke W. Differential response of antioxidant compounds to salinity stress in salt-tolerant and salt-sensitive seedlings of foxtail millet (*Setaria italica*). Physiol Plant 2000; 109: 435-442

Sreenivasulu N., Miranda M., Prakash H.S., Wobus U., Weschke W. Transcriptome changes in foxtail millet genotypes at high salinity: Identification and characterization of a PHGPX gene specifically up-regulated by NaCl in a salt-tolerant line. Jour Plant Physiol 2004b; 161: 467-477

Sreenivasulu N., Ramanjulu S., Ramachandra-kini K., Prakash H.S., Shekar Shetty K., Savithri H.S., Sudhakar C. Total peroxidase activity and peroxidase isoforms as modified by salt stress in two cultivars of foxtail millet with differential salt tolerance. Plant Sci 1999; 14: 1-9

Steponkus P.L., Uemura M., Joseph R.A., Gilmour S.J., Thomashow M.F. Mode of action of the COR15a gene on the freezing tolerance of *Arabidopsis thaliana*. Proc Nat Acad Sci USA 1998; 95: 14570-14575

Sutton F., Ding X., Kenefick D.G. Group 3 LEA gene *HVA1* regulation by cold acclimation and deacclimation in two barley cultivars with varying freeze resistance. Plant Physiol 1992; 99: 338-340

Takabe T., Hayashi Y., Tanaka A., Takabe T., Kishitani S. Evaluation of glycine betaine accumulation for stress tolerance in transgenic rice plants. In: *Proc Intern Workshop on Breeding and Biotechnology for Environmental Stress in Rice*, Hokkaido Agricultural Experiment Station, Sapporo, Japan, 1998, pp 63-68

Talamè V., Sanguineti M.C., Chiapparino E., Bahri H., Ben Salem M., Ellis R., Forster B.P., Rhouma S., Zoumarou W. Advanced backcross QTL analysis for the identification of agronomically valuable alleles in wild barley (*Hordeum spontaneum*). In: *In the Wake of*

the Double Helix from Green Revolution to the Gene Revolution, May 27-31, 2003, Bologna, Italy, p 123

Tanaka Y., Hibino T., Hayashi Y., Tanaka A., Kishitani S., Takabe T., Yokota S., Takabe T. Salt tolerance of transgenic rice overexpressing yeast mitochondrial Mn-SOD in chloroplasts. Plant Sci 1999; 148: 131-138

Tanksley S.D., Grandillo S., Fulton T.M., Zamir D., Eshed Y., Petiard V., Lopez J., BeckBunn T. Advanced backcross QTL analysis in a cross between an elite processing line of tomato and its wild relative *L. pimpinellifolium*. Theor Appl Genet 1996; 92: 213-224

Tanksley S.D., Nelson J.C. Advanced backcross QTL analysis: A method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. Theor Appl Genet 1996; 92: 191-203

Thomashow M.F. Plant cold acclimation: Freezing tolerance genes and regulatory mechanisms. Annu Rev Plant Physiol plant mol Biol 1999; 50: 571-599

Tilman D., Cassman K.G., Matson P.A., Naylor R., Polasky S. Agricultural sustainability and intensive production practices. Nature 2002; 418: 671-677.

Tzeng S.S., Yeh K.W., Chen Y.M., Lin C.Y. Two *Oryza sativa* genomic DNA clones encoding 16.9-kilodalton heat-shock proteins. Plant Physiol 1992; 99:1723-1725

Van Breusegem F., Slooten L., Stassart J.M., Botterman J., Moens T., Van Montagu M., Inze D. Effects of overproduction of tobacco MnSOD in maize chloroplasts on foliar tolerance to cold and oxidative stress. Jour Exp Bot 1999; 50: 71-78

Varshney R.K., Prasad M., Graner A. Molecular marker maps of barley: a resource for intra- and interspecific genomics. In: G. Wenzel and L. Horst (eds.) *Molecular Markers in Improvement of Agriculture and Forestry*, Spinger Verlag, Germany, 2004 (in press)

Vilardell J., Goday A., Freire M.A., Torent M., Martinez M.C., Torne J.M., Pages M. Gene sequence, development expression and protein phosphorylation of RAB-17 in maize. Plant Mol Biol 1990; 14: 423-432

Weigel D., Ahn J.H., Blazquez M.A., Borevitz J.O., Christensen S.K., Fankhauser C., Ferrandiz C., Kardailsky I., Malancharuvil E.J., Neff M.M., *et al.*, Activation tagging in Arabidopsis. Plant Physiol 2000; 122: 1003-1013

Wood A.J., Saneoka H., Rhodes D., Joly R.J., Goldsburough P.B. Betaine aldehyde dehydrogenase in *Sorghum*-molecular cloning and expression of two related genes. Plant Physiol 1996; 110: 1301-1308

Xiao J.H., Li J.M., Grandillo S., Ahn S.N., Yuan L.P., Tanksley S.D., McCouch S.R. Identification of trait-improving quantitative trait loci alleles from a wild rice relative, *Oryza rufipogon*. Genetics 1998; 150: 899-909

Xiong L.M., David L., Stevenson B., Zhu J.K.. High throughput screening of signal transduction mutants with luciferase imaging. Plant Mol Biol Rep 1999; 17: 159-170

Xu D., Duan X., Wang B., Hong B., Ho T.H., Wu R. Expression of a late embryogenesis abundant protein gene, HVA1, from barley confers tolerance to water deficit and salt stress in transgenic rice. Plant Physiol 1996; 110: 249-257

Yamaguchi-Shinozaki K., Shinozaki K. A novel *cis*-acting element in an *Arabidopsis* gene is involved in responsiveness to drought, low-temperature, or high-salt stress. Plant Cell 1994; 6: 251-264

Yokoi S., Higashi S., Kishitani S., Murata N., Toriyama K. Introduction of the cDNA for Arabidopsis glycerol-3-phosphate acyltransferase (GPAT) confers unsaturation of fatty acids and chilling tolerance of photosynthesis on rice. Mol Breed 1998; 4: 269-275

Yu L., Setter T.L. Comparative transcriptional profiling of placenta and endosperm in developing maize kernels in response to water deficit. Plant Physiol 2003; 131: 568-582

Zhang H.X., Blumwald E. Transgenic salt-tolerant tomato plants accumulate salt in foliage but not in fruit. Nature Biotech 2001; 19: 765-768

Zhang H.X., Hodson J.N., Williams J.P., Blumwald E. Engineering salt tolerant Brassica plants: characterization of yield and seed oil quality in transgenic plants with increased vacuolar sodium accumulation. Proc Nat Acad Sci USA 2001; 98: 12832-12836

Zhu J.K. Cell signaling under salt, water and cold stresses. Curr Opin Plant Biol 2001; 4: 401-406

Zhu J.K. Salt and drought stress signal transduction in plants. Ann Rev Plant Biol 2002; 53: 247-273

Zhu J.K. Regulation of ion homeostasis under salt stress. Curr Opin Plant Biol 2003; 6: 441-445

Zhu B.C., Su J., Chan M.C., Verma D.P.S., Fan Y.L., Wu R. Overexpression of a delta (1)-pyrroline-5-carboxylate synthase gene and analysis of tolerance to water and salt stress in transgenic rice. Plant Sci 1998; 139: 41-48

Chapter 17

THE ARABIDOPSIS GENOME AND ITS USE IN CEREAL GENOMICS

Klaus Mayer*, Stephen Rudd and Heiko Schoof

MIPS, Institute for Bioinformatics, National Research Centre for Environment and Health (GSF), Ingolstädter Landstrasse 1, 85764 Neuherberg, Germany
*Author for correspondence: Kmayer@gsf.de

1. INTRODUCTION

Arabidopsis thaliana (L.) Heynhold, popularly known as thale cress, is a self pollinated diploid weed of small size with short generation time, small chromosome number ($n = 5$) and small genome size (125 Mb), thus making it a model system for a variety of studies. The steep rise of Arabidopsis as a plant model system dates back to pioneering cytogenetic work (see Redei, 1992; Redei and Koncz, 1992 and references therein), although its genome was fully characterized and described only in the late 1980s (Pruitt and Meyerowitz, 1986). Highly saturated genetic and physical maps of the genome of this model plant species have been generated and utilized for gene tagging, QTL analysis, association analysis and map based cloning (Alonso *et al.*, 2003; Lukowitz *et al.,* 2000; Schmid *et al.*, 2003). Arabidopsis has also been found to be easily transformable, and has the unique honour of being the first flowering plant whose genome has been fully sequenced (TAGI, 2000). Consequently, at the genetic and molecular levels, this weed with no economic value has become the most extensively studied species of flowering plants. In the realm of genomics, Arabidopsis is now treated as a model flowering plant and deserves an honourable position with other pre-eminent model eukaryotes like yeast, fruitfly, *C. elegans* and the human, whose genomes have already been fully sequenced (Anon 1998; Adams *et al.*, 2000; Lander *et al.*, 2001; Venter *et al.*, 2001). The information generated from the study of Arabidopsis genome has been compared with and utilized for the study of the genomes of other flowering plants including

P.K. Gupta and R.K. Varshney (eds.), Cereal Genomics, 515-534.

those of tomato, mustard, *Capsella*, among the dicotyledons (Ku *et al.*, 2000; Rossberg *et al.*, 2001) and those of rice and maize among the monocotyledons (Yano *et al.*, 2000; Mayer *et al.*, 2001; Salse *et al.*, 2002; vanDodeweerd *et al.*, 2001; Liu *et al.*, 2001; Brendel *et al.*, 2002). More recently with the availability of several drafts of the rice genome sequences (Goff *et al.*, 2002; Yu *et al.*, 2002), the high quality sequences of rice chromosomes 1, 4 and 10 (Sasaki *et al.*, 2002; Feng *et al.*, 2002; The Rice Chromosome 10 Consortium, 2003), and increasing amounts of genomic survey sequences (GSS) from maize (Chandler and Brendel, 2002), it has become possible to compare and utilize the Arabidopsis genome for a study of cereal genomes, particularly those of rice and maize. In this chapter, an attempt will be made to review the present status and future prospects of utilizing data of the Arabidopsis genome (both genomic and cDNA) for an analysis of cereal genomes, which are themselves very closely related amongst each other (for details see Chapter 5 by Paterson in this book).

2. A BRIEF OVERVIEW OF THE ARABIDOPSIS GENOME

The completion of high quality sequence of 115 Mb of the 125 Mb genome of *Arabidopsis thaliana* marks a milestone in plant biology (TAGI, 2000). The sequenced part of the genome includes a majority of the euchromatic portion of the genome, with only few gaps. The repetitive sequences that are characteristic of cereals and other plant genomes are less prominent in Arabidopsis, and are mainly restricted to the centromeric regions, pericentromeric regions, and at a few cytogenetically characterized regions, the so called *knob* regions (Mayer *et al.* 1999; Fransz *et al.*, 2000; McCombie *et al.*, 2000). Therefore Arabidopsis has a densely organized gene rich genome, which is estimated to have 25,500 gene*s*, a number which has been revised slightly since the initial report (TAGI, 2000). Surprisingly this number exceeds the number of genes found in *Drosophila* and *C. elegans* genomes and is close to that for the human genome. For an estimate of the number of genes in a genome, knowledge of the extent of pseudogenes and that of alternative splicing is also necessary. Although the Arabidopsis genome has not been subjected to any detailed analysis of elements like pseudogenes, it has been compared with animal genomes for alternative splicing. Alternatively spliced genes appear to be more frequent in animal than in plant genomes (Brett *et al.*, 2002; Seki *et al.*, 2002; Kikuchi *et al.*, 2003).

An analysis of the Arabidopsis genome and its comparison with animal genomes resolved a number of gene families. Some of these gene families are absent in plants, others are restricted to plants and still others are found in both plants and animals (TAGI, 2000). Domains have been identified, some of which are over-represented, and some are absent in the Arabidopsis/plant gene inventory. Some of the prominent examples restricted to plants include Cytochrome P450, AP2/ethylene responsive element binding factor signature, MADS boxes or leucine rich repeat signatures (TAGI, 2000). Others are absent or rare in plants, but are largely expanded in the animal kingdom (TAGI, 2000).

Analysis of the Arabidopsis genome has also resolved preferential combinations of different functional domains in the same gene, a feature which is particularly insightful. It has been shown that some combinations of functional domains are specific for individual kingdoms, e.g. only 9% of the combinations occur in plants, metazoans and fungi. In contrast, the domain combinations that are shared between Arabidopsis and only one of the other kingdoms can be as many as 25% (Shiu and Mayer unpublished, Apic *et al.*, 2001; Chothia *et al.*, 2003). Thus not only the presence or absence of individual functional domains, but even more specifically the combinatorial relationship of functional domains, reflects organismal diversity between plants and metazoans at the molecular level.

It is also well known that a large proportion of higher plants have undergone at least one round of polyploidisation (genome duplication) during their evolutionary history. This genome duplication appears to be a widespread mechanism for the generation of new genes for speciation in plants (Wendel, 2000; Bowers *et al.*, 2003). Several detailed analyses have been conducted in Arabidopsis to identify molecular relics of ancient polyploidisation events on a whole genome scale, leading to a better insight into the dynamics of alterations due to polyploidy (Pikaard, 2001). Intelligent systems have also been developed for large scale analyses and exploitation of synteny even between evolutionarily distant species such as rice and Arabidopsis (Vandepoele *et al.*, 2002a). These systems will be utilized not only for Arabidopsis-"centric" research, but also for the analysis of evolutionary relations between Arabidopsis and cereals, and also among different cereals. These tools will also be useful for transfer of knowledge from Arabidopsis to different cereals.

Large-scale cDNA sequencing projects for the Arabidopsis genome allowed confirmation or correction of the intron-exon structures of thousands of genes (Haas *et al.*, 2002; Seki *et al.*, 2002). The first genome-wide annotation sets for rice are also emerging (Schoof and Karlowski, 2003; Kikuchi *et al.*

2003). In view of this, it is now appropriate to compare the results of the Arabidopsis genome with rice and other cereals. This will also allow planning of future strategies for the annotation of cereal genomes.

3. THE 2010 VISION AND LARGE SCALE PROJECTS TO ELUCIDATE THE FUNCTION OF ALL GENES

A study of the Arabidopsis whole-genome sequence suggested that only 8% of the individual genes had been described by pre-genomic approaches. Only a fraction of these had been examined in-depth by biochemical means (TAGI, 2000). Thus, for a major part of the Arabidopsis genome, reliable functional information is not yet available. Keeping this in view, the 2010 Project, initiated in USA, is targeted to determine the functions of all Arabidopsis genes by the year 2010 (Chory *et al.*, 2000). The work involved in deciphering the functions of all genes can be considerably reduced by bioinformatic strategies, which can deduce potential biological function on the basis of similarity. However, these strategies leave a large number of genes without tentative function assignments (Mayer *et al.*, 1999; TAGI, 2000; Feng *et al.*, 2002; Sasaki *et al.*, 2002; Kikuchi *et al.*, 2003). Naturally, these strategies are limited to the transfer of available knowledge and are inherently restricted. For instance, when expanded gene families are detected, usually only few members have been functionally characterized earlier. The biochemical or molecular function (e.g., transcriptional regulation, signal perception) can be transferred quite reliably. However, biological roles cannot be assigned reliably. For instance, processes like flower development, hormone signalling or pathogen defence may involve genes of the same family for quite different biological tasks. Nevertheless, these are all based on the same molecular function.

Large-scale functional genomics projects have been launched to elucidate the function and role of each individual gene within the Arabidopsis genome. The efforts and approaches include transcriptional profiling for varying developmental phases, treatments, organs, etc. Detailed experimental investigations involving individual gene-families for different specificities, and studies involving regulation and interplay of genes are also being used for elucidating functions of different genes (Chory *et al.*, 2000). Although these approaches and analyses will increase our knowledge on the molecular lifestyle of plants, specific aspects of plant biology won't be approachable using the weed model. Some traits, including many agronomically important characteristics, are not observable in Arabidopsis. For instance the

Arabidopsis genome may not prove useful in answering questions on tuber growth and development (as in potato), root nodule formation (as in legumes), storage of carbohydrates within roots (as in sugar beet), grain filling (as in cereals), etc. Nevertheless the knowledge, the technology and the associated infrastructure resulting from Arabidopsis genome research already has a strong impact on cereal research.

4. LESSONS LEARNED FROM THE ARABIDOPSIS GENOME

Besides its scientific impact, the Arabidopsis genome project can also serve as a case study for future international genome-sequencing projects in other plant species. Lessons can be learned from this model system for planning and organizing future collaborative genome sequencing and annotation efforts. The 'whole genome shotgun' and 'clone by clone' techniques for whole genome sequencing are now well known and are being routinely used (see Chapter 13 by Yu and Wing in this book).

Annotation in the Arabidopsis genome project was conducted by several groups, so that it was necessary to compare the results of these different groups. For instance, gene prediction was performed through manual integration of several gene-finding programs and also through the use of other extrinsic information like protein or EST matches. In a test case, where three research groups annotated the same sequence, the resulting gene models showed only rare deviations, and the overall quality was very similar (TAGI, 2000). On the other hand, the quality of gene modelling varied over time, and the contigs annotated later show an improved quality. This was mainly due to an improvement in gene finding algorithms and the availability of massive data on protein and DNA sequences. For instance, recently large collections of full-length cDNA sequences greatly improved the accuracy of gene models (Haas *et al.*, 2002; Seki *et al.*, 2002). These full-length cDNA sequences have also confirmed a majority of manual gene models, thus demonstrating that manual curation of gene models greatly improves quality as compared to fully automated procedures. This approach may, however, not be feasible for future projects due to the enormous manpower needed to check every gene. Automatic annotation, on the other hand, holds the promise to be constantly updated, thus offering a consistent quality across the whole genome (Schoof and Karlowski, 2003).

Most annotation projects have over time recognized the need for the definition of standard operating procedures. Definitions of what is meant by

certain gene names or terms like "strong similarity" are necessary to provide consistent annotation and make it useful to others (Ware *et al.*, 2002). Especially in a situation where more and more annotation is based purely on *in silico* knowledge transfer, the basis of assigning function must be represented. Controlled vocabularies offer a more strict control over such naming issues, and the Gene Ontology project strives to make the necessary tools available (Ashburner *et al.*, 2000).

During the Arabidopsis project, many errors in data transfer could have been avoided by arranging common specifications beforehand, but procedures and formats for data exchange were developed as need arose. However, in the bioinformatics community, it is now recognized that there is a need for interoperability of databases, and the development of standard procedures for data exchange (Schoof, 2003; Stein, 2003). As a first step towards this goal The Arabidopsis Genome Initiative (TAGI) decided that stable, unique and general identifiers should be assigned to all Arabidopsis genes. This practice has proven invaluable for data exchange and integration of results or for the purpose of referring to research publications. In view of this, there is an urgent need of curation, such as a centralized naming service.

4.1. Molecular Peculiarities of Monocot Genes

One of the most interesting results from the rice genome annotation analysis is the discovery of a % GC gradient within coding sequences. This is reported to be a specific feature of almost all monocot genes (Wong *et al.*, 2002). This phenomenon was initially detected on the basis of a comparison between the genomic sequences of rice and Arabidopsis and was later extended to several *Gramineae* plants (e.g. maize, barley and wheat) and eudicots (tobacco). When the GC content of rice protein-coding regions is plotted as a function of position along the transcript, starting from the 5'-end, a negative gradient in the %GC can be observed (Yu *et al.*, 2002). Typically, the 5' end is up to 25% GC richer than the 3' end. This is also manifested in the patterns of codon and amino acid usage. A similar gradient was also observed in introns of *Gramineae* genes. Notably, the magnitude of GC gradients in rice (monocot) genes is large enough to interfere with the gene prediction programmes, and the detection of protein homologies across the monocot-eudicot divide (Wong *et al.*, 2002).

4.2. Conservation Since the Monocot/ Dicot Diverge

Information and knowledge transfer between species, particularly between Arabidopsis and the cereals, requires identification of orthologous genes. A large percentage of genes belong to highly expanded gene families (TAGI, 2000; Goff *et al.*, 2002). Some of these families share high overall sequence similarity. In contrast, others, e.g. certain families of transcription factors, seem to evolve faster, and only functionally important domains appear to be conserved, (TAGI, 2000; Goff *et al.*, 2002). In both cases, homology based orthology assignment is problematic, if not impossible. To overcome this limitation, synteny assisted orthology assignment has been proposed (Acarkan *et al.*, 2000). Synteny has long been observed within cereals (Gale and Devos, 1998), and has been used as an instrument to facilitate comparative QTL mapping, positional cloning and directed candidate gene approaches (Ramalingam *et al.*, 2003). However, at the microscale, important differences have been observed between corresponding genomic stretches even between closely related species (Acarkan *et al.*, 2000; Schmidt 2000; Rossberg *et al.*, 2001; Bennetzen, 2002; Bennetzen and Ramakrishna, 2002; Ramakrishna *et al.*, 2002). One may thus ask the question, if synteny assisted orthologue detection is feasible over large evolutionary distances, such as the monocot /dicot diverge? Initially, syntenic conservation between rice and Arabidopsis was predicted to be extensive (Paterson *et al.*, 1996). However, later with the availability of extensive genomic sequence stretches from both rice and Arabidopsis, and finally with the availability of the draft sequence from rice, sequence-based analysis has led to re-evaluation of the degree of synteny conservation (Liu *et al.*, 2001; Mayer *et al.*, 2001; Vandepoele *et al.*, 2002). On a dataset covering 38% of the rice genome, Vandepoele *et al.* (2002) detected >20 regions with biologically relevant collinearity, each region consisting of 4 to 11 conserved genes (Vandepoele *et al.* 2002). Using the *Oryza sativa* ssp. *japonica* draft sequence, Goff *et al.* (2000) reported 137 / 508 / 1166 syntenic groups, depending on the significance threshold (99.9%, 99% and 95% significance, respectively). Similar findings have been reported by Liu *et al.* (2001) using Monte Carlo simulations to evaluate the significance. Thus, although Arabidopsis and rice diverged 200 Mya ago, some collinearity is still detectable between these genomes, and naïve expectation would predict a similar degree of conservation between Arabidopsis and other cereal genomic sequences. However, this collinearity between Arabidopsis and cereal genomes is scant when compared to the extent of conservation among eudicots. Consequently, information transfer solely based on syntenic conservation between Arabidopsis and cereal genomes, will only be able to cover a small fraction

of genes. For members of many gene families, identification of orthology between Arabidopsis and cereal genomes will need evidence other than syntenic conservation. This can be achieved by the study of expression patterns, expression dynamics, protein interactions or metabolite profiles.

There is no doubt that large number of Arabidopsis genes have counterparts in cereals. Initial estimations of this number were complicated by problems with the correct detection of genes within the rice genome (Goff *et al.*, 2002; Wong *et al.*, 2002; Yu *et al.*, 2002). Even then, upon allowing permissive thresholds, 80% and 85% of Arabidopsis genes were reported to have counterparts in the *Oryza sativa ssp. indica* and *ssp. japonica* genomes, respectively. The reciprocal rice analysis, however, had initially suggested that there are 23% to 50% (depending on the quality criteria imposed) genes without an obvious counterpart in Arabidopsis (Goff *et al.*, 2002; Yu *et al.*, 2002). This can be explained as the rice genome is estimated to have more than double the number of genes estimated for the Arabidopsis genome. However, to overcome the apparent restrictions and complications in grass gene detection and modelling, full length cDNA isolation and sequencing projects have been launched and isolation of approximately 28,000 nonredundant cDNAs has been reported (Kikuchi *et al.*, 2003). Although this set might be slightly biased towards genes, which are constitutively expressed or strongly expressed in the sampled tissues, the data allow a more reliable approximation of the number of conserved genes, which can be detected reciprocally. Employing this dataset, 64% of rice genes turn out to have a detectable counterpart within Arabidopsis, whereas 75% of Arabidopsis genes have a counterpart within rice (Kikuchi *et al.*, 2003). Thus, while a significant number of genes in rice are unique to grasses, almost two thirds of rice genes have a counterpart in Arabidopsis. These are potential targets for cross species information transfer.

The dynamics of plant genomes can be analysed on a different level of detail on the basis of individual genes of interest or functional modules (Goff *et al.*, 2002). The results demonstrate examples of the generation of paralogues through duplication. Other examples witness the evolution of functional networks. For instance, clear rice homologues for the Arabidopsis floral meristem and organ identity genes *LFY*, *AP3*, *PI* and *AG* exist, but rice orthologues of the Arabidopsis *CAL*, *UFO* and *SUP* genes have not yet been characterized (Goff *et al.*, 2002). Similar observations have been made for the relationship of genes involved in disease resistance. Whereas clear counterparts for Arabidopsis signal transduction genes (*NDR1*, *PAD4*, *EDS1*) can be assigned in rice, for other components such as *COI1*, *NPR1* and *LSD1* clear one-to-one orthology relations can not be easily assigned (Goff *et al.*, 2002). Similarly, although counterparts for the individual genes of various

functional networks can be easily assigned, in some cases like the control of flowering time it is not evident if the network that integrates them is conserved (Goff *et al.*, 2002 and references therein). Thus further functional analysis is required to demonstrate conservation of specific regulatory circuits.

The identification and characterisation of individual protein encoding genes is becoming more and more routine. The focus and scientific challenge is shifting towards detection and functional analysis of conserved regions outside the protein-coding portion of the genes. A research area which has attracted major attention recently has been the discovery of microRNAs (miRNAs) in plants and their role in regulation of developmental processes at the molecular level (Llave *et al.*, 2002; Boutet *et al.*, 2003; Kasschau *et al.*, 2003; Kidner and Martienssen, 2003). miRNA regulate their targets through triggering degradation of the target mRNAs. It has been shown that miRNAs as well as their targets for binding are conserved over long evolutionary distances, e.g. between Arabidopsis and rice (Reinhart *et al.*, 2002; Rhoades *et al.*, 2002). Thus, the detection of miRNAs using comparative approaches as well as the study of their action in a comparative framework is now feasible.

A future emphasis in plant genome analysis is the study of conserved elements adjacent to transcribed regions. Following evolutionary logic, these are of functional importance and thus are the prime candidates for *cis* regulatory elements (for a review, see Wray *et al.*, 2003). A prominent and probably the most powerful approach to detect and analyse such elements is *phylogenetic footprinting* (Blanchette *et al.*, 2000; Wasserman *et al.*, 2000; Tompa, 2001). Concerns about the prospects of success for the detection of conserved *cis* elements for divergent orthologous genes separated for long evolutionary time have been raised (Tompa, 2001). However, comparison of the promoter regions of selected orthologous genes from rice and Arabidopsis for the detection of conserved *cis* elements has proven successful (Blanchette and Tompa, 2002). However, only a limited number of *cis* elements seem to be detectable over this long evolutionary distance (Tompa, 2001; Blanchette and Tompa, 2002). Technology in this area is rapidly evolving, and Arabidopsis is still a major model platform. Consequently, despite restricted utility of a direct rice vs. Arabidopsis analysis, technology developed and established for *cis*-element analysis among dicots can certainly be utilized for cereals. In fact, a first, promising *cis* element analysis on limited datasets from monocotyledonous plants has already been reported (Guo *et al.*, 2003).

5. RECONSTRUCTOMICS AS A STRATEGY FOR THE STUDY OF CEREAL TRANSCRIPTOMES

Among cereal genomes, the rice genome is the only representative for which a nearly complete genome sequence is available. Due to the large size of other cereal genomes, no other cereal genome is likely to be fully sequenced in the foreseeable future. Under these conditions, sequencing of only the transcribed part (the transcriptome) of the respective genomes is an economical alternative. This approach is limited by the inherent quantitative heterogeneity within the transcriptome, with some genes highly represented and others hardly detectable. This affects interpretation of any transcriptome data. Under these conditions, an approximation of the transcriptome is obtained from expressed sequence tags (ESTs), which provide an economic way to analyze the transcriptome even though the sequence quality is restricted (Rudd, 2003). The typical nucleotide error rate has been estimated at ~3% for human ESTs (Hillier *et al.*, 1996), the coverage is incomplete, and there is considerable redundancy. Nevertheless bioinformatic methods to tackle these complications have been developed (see Chapter 14 by Matthews *et al*. in this book).

For some cereals, large EST data are now available (e.g. *Triticum aestivum*, 503,828; *Hordeum vulgare*, 348,282; and *Zea mays*, 384,103 sequences as of November, 2003). An exercise involving structuring and comprehensive analysis of this data can yield valuable snapshots of the genome (see below). The inherent redundancy actually serves to overcome some of the quality problems. A dataset, the "reconstructome", (a reconstructed, partial transcriptome) can be constructed, which, though based on fragmentary and necessarily incomplete sequences, provides a highly informative view of the corresponding genome.

5.1. Sequence Clustering and Assembly of Tentative Consensus Sequences (TCs)

The “tentative consensus (TC) sequence” is the most useful commodity extracted from the large EST resources. Within an EST processing pipeline, xenocontaminants are removed and vector or polylinker remnants are stripped off from the sequences. The ESTs are then assembled into contigs of overlapping ESTs (Liang *et al.*, 2000; Rudd *et al.*, 2003). A typical consensus sequence yielded from such an assembly will be longer than any of the ESTs. Misread sequence bases and frameshift errors will be averaged

Table 1. Clustering of barley EST sequences results in TCs and singletons

	Number of sequences	Total sequence (kb)	Average sequence length (bp)
ESTs	339,027	180,774	533
Clustered assemblies	63,378	50,065	790
Tentative consensus	26,432	25,241	955
Singletons	36,946	24,824	672

out when redundancy is present and this leads to a higher sequence quality. When a typical EST collection is clustered and assembled, a large number of sequences can be assembled into such TCs (see Table 1 for more details, and consult Chapter 14 by Matthews *et al.* in this book). It can be seen from Table 1, that in barley, 180 Mbp of sequences have been reduced to a unigene set (TCs and singletons) of 50 Mbp, and the average length of the sequence increased by almost 50% with associated increase in quality.

5.1.1. *Arabidopsis thaliana* as a Source for Surrogate TC Annotation

Cereal genomics or reconstructomics requires structuring of data and transfer of knowledge, particularly from the Arabidopsis genome. On the basis of orthology, and in many cases on the basis of homology, functional attributes are assigned to a sequence, which shows significant identity to an Arabidopsis sequence. The Arabidopsis protein collection has been annotated for a wide variety of sequence attributes. For instance, the pertinent annotated features include the assignments of protein function. By comparing sequences from cereals to the functionally annotated Arabidopsis proteins, distribution of functional attributes within the sequence collection can be uncovered. Such analyses require nonredundant TC sequences rather than raw ESTs to avoid skew of the proportions of assignments (see Table 2). With sequencing of additional plant genomes well underway (maize, barley, medicago, poplar), and with swelling EST databases, the future analysis, will improve our knowledge further about the function of genome sequences.

5.1.2. TCs and Protein Sequences

The availability of a reconstructome provides a jumping board to move beyond the nucleotide sequences. While nucleotide sequences are a highly suitable frame for the detection of similarity between related sequences,

Table 2. A rolling classification of 63,378 TCs from barley*

Sequence comparison	Sequence matches	Uncharacterised sequences classified	Unclassified sequences
Arabidopsis protein matches	28,941	28,941	34,437
Arabidopsis genome matches	5,953	2,074	32,363
Rice protein matches	27,481	7,220	25,143
Chloroplast matches	1,153	22	25,121
Mitochondria matches	489	3	25,118
Contaminant matches	2,721	439	24,679
Swissprot matches	18,983	1,489	23,190
Comparative matches	48,731	12,707	10,483
Short matches	9,518	3,573	6,910

*The original 63,378 tentative sequences have been assigned different sequence attributes. Protein matches (Arabidopsis protein, Rice protein and Swissprot) have been determined by BLASTX homology using an expectation value of 10e-15. Nucleotide matches (Arabidopsis genome, Chloroplast, Mitochondria, Contaminant and Comparative) have been determined by BLASTN homology using an expectation value of 10e-15. Short sequences have an associated peptide sequence of less than 50 amino-acids as determined by the framefinder coding potential assessment of the TC sequence. The column 'sequence matches' describes the number of TC sequences than can be matched to the sequence class investigated. The column 'uncharacterized sequences classified' shows the number of sequences for this class that can be explained using this criteria and that have not already been explained elsewhere. The column 'unclassified sequences' lists the number of barley TC sequences for which the origin cannot be explained. The contaminant collection includes a collection of human, *E. coli*, vector, phage and non-protein coding RNAs. The Swissprot collection is the complete Swissprot protein database. The comparative collection is a collection of TCs from all other plant EST collections with more than 5,000 ESTs as available within Sputnik (Rudd et al., 2003). The final number of 6,910 is the number of barley TCs that encode a long peptide sequence and that are apparently unique to the barley EST collection.

increased evolutionary distances render this measure inappropriate, so that protein sequence is required for sensitive analysis. As sequences become more dissimilar, protein domains and structural features become the most valuable landmarks from which information can be inferred. Such structural features and domains are often cryptic at the nucleotide level, and are generally not readily characterized on the basis of nucleotide sequence alone. It is therefore necessary to examine the peptide sequence. This is however complicated by the lack of a clear ORF (open reading frame) in many TCs as well as the presence of UTRs (untranslated regions) within the consensus sequences. While the longest ORF remains a good indicator of true CDS (coding DNA sequence), this measure becomes blurred in the presence of frameshifts, which are an inherent feature of ESTs and TCs. Alternative methods have been developed that extract the coding sequence from a TC or EST on the basis of coding potential and are therefore largely tolerant of a background of low sequence error and frameshift (Iseli *et al.*, 1999). However, a problem results from the pronounced 5′ to 3′ slope in GC content in most of the cereal genes (Wong *et al.*, 2002). This considerably complicates and decreases the fidelity of frame finding in cereal TCs. However, through the large-scale prediction of CDS, the (partial) proteome

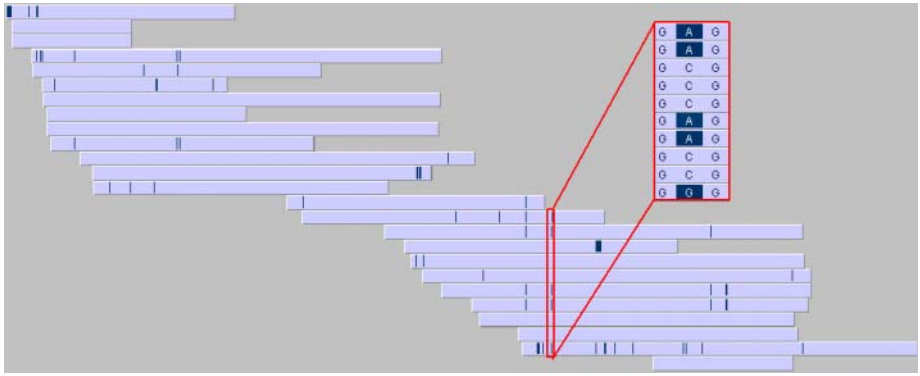


Figure 1. A graphic depicting how 25 barley ESTs have been clustered and assembled to form a single tentative consensus sequence.
The light grey boxes represent where an EST is present relative to the TC consensus sequence. A dark grey shaded bar represents where an individual nucleotide (or lack of) is in disagreement with the consensus sequence. While some of the disagreements relate to sequencing errors or base-calling errors, others with conserved patterns of disagreement relate to putative SNPs. A C to A transversion is highlighted in the panel shown.

rather than the nonredundant DNA sequence can be subjected to analyses. This enables more sensitive *de novo* sequence annotation, e.g. function or structure prediction, based on intrinsic properties of the amino acid sequence.

5.1.3. TCs for Developing Molecular Markers

For the generation of EST sequences, often a variety of tissues, developmental stages and even different ecotypes or varieties are sampled. This broad coverage of the transcriptome allows immediate access to a variety of materials suitable for the selection of molecular markers. The easiest markers to score are microsatellites or *simple sequence repeats* (SSRs), which are ubiquitous and are found within both UTR and coding sequences (Cardle *et al.*, 2000; Varshney *et al.*, 2002; Schmid *et al.*, 2003; see Chapter 4 by Schulman *et al.* in this book). Other markers that can be extracted from EST and TCs are *conserved orthologous sequence* (COS) markers (Fulton *et al.*, 2002). Computer programmes or scripts can also be developed to screen *single nucleotide polymorphism* (SNP) between the ESTs of two or more than two species of an organism (see Chapter 3 by Varshney *et al.* in this book). For instance, larger EST collections, when sampled from different genotypes, allow for the selection of likely SNP markers (Fig. 1; Kota *et al.*, 2003). Consequently methods have been developed to harvest *in silico* these different types of markers from EST and TC collections (for a review see Gupta and Rustgi, 2004).

5.1.4. TCs for Comparative Analysis of Genomes

TC collections from a wide range of organisms enable us to perform first comparative genomic analyses. The most sensitive analyses would require a few complete crop genomes as the ideal reference points. However, reconstructome sequence alone is enough for a first peek into similarities and differences within cereal genomes. As an example, Table 3 shows how barley EST sequences overlap with the other large plant EST collections. As expected, clear patterns of relatedness can be observed between the crops.

5.2. New Methodologies

While EST sequencing is an economic and powerful way to get an overview of a particular cereal genome, there are inherent problems especially with regard to genome coverage and the fragmentary nature of EST and TC sequence information. However the reconstructomic methods developed to structure, analyse and mine the sequence information are not restricted to use in EST/TC analysis. To a large extent they can be applied to new methods and strategies that are being used, and which produce short sequences. The most contemporary and exciting alternatives to complete genome sequencing currently include RescueMu insertion site sequencing and sequencing of Cot and methyl filtrated libraries (Rabinowicz *et al.*, 1999; Raizada *et al.*, 2001; Yuan *et al.*, 2003). These methodologies enrich genome sequencing libraries for gene containing regions and are highly amenable to automated analysis within sequence annotation pipelines initially designed for EST and TC analysis. The enrichment for gene rich islands is an exciting and promising methodological advancement and will allow to overcome some of the limitations associated with EST/TC analysis.

6. SUMMARY AND OUTLOOK

Arabidopsis has been the best-characterised plant at the molecular level. A complete genome sequence of this model plant allows insight into the basic plant toolbox. Beside the identification of each individual gene within the genome, increasing emphasis is laid on comprehensive analysis of genome features beyond the genes, that is the regulation, topology, structure, as well as the functional networks and their interplay. While in-depth, full-scale and detailed analyses remain feasible only in the model plant, strategies are being developed and refined to transfer the knowledge gained within the model plant Arabidopsis to crops like cereals. Furthermore, the strategies developed

Table 3. A comparative analysis of the 63,378 barley TCs with other plant species*

Comparative species	Sequence matches	As a %
Triticum aestivum	45,119	71.2
Zea mays	28,821	45.5
Sorghum bicolor	27,946	44.1
Oryza sativa	26,563	41.9
Secale cereale	17,130	27.0
Glycine max	8,752	13.8
Gossypium arboreum	8,367	13.2
Medicago truncatula	6,371	10.1
Lycopersicon esculentum	6,157	9.7
Solanum tuberosum	5,419	8.6
Arabidopsis thaliana	5,150	8.1
Ipomoea nils	4,545	7.2
Lotus japonicus	4,197	6.6
Lactuca sativa	4,060	6.4
Mesembryanthemum	3,504	5.5
Populus balsamifera	3,347	5.3
Capsicum annuum	3,116	4.9
Prunus persicus	3,088	4.9
Beta vulgaris	3,048	4.8
Vitis vinifera	2,712	4.3
Pinus taeda	1,960	3.1
Helianthus annuus	1,748	2.8
Chlamydomonas spp.	1,728	2.7
Physcomitrella patens	1,541	2.4
Porphyra yezoensis	188	0.3

*The TC sequences have been compared to the EST collections from other plant species with more than 5000 ESTs. The number of sequences determined as overlapping by BLASTN of the TCs filtered using the expectation value of 10e-10 are shown for barley : comparative genome. The number of sequences is also expressed as a % and has been ranked by the barley: comparative genome overlap. There is a clear trend for more relatedness within the other crop cereal genomes.

to exploit partial genome sequence can be tested and fine-tuned on this model plant and then applied to cereal datasets. While both these are necessarily based on approximations, they hold the promise of overcoming some of the limitations that restrict analyses in cereals. Although the demonstrated similarity between the species can be the immediate target for knowledge transfer, from a systematic comparison we will also be able to recognize the differences between the model Arabidopsis genome and cereals. The accumulation of data and experience both within the model plant

and while utilizing this data for cereals will add up to an enrichment of cereal genomics with a body of knowledge that could not have been obtained with reasonable effort by other means.

7. REFERENCES

Acarkan A., Rossberg M., Koch M., Schmidt R. Comparative genome analysis reveals extensive conservation of genome organisation for *Arabidopsis thaliana* and *Capsella rubella*. Plant Jour 2000; 23:55-62

Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D., Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F., *et al.* The genome sequence of *Drosophila melanogaster*. Science 2000; 287:2185-2195

Alonso J.M., Stepanova A.N., Leisse T.J., Kim C.J., Chen H., Shinn P., Stevenson D.K., Zimmerman J., Barajas P., Cheuk R., *et al.* Genome-Wide insertional mutagenesis of *Arabidopsis thaliana*. Science 2003; 301:653-657

Anon. Genome sequence of the nematode *C. elegans*: A platform for investigating biology. Science 1998; 282:2012-2018

Anon.. The Cold Spring Harbor Laboratory, Washington University Genome Sequencing Center, and PE Biosystems Arabidopsis Sequencing Consortium. The complete sequence of a heterochromatic island from a higher eukaryote. Cell 2000; 100:377-386

Apic G., Gough J., Teichmann S.A. Domain combinations in archaeal, eubacterial and eukaryotic proteomes. Jour Mol Biol 2001; 310:311-325

Ashburner M., Ball C.A., Blake J.A., Botstein D., Butler H., Cherry J.M., Davis A.P., Dolinski K., Dwight S.S., Eppig J.T., *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nature Genet 2000; 25:25-29

Bennetzen J.L. Mechanisms and rates of genome expansion and contraction in flowering plants. Genetica 2002; 115:29-36

Bennetzen J.L., Ramakrishna W. Numerous small rearrangements of gene content, order and orientation differentiate grass genomes. Plant Mol Biol 2002; 48:821-827

Blanchette M., Schwikowski B., Tompa M. An exact algorithm to identify motifs in orthologous sequences from multiple species. Proc Int Conf Intell Syst Mol Biol 2000; 8:37-45.

Blanchette M., Tompa M. Discovery of regulatory elements by a computational method for phylogenetic footprinting. Genome Res 2002; 12:739-748

Boutet S., Vazquez F., Liu J., Beclin C., Fagard M., Gratias A., Morel J.B., Crete P., Chen X., Vaucheret H. Arabidopsis HEN1. A Genetic link between endogenous miRNA controlling development and siRNA controlling transgene silencing and virus resistance. Curr Biol 2003; 13:843-848

Bowers J.E., Chapman B.A., Rong J., Paterson A.H. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature 2003; 422:433-438

Brendel V., Kurtz S., Walbot V. Comparative genomics of Arabidopsis and maize: prospects and limitations. Genome Biol 2002; 3:REVIEWS1005

Brett D., Pospisil H., Valcarcel J., Reich J., Bork P. Alternative splicing and genome complexity. Nature Genet 2002; 30:29-30

Cardle L., Ramsay L., Milbourne D., Macaulay M., Marshall D., Waugh R. Computational and experimental characterization of physically clustered simple sequence repeats in plants. Genetics 2000; 156:847-854

Chandler V.L., Brendel V. The Maize genome sequencing project. Plant Physiol 2002; 130:1594-1597

Chory J., Ecker J.R., Briggs S., Caboche M., Coruzzi G.M., Cook D., Dangl J., Grant S., Guerinot M.L., Henikoff S., *et al.* National Science Foundation-sponsored workshop report: "The 2010 Project" functional genomics and the virtual plant. A blueprint for understanding how plants are built and how to improve them. Plant Physiol 2000; 123:423-426

Chothia C., Gough J., Vogel C., Teichmann S.A. Evolution of the Protein Repertoire. Science 2003; 300:1701-1703

Feng Q., Zhang Y.J., Hao P., Wang S.Y., Fu G., Huang Y.C., Li Y., Zhu J.J., Liu Y.L., Hu X., *et al.* Sequence and analysis of rice chromosome 4. Nature 2002; 420:316-320

Fransz P.F., Armstrong S., de Jong J.H., Parnell L.D., van Drunen G., Dean C., Zabel P., Bisseling T., Jones G.H. Integrated cytogenetic map of chromosome arm 4S of A-thaliana: Structural organization of heterochromatic knob and centromere region. Cell 2000; 100:367-376

Fulton T.M., Van der Hoeven R., Eannetta N.T., Tanksley S.D. Identification, analysis, and utilization of conserved ortholog set markers for comparative genomics in higher plants. Plant Cell 2002; 14:1457-1467.

Gale M.D., Devos K.M. Plant comparative genetics after 10 years. Science 1998; 282:656-659

Goff S.A., Ricke D., Lan T.H., Presting G., Wang R.L., Dunn M., Glazebrook J., Sessions A., Oeller P., Varma H., *et al.* A draft sequence of the rice genome (*Oryza sativa* L. *ssp japonica*). Science 2002; 296:92-100

Guo H., Moose S.P. Conserved noncoding sequences among cultivated cereal genomes identify candidate regulatory sequence elements and patterns of promoter evolution. Plant Cell 2003; 15:1143

Gupta P.K., Rustgi S. Molecular markers from transcribed/expressed region of genome in higher plants. Funct Integr Genomics 2004 (in press)

Haas B.J., Volfovsky N., Town C.D., Troukhan M., Alexandrov N., Feldmann K.A., Flavell R.B., White O., Salzberg S.L. Full-length messenger RNA sequences greatly improve genome annotation. Genome Biol 2002; 3:RESEARCH0029

Hillier L.D., Lennon G., Becker M., Bonaldo M.F., Chiapelli B., Chissoe S., Dietrich N., DuBuque T., Favello A., Gish W., *et al.* Generation and analysis of 280,000 human expressed sequence tags. Genome Res 1996; 6:807-828

Iseli C., Jongeneel C.V., Bucher P. ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. Proc Int Conf Intell Syst Mol Biol 1999:138-148

Kasschau K.D., Xie Z., Allen E., Llave C., Chapman E.J., Krizan K.A., Carrington J.C. P1/HC-Pro, a viral suppressor of RNA silencing, interferes with Arabidopsis development and miRNA unction. Dev Cell 2003; 4:205-217

Kidner C.A., Martienssen R.A. Macro effects of microRNAs in plants. Trends Genet 2003; 19:13-16

Kikuchi S., Satoh K., Nagata T., Kawagashira N., Doi K., Kishimoto N., Yazaki J., Ishikawa M., Yamada H., Ooka H., *et al.* Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice. Science 2003; 301:376-379

Kota R., Rudd S., Facius A., Kolesov G., Thiel T., Zhang H., Stein N., Mayer K., Graner A. SniPping polymorphisms from the large barley EST collections. Mol Gen Genom 2003; 270: 24-33

Ku H.M., Vision T., Liu J., Tanksley S.D. Comparing sequenced segments of the tomato and Arabidopsis genomes: large-scale duplication followed by selective gene loss creates a network of synteny. Proc Natl Acad Sci USA 2000; 97:9121-9126

Lander E.S., Linton L.M., Birren B., Nusbaum C., Zody M.C., Baldwin J., Devon K., Dewar K., Doyle M., FitzHugh W., *et al.* Initial sequencing and analysis of the human genome. Nature 2001; 409:860-921

Liang F., Holt I., Pertea G., Karamycheva S., Salzberg S.L., Quackenbush J. An optimized protocol for analysis of EST sequences. Nucl Acid Res 2000; 28:3657-3665

Liu H., Sachidanandam R., Stein L. Comparative genomics between rice and Arabidopsis shows scant collinearity in gene order. Genome Res 2001; 11:2020-2026

Llave C., Xie Z., Kasschau K.D., Carrington J.C. Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA. Science 2002; 297:2053-2056

Lukowitz W., Gillmor C.S., Scheible W.R. Positional cloning in Arabidopsis. Why it feels good to have a genome initiative working for you. Plant Physiol 2000; 123:795-806

Mayer K., Murphy G., Tarchini R., Wambutt R., Volckaert G., Pohl T., Dusterhoft A., Stiekema W., Entian K.D., Terryn N., *et al.* Conservation of microstructure between a sequenced region of the genome of rice and multiple segments of the genome of *Arabidopsis thaliana*. Genome Res 2001; 11:1167-1174

Mayer K., Schuller C., Wambutt R., Murphy G., Volckaert G., Pohl T., Dusterhoft A., Stiekema W., Entian K.D., Terryn N., *et al.* Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana*. Nature 1999; 402:769-+

McCombie W.R., de la Bastide M., Habermann K., Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L., Yordan C., *et al.* The complete sequence of a heterochromatic island from a higher eukaryote. Cell 2000; 100:377-386

Paterson A.H., Lan T.H., Reischmann K.P., Chang C., Lin Y.R., Liu S.C., Burow M.D., Kowalski S.P., Katsar C.S., DelMonte T.A., *et al.* Toward a unified genetic map of higher plants, transcending the monocot-dicot divergence. Nature Genet 1996; 14:380-382

Pikaard C.S. Genomic change and gene silencing in polyploids. Trends Genet 2001; 17:675-677

Pruitt R.E., Meyerowitz E.M. Characterization of the genome of *Arabidopsis thaliana*. Jour Mol Biol 1986; 187:169-183

Rabinowicz P.D., Schutz K., Dedhia N., Yordan C., Parnell L.D., Stein L., McCombie W.R., Martienssen R.A. Differential methylation of genes and retrotransposons facilitates shotgun sequencing of the maize genome. Nature Genet 1999; 23:305-308

Raizada M.N., Nan G.L., Walbot V. Somatic and germinal mobility of the RescueMu transposon in transgenic maize. Plant Cell 2001; 13:1587-1608

Ramakrishna W., Emberton J., SanMiguel P., Ogden M., Llaca V., Messing J., Bennetzen J.L. Comparative sequence analysis of the sorghum Rph region and the maize Rp1 resistance gene complex. Plant Physiol 2002; 130:1728-1738

Ramalingam J., Vera Cruz C.M., Kukreja K., Chittoor J.M., Wu J.L., Lee S.W., Baraoidan M., George M.L., Cohen M.B., Hulbert S.H., *et al.* Candidate defense genes from rice, barley, and maize and their association with qualitative and quantitative resistance in rice. Mol Plant Microbe Interact 2003; 16:14-24

Redei G.P. A heurisic glance at the past of Arabidopsis genetics. In: C. Koncz, N.-H. Chua, J. Schell, *Methods in Arabidopsis Research*, World Scientific Publishing, Singapore, 1992, pp. 1-15

Redei G.P., Koncz C. Classical mutagenesis, In: C. Koncz, N.-H. Chua, J. Schell, *Methods in Arabidopsis Research*, World Scientific Publishing, Singapore, 1992, pp. 16-82

Reinhart B.J., Weinstein E.G., Rhoades M.W., Bartel B., Bartel D.P. MicroRNAs in plants. Genes Dev 2002; 16:1616-1626

Rhoades M.W., Reinhart B.J., Lim L.P., Burge C.B., Bartel B., Bartel D.P. Prediction of plant microRNA targets. Cell 2002; 110:513-520

Rossberg M., Theres K., Acarkan A., Herrero R., Schmitt T., Schumacher K., Schmitz G., Schmidt R. Comparative sequence analysis reveals extensive microcolinearity in the lateral suppressor regions of the tomato, Arabidopsis, and Capsella genomes. Plant Cell 2001; 13:979-988

Rudd S. Expressed sequence tags: alternative or complement to whole genome sequences? Trends Plant Sci 2003, 8:7:321-329

Rudd S., Mewes H.W., Mayer K.F. Sputnik: a database platform for comparative plant genomics. Nucl Acids Res 2003; 31:128-132

Salse J., Piegu B., Cooke R., Delseny M. Synteny between *Arabidopsis thaliana* and rice at the genome level: a tool to identify conservation in the ongoing rice sequencing project. Nucleic Acids Res 2002; 30:2316-2328

Sasaki T., Matsumoto T., Yamamoto K., Sakata K., Baba T., Katayose Y., Wu J.Z., Niimura Y., Cheng Z.K., Nagamura Y., *et al.* The genome sequence and structure of rice chromosome 1. Nature 2002; 420:312-316

Schmid K.J., Sorensen T.R., Stracke R., Torjek O., Altmann T., Mitchell-Olds T., Weisshaar B. Large-Scale Identification and analysis of genome-wide single-nucleotide polymorphisms for mapping in *Arabidopsis thaliana*. Genome Res 2003; 13:1250-1257

Schmidt R. Synteny: recent advances and future prospects. Curr Opin Plant Biol 2000; 3:97-102

Schoof H. Towards interoperability in genome databases: The MatDB (MIPS Arabidopsis thaliana Database) experience. Comp Funct Genom 2003; 4: 255-258

Schoof H., Karlowski W.M. Comparison of rice and Arabidopsis annotation. Curr Opin Plant Biol. 2003; 6:106-12

Seki M., Narusaka M., Kamiya A., Ishida J., Satou M., Sakurai T., Nakajima M., Enju A., Akiyama K., Oono Y., *et al.* Functional annotation of a full-length Arabidopsis cDNA collection. Science 2002; 296:141-145

Stein L.D. Integrating biological databases. Nature Rev Genet 2003; 4:337-345

TAGI. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 2000; 408:796-815

The Rice Chromosome 10 Consortium In-depth view of structure, activity, and evolution of rice chromosome 10. Science 2003; 300:1566-1569

Tompa M. Identifying functional elements by comparative DNA sequence analysis. Genome Res 2001; 11:1143-1144

Vandepoele K., Saeys Y., Simillion C., Raes J., Van De P.Y. The automatic detection of homologous regions (ADHoRe) and its application to microcolinearity between Arabidopsis and rice. Genome Res 2002; 12:1792-1801

van Dodeweerd A.M., Hall C.R., Bent E.G., Johnson S.J., Bevan M.W., Bancroft I. Identification and analysis of homoeologous segments of the genomes of rice and *Arabidopsis thaliana*. Genome 2001; 42:887-892

Varshney R.K., Thiel T., Stein N., Langridge P., Graner A. *In silico* analysis on frequency and distribution of microsatellites in ESTs of some cereal species. Cell Mol Biol Lett 2002; 7:537-546

Venter J.C., Adams M.D., Myers E.W., Li P.W., Mural R.J., Sutton G.G., Smith H.O., Yandell M., Evans C.A., Holt R.A., *et al.* The sequence of the human genome. Science 2001; 291:1304-+

Ware D., Jaiswal P., Ni J., Pan X., Chang K., Clark K., Teytelman L., Schmidt S., Zhao W., Cartinhour S., *et al.* Gramene: a resource for comparative grass genomics. Nucleic Acids Res 2002; 30:103-105

Wray G.A, Hahn M.W., Abouheif E., Balhoff J.P., Pizer M., Rockman M.V., Romeno L.A. The evolution of transcriptional regulation in eukaryotes. Mol. Biol. Evol. 2003; 20: 1377-1419.

Wasserman W.W., Palumbo M., Thompson W., Fickett J.W., Lawrence C.E. Human-mouse genome comparisons to locate regulatory sites. Nature Genet 2000; 26:225-228

Wendel J.F. Genome evolution in polyploids. Plant Mol Biol 2000; 42:225-249

Wong G.K.S., Wang J., Tao L., Tan J., Zhang J.G., Passey D.A., Yu J. Compositional gradients in Gramineae genes. Genome Res 2002; 12:851-856

Yano M., Katayose Y., Ashikari M., Yamanouchi U., Monna L., Fuse T., Baba T., Yamamoto K., Umehara Y, Nagamura Y., *et al. Hd1*, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the Arabidopsis flowering time gene *CONSTANS*. Plant Cell 2000; 12:2473-2483

Yu J., Hu S.N., Wang J., Wong G.K.S., Li S.G., Liu B., Deng Y.J., Dai L., Zhou Y., Zhang X.Q., *et al.* A draft sequence of the rice genome (*Oryza sativa* L. *ssp indica*). Science 2002; 296:79-92

Yuan Y., SanMiguel P.J., Bennetzen J.L. High-Cot sequence analysis of the maize genome. Plant Jour 2003; 34:249-255

Chapter 18

RICE GENOME AS A MODEL SYSTEM FOR CEREALS

Takuji Sasaki* and Baltazar A. Antonio

Rice Genome Research Program, National Institute of Agrobiological Sciences, 1-2, Kannondai 2-chome, Tsukuba, Ibaraki 305-8602, Japan
*Author for correspondence: tsasaki@nias.affrc.go.jp

1. INTRODUCTION

The genome size in major cereals ranges from 430 Mb in rice (*Oryza sativa*) to 16 Gb in bread wheat (*Triticum aestivum*), so that rice genome is the smallest among genomes of food crops in general, and among those of cereals in particular. However, rice genome is still three times the size of the genome of model plant, thale cress (*Arabidopsis thaliana*) that was the first genome of a higher plant to be fully sequenced. However, ever since the whole genome sequencing in model plant, Arabidopsis was initiated in early 1990s, there was a growing realization that this genome may not serve the purpose of a model genome for cereals, which constitute the major staple food crops of the world. Keeping this in view, the work on analysing the rice genome was initiated in Japan in the year 1991. Later, sequencing of the entire rice genome was initiated with the establishment of the International Rice Genome Sequencing Project (IRGSP) in the year 1997. The events that were involved in mapping and sequencing of the rice genome during 1990-2004 are listed in Table 1.

The rice genome consists of a haploid complement of 12 tiny chromosomes that were subjected to extensive cytogenetic and molecular analysis during the last two decades. For instance, karyotype analysis of rice chromosomes was undertaken using pachytene chromosomes, and trisomics were produced, which later became useful for cytogenetic analyses (Khush *et al.*, 1984, Misra *et al.*, 1986). A complete series of alien addition lines involving addition of individual chromosomes from *Oryza officinalis* to the complete set of

P.K. Gupta and R.K. Varshney (eds.), Cereal Genomics, 535-557.

Table 1. A summary of the year-wise progress of rice genome sequencing

Year	Event
1991	A seven year rice genome project at a cost of 25 million was initiated
1991	Collinearity among cereal genomes established
1993	A five year rice genome mapping project was initiated at a cost of 3.8 million in China
1997	An international consortium for sequencing of cultivar Nipponbare of *japonica* rice was formed
1998	IRGSP formally took shape and sequencing was targeted to be completed by 2008
2000 (April)	Monsanto announced completion of a draft of rice genome sequence with four times (x 4) coverage and promised to share the data with IRGSP and individual researchers
2000 (May)	Beijing Genomic Institute (BGI) announced that it plans to sequence the genome of an *indica* rice using whole genome shotgun (WGS) approach within 2 years
2001 (January)	Syngenta announced sequencing of Nipponbare genome with six times (x 6) coverage using WGS
2001 (October)	IRGSP decided to aim at 10 times (x10) coverage by the end of 2002
2002	Science published Syngenta and BGI draft sequences of *japonica* and *indica* rice genomes
2003	Nature and Science published nearly complete high quality sequences of chromosomes 1, 4 and 10

from *O. sativa* was also produced and utilized (Jena and Khush, 1986). However, the molecular analysis of these chromosomes was initiated later with the publication of the first molecular map of rice chromosomes in late 1980s (McCouch *et al*., 1988). This map was followed with the publication of several molecular maps including a high-resolution molecular genetic map having 2,275 markers at an average interval of 190 kb (Harushima *et al*., 1998; also see Chapter 3 by Varshney *et al.* in this book).

The availability of large insert YAC and BAC libraries also provided useful tools for generating molecular physical maps of complex genomes, so that BAC-based genome-wide physical maps became available for all the 12 rice chromosomes (Chen *et al.*, 2002), which later facilitated whole genome sequencing (WGS). These genetic and physical molecular maps were compared with molecular maps of other grass species including cereals,

millets and sugarcane. Such a comparison exhibited synteny conservation and collinearity of markers between different grass genomes. More recently, several drafts of whole genome sequence of rice also became available, which further increased the value of rice genome as a model for a study of other cereal genomes.

The above wealth of information about the rice genome and continued research on this genome at the molecular level has proved to be of immense value. The recent large-scale dense mapping and genome sequencing has led to the elucidation of the genome structure of not only rice, but also that of wheat, maize and several other grass species. This provides a glimpse of the events, which define the evolution of each individual grass species, particularly of cereal genomes, after divergence from a common ancestor. It also has a strong bearing on the cereal improvement programmes, opening out avenues for further improvement of cereal crops in an unprecedented manner. In this review article, the potential of using rice as a reference genome is discussed with emphasis on the rice genome sequencing and its impact not only on an understanding of the collinearity of gene orders among grass species, but also on the identification of orthologous genes even in those grass genomes, where sufficient genomic information is currently not available.

2. EVOLUTION WITHIN THE GRASS FAMILY

The basic chromosome number of rice, wheat and maize are $x = 12$, $x = 7$, and $x = 10$, respectively. As mentioned earlier, the genome size of these species varied from 430 Mb for rice (*Oryza sativa*), a diploid species, to 16 Gb for bread wheat (*Triticum aestivum*), a hexaploid species (Arumuganathan and Earle, 1991). Although the driving force which led to such varied chromosome numbers and genome size among closely related species is still unknown, it can be speculated that the derivation of the genome of each species from a common ancestral chromosome must have been an adaptive mechanism that allowed their survival and sustained reproduction to the benefit of mankind in varied environments. Furthermore, the pressure exerted by continuous human manipulation through breeding and selection must have caused many other structural changes in chromosomes, such as duplications, rearrangements by inversions or translocations, and expansion with the introduction of transposable elements. Domestication of wild species is generally aimed at generating plants, which produce a much larger quantity of seeds or fruits for human needs, than what the plant needs to produce for

its survival. Occasionally, this change was accompanied by increase in the number of chromosomes and increase in genome size. It is still difficult to visualize, how and why during the course of evolution, duplications in genomic segments occurred in bread wheat and maize at such a large scale, relative to those in the rice genome. Nevertheless, during the last ten years, genome analysis of major grass species has significantly improved our understanding of the evolutionary relationships among different cereal genomes. The studies at the molecular level also allowed elucidation of presence/absence of orthologous genes and their conserved or altered functions in each species.

3. GENETIC ANALYSIS OF THE RICE GENOME

In the past, detailed comparative studies at the molecular level have been conducted in all cereal genomes. These studies proved extremely useful for understanding the general structure of cereal genomes. For such a comparative analysis, an appropriate model genome was needed, and the rice genome proved to be an ideal genome for this purpose, even though rice is distantly related to wheat and maize (Grass Phylogeny Working Group, 2000). A single factor that weighed heavily in selecting rice genome as the model was its small size relative to all other cereal genomes (Arumuganathan and Earle, 1991), thus making molecular dissection of this model genome much easier (Sasaki *et al.*, 1996). Its small size also encouraged its selection as the first cereal genome to be subjected to high quality whole genome sequencing through an international effort (Sasaki and Burr, 2000). During the course of extensive molecular genetic analysis of the genomes of rice and other cereals, a variety of tools have also become available, which facilitated further studies. For instance, the discovery of collinearity by comparative mapping of cereal genomes was initially facilitated by using a number of same anonymous DNA markers like RFLPs (restriction fragment length polymorphism) for mapping across a number of genomes (McCouch and Tanksley, 1991). Later, SSRs (simple sequence repeats), AFLPs (amplified fragment length polymorphisms), and SNPs (single nucleotide polymorphisms) were used as other DNA-marker systems for comparative mapping. These markers may though sample both the coding and the non-coding regions of the genome for suggesting synteny and collinearity, but polymorphisms only in the expressed region proved useful for a comparison of gene order. In all such studies in cereals, it is the rice genome, which has been used as a model (see also Chapter 5 by Paterson and Chapter 11 by Stein and Graner in this book).

3.1. Genetic Maps of Rice and Other Genomes

Although, among cereals, maize genome was the first to be used for construction of an RFLP map (Helentjaris *et al.*, 1986), construction of molecular map of the rice genome was only the next to follow (McCouch *et al*., 1988; McCouch and Tanksley, 1991). The DNA markers that were used for mapping the rice genome also included the RFLPs that were detected using maize cDNAs as probes, so that the maps of rice and maize genome could be easily compared to detect collinearity between rice and maize genomes (Ahn and Tanksley, 1993). This also provided evidence to suggest that genomic structure among grass species may be largely conserved (Helentjaris, 1993; Moore *et al*., 1993). A high-density functional linkage map with nearly 900 expressed genes was later constructed for the rice genome (Kurata *et al.*, 1994a), which facilitated the determination of segmental collinearity between rice and wheat genomes (Kurata *et al.*, 1994b). These studies on comparative mapping encouraged further research in the field of comparative genetics and genomics, so that the collinearity in the order of DNA markers and that of the known genes among rice, wheat, and maize genomes also reflected the existence of a common ancestral chromosome structure, which diverged 40-50 million years ago into chromosomes that are found in the other present-day cereals.

3.2. Collinearity between Rice Genome and Other Grass Genomes

Although the terms collinearity and synteny have sometimes been used as synonyms, more recently a distinction has been made between these two terms. Synteny means the ‘same thread’ or a state of being together in location in the same manner that synchrony means being together in time (Renwick, 1971). For example, rice chromosome 1 is known to share synteny with wheat group 1 and with maize chromosomes 3 and 8 (Devos and Gale, 1997). In contrast, collinearity refers to the same linear order of markers or genes on syntenic chromosomes from two different genomes. However, nowadays, the term synteny is also used to refer to the gene loci in different organisms located on a chromosomal region of common evolutionary history. Moore (1995) had also shown that the genetic maps of major cereal genomes could be combined in an integrated grass genome map and that the 12 rice chromosomes could be divided into 19 linkage blocks that form the basis of all chromosomes available in the grass family, by assembling in different

ways (Moore 1995; Moore *et al.*, 1997). Later, Gale and Devos (1998) prepared an extended and more detailed version of the consensus grass map based on 25 rice linkage blocks. On the basis of further more extensive studies, it was suggested that as many as about 30 rice linkage blocks are needed to represent all cereal genomes (Devos and Gale, 2000). The maize genome with 10 chromosomes was also shown to have duplicated structure, so that each constituent rice block is represented twice in the maize genome. Several other grass species such as foxtail millet, pearl millet, sugarcane, sorghum, and oats were also found to carry genome structure that was syntenic with that of rice (Gale and Devos, 1998). Therefore, the genome rearrangement that is syntenic to rice genome was observed in maize, millet, sugarcane and sorghum, which all belong to *Panicoideae* group. Such a relationship must have resulted due to divergence of a common ancestor of this group from those of other groups of grass genera.

4. ELUCIDATING THE GENOME SEQUENCE OF RICE

The efforts to obtain the rice genome sequence, although competitive at first, later became a collaborative effort between the public and private sectors. The International Rice Genome Sequencing Project (IRGSP) in the public sector, and Monsanto and Syngenta in the private sector undertook sequencing of the genome of *japonica* rice variety Nipponbare, and Beijing Genomics Institute (BGI), China, in the public sector undertook the sequencing of the genome of an *indica* rice variety, 93-11, so that a comparison of the genome sequences of *japonica* and *indica* rice varieties was also possible.

4.1. Genome Sequences of *japonica* Rice (Nipponbare) by IRGSP

The International Rice Genome Sequencing Project (IRGSP) organized in September 1997 (http://demeter.bio.bnl.gov/rice.html; Sasaki and Burr, 2000) initially spearheaded the public effort. The members of the consortium included Japan, U.S.A., China, France, Taiwan, Korea, Thailand, U.K, India and Brazil and aimed to elucidate an accurate map-based sequence of the *japonica* rice variety, Nipponbare. A clone-by-clone strategy was adopted and PAC/BAC clones aligned on the physical map of the rice genome served as templates for sequencing (Chen *et al.*, 2002).

4.2. Genome Sequences of *japonica* Rice (Nipponbare) by Private Sector (Monsanto and Syngenta)

The effort of sequencing the rice genome was also pursued by two private companies. Concurrently with IRGSP, Monsanto Co. in USA constructed a physical map of the same Nipponbare variety, which was used by IRGSP. The clones aligned in the physical map with 65% genome coverage were used for sequencing using the clone-by-clone strategy. The resulting draft sequence however was not necessarily of high quality and Monsanto had no plan whatsoever to raise the quality or finish the sequence. Instead, Monsanto decided in early 2000 to donate its data to IRGSP to accelerate the production of high-quality rice genome sequence (Barry, 2001). This greatly facilitated the efforts of the international consortium to accelerate sequencing and the immediate release of sequence data in public databases.

In order to obtain the rice genome sequence, another private company, Syngenta used an approach that was entirely different from that used by IRGSP and Monsanto. This strategy known as whole genome shotgun (WGS) sequencing was widely used for sequencing bacterial genomes, but it was considered to be difficult to use this approach for sequencing a complex genome like that of rice, which is rich in repetitive sequences. In view of this, availability of many reliable markers from repetitive as well as unique regions of the genome was considered necessary for correct assembly of sequence fragments of such complex genomes. Therefore, a BAC library was constructed to generate markers using the BAC-end sequence data and fingerprints of these BACs to allocate positions to the end sequences. As a result, 93% of the entire genome was sequenced based on the total length of independent contigs (Goff *et al*., 2002). The estimated number of genes ranged from 32,277 to 61,668 depending on the minimum gene length chosen. Although homologs of almost 98% of the known maize, wheat and barley proteins were found in rice, synteny with Arabidopsis was found to be less extensive than previously expected (Goff *et al.*, 2002). However, an estimate of the degree of accuracy of Syngenta's contig sequences can be obtained only by direct comparison with the corresponding sequence data from IRGSP.

4.3. Genome Sequences of *indica* Rice (93-11) by BGI, China

Another effort towards complete genome sequence of rice, undertaken by Beijing Genomics Institute in China, led to release of the draft sequence of an

indica variety, 93-11 (Yu *et al.*, 2002). BGI used a whole genome shotgun approach to sequence this rice variety and assembled the data by eliminating the repeated sequences. After assemblage, sequences of about 100,000 contigs with a total length of 362 Mb were obtained. Adding the repeat sequences, the total genome size of this rice variety was estimated to be 466 Mb. The estimated number of genes ranged from 46,022 to 55,615 and functional coverage determined by mapping STS markers from Nipponbare and ESTs from 93-11 was estimated to be 92%. In addition, about 49.4% of predicted rice genes had each a homologue in Arabidopsis.

Comparison of the BGI sequence with a 493 kb region in the distal end of chromosome 1 from IRGSP sequence showed that only about 78% of the *indica* sequence corresponded with the *japonica* sequence (Sasaki *et al.*, 2002). The discrepancy could be attributed either to the large number of gaps in the BGI sequence or to the sequence differences between the two subspecies. Most contigs are very short which hinder a comprehensive analysis of entire regions of the chromosomes. Also the BGI sequence does not provide positional information and most of the predicted genes are incomplete.

4.4. Genome Features Derived from the Whole Genome Draft Sequences

Since both the Syngenta and BGI used a whole genome shotgun strategy, automatic assembly of the reads has been a major problem, since nearly half of the rice genome sequence is made up of repetitive sequences. Furthermore unmapped genomic reads may not be very useful in obtaining highly accurate genome information. Nevertheless, several general biologically interesting features of the rice genome sequence were elucidated by the whole genome draft sequences. In particular, the rice genome was found to be more GC-rich than that of Arabidopsis, and a large number of GC-rich islands were detected in the draft sequences. As to repetitive sequences, all the expected types of repeats were found in the rice genome. About 60% of the fully masked sequences in the BGI data consists of transposable elements particularly retroelements and gypsy-type elements. The estimated number of genes from the draft sequences range from 46,022 to 61,668 but these wide variations are mainly due to the use of different approaches and to the poor quality of the whole genome draft sequences.

Thus, although several draft sequences of the rice genome are currently available including the IRGSP high-quality draft sequence, which was

completed in December 2002, a finished high quality sequence of the rice genome is still awaited. Monsanto's and Syngenta's collaboration with IRGSP would facilitate the completion of this goal.

5. ANNOTATION OF THE STANDARD RICE GENOME SEQUENCE

The accuracy of genome sequence should be evaluated by the quality of annotation, i.e. assignment of biological function to the sequence. Gene modeling for a given sequence using gene prediction and similarity search programmes facilitates gene discovery in a systematic and comprehensive manner. In case of rice, an automated annotation system called RiceGAAS (Rice Genome Automated Annotation System) has been developed by combining mainly coding region prediction programmes, splice site prediction programmes, t-RNA gene prediction programme, and similarity search analysis programmes (Sakata *et al.*, 2002). The interpretation of the coding region is though fully automated, gene modeling is accomplished with manual evaluation and modification using all existing resources such as ESTs, full-length cDNAs and protein homologies. In this connection, the recent release of sequence information for about 30,000 full-length rice cDNAs of Nipponbare is significant (The Rice Full-Length cDNA Consortium, 2003). A combination of high-quality genome sequence and the sequence of full-length cDNAs would identify more precisely the coding regions of the genome and would also resolve the sites of alternative splicing, wherever they occur. However, homology of a genome sequence to a cDNA must be interpreted with caution, keeping in view that errors may be introduced during synthesis of cDNA from mRNA, due to chimera formation or secondary structure of mRNA.

6. INTEGRATION OF ANNOTATIONS WITH GENETIC AND PHYSICAL MAPS

The annotated genome sequences obtained as above are integrated with other rice genomics information including the genetic maps, physical maps and transcript maps. This information is available in the database named INE (INtegrated rice genome Explorer), which was constructed using Java script (Sakata *et al.*, 2000; http://rgp.dna.affrc.go.jp /giot/INE.html). Basically in INE, the DNA markers on the molecular genetic maps are used to indicate position of YACs, PACs or BACs in the physical map. A total of 6,500 EST markers were also generated by mapping ESTs on the YAC-based physical

map (Wu *et al.*, 2002). These markers add to the 3,000 mapped markers and would facilitate the alignment of PACs and BACs along the chromosomes. The annotation results are shown in INE either as individual clone sequences, or by about 150 kb interval. After finalizing the annotation of adjacent clones, the overlapping regions are deleted to show the annotation of long contiguous sequences.

The results of BLAST search are also shown for each predicted gene to facilitate the characterization of function and other features of a gene being annotated. The prediction programmes can identify genes along the genome sequence, irrespective of its biological function, if the sequences satisfy the parameters for prediction. The minimum threshold on the number of amino acids encoded in the transcript of predicted gene was set at 50, which is the minimum size of a protein reported in plants. Among the 6,756 genes predicted for chromosome 1, 54% are hypothetical genes with no significant homology to amino acid sequences or ESTs in public databases (Sasaki *et al.*, 2002). A combined search programme of motif structure, the InterPro could, however, identify known motif structures among 33% of these hypothetical proteins, still leaving >20% of the predicted genes with no signature in any known gene (Sasaki *et al.*, 2002). The annotations of chromosomes 4 and 10 also gave 42% and 40% hypothetical proteins, respectively, among predicted genes (Feng *et al.*, 2002; The Rice Chromosome 10 Sequencing Consortium, 2003). In contrast to this high proportion of hypothetical genes in rice chromosomes already sequenced, in case of Arabidopsis, only 17% of the predicted genes are hypothetical (TAGI, 2000). In view of this, we can only speculate that in the rice genome there are many genes that are either unique to the genome of rice, or unique to cereal genomes. Future progress in the study of wheat and maize genomes may provide more information concerning the nature of these sequences.

7. REFERENCE GENOME FOR GENE DISCOVERY IN CEREALS

The value of the rice genome sequence extends far beyond an understanding of the genetic organization of the rice plant. The syntenous relationship of grasses, as discussed earlier in this chapter, has facilitated gene identification among grass genomes using rice as a reference. Some of the commonly identified traits among grass family are controlled by genes that are syntenic among several grass genomes suggesting the possibility of their being orthologous in nature. Some of the important traits that are controlled by orthologous genes among all grass species include plant height, flowering

time, shattering habit and waxy leaves (Devos and Gale, 1997). The *waxy* locus in particular has been characterized to code for granule-bound starch synthase I (Shure *et al.*, 1983).

7.1. Analysis of Dwarfing Genes

Dwarfing genes in wheat and maize have particularly been identified and cloned using information about their synteny and orthologous relationship with genes in the model genomes (Peng *et al.*, 1999). For instance, the *Rht-1* gene widely used for wheat breeding to introduce dwarf character (Gale and Youssefian, 1985) and the maize *dwarf-8* (*d8)* were found to be orthologous to gibberellin insensitive (*GAI*) gene from Arabidopsis (Peng *et al.*, 1997). Although there exists no distinct synteny between grass and Arabidopsis genomes, amino acid sequence of gene product and/or nucleotide sequences are conserved and sometimes orthologous genes could be identified in grass species using the sequence information from Arabidopsis. The sequence of *GAI* also exhibited homology to a rice EST, which was then used to identify the genes corresponding to *Rht-1* and *d8*. The wheat genomic clone, which hybridized with the above rice EST was mapped to the *Rht-D1b* locus, which therefore became the target for further study. The synteny between wheat group 4 chromosome and maize chromosome 1 could identify orthologous maize gene, *d8*. The synteny between wheat and rice also predicted the presence of a rice orthologue, or the EST used for *Rht-1* gene cloning on rice chromosome 3. Although the dwarf mutant gene on the corresponding rice chromosomal region has not been identified, a rice slender mutant (*slr1-1*) gene that is also an orthologue of the height-regulating genes in wheat and maize has been cloned and characterized at the molecular level (Ikeda *et al.*, 2001). The above genes including *GAI/RHT/D8/SLR1* have been shown to code for similar transcription factors, but they seem to differ in the position of mutations in their nucleotide sequences, thus producing motifs that are specific for their biological functions in individual plant species.

7.2. Analysis of Vernalization Gene

The syntenous relationship among cereal genomes has also been utilized for cloning of wheat vernalization gene, *VRN1* (Yan *et al.*, 2003). It is expected that any gene in rice, which may be orthologous to wheat *VRN1*, will not have any biological function, because rice does not require low temperature for induction of flowering. Yan *et al.*. (2003) performed genetic mapping of *VRN1* locus of wheat in detail using the annotated rice genome information to

generate new markers adjacent to *VRN1*. As a result, two candidate genes of *VRN1* were identified in rice, which are syntenous to *VRN1* gene region having a gene order similar to that in wheat. The same gene order was also observed in sorghum genome. This clearly indicates the utility of synteny in identifying candidate genes within grass species, even if the corresponding phenotype is not expressed in an individual species.

7.3. Analysis of Disease Resistance Genes

In contrast to the above two examples of genes, syntenous relationships may not be of much help for rapidly evolving genes like disease resistance genes (R-genes). Several of these disease resistance genes have been isolated and studied; these genes include *Pto* gene, conferring resistance against *Pseudomonas syringe* (Martin *et al.*, 1993), and several other genes carrying nucleotide binding site (NBS) and leucine-rich repeat (LRR) motifs, e.g. *RPS2* of Arabidopsis (Bent *et al.*, 1994; Mindrinos *et al.,*1994), and *Xa1* of rice (Yoshimura *et al.,*1998). The NBS includes the conserved sequence called P-loop (Grant *et al.,* 1995), so that the universal PCR primers designed using this nucleotide sequence have been used for detection of genes carrying NBS (Leister *et al.,* 1998; Yu *et al.,* 1996; Pan *et al.,* 2000). When such genes were used for genetic mapping, it could be shown that no synteny for these genes existed between rice and barley or between rice and foxtail-millet or between maize and foxtail-millet. This suggests that R-genes are rapidly reorganized in each cereal genome in order to respond to coresponding avirulence genes in the pathogen (Leister *et al.,* 1998).

The most extensive effort, which met with failure to isolate a rice R-gene using synteny between rice and barley is exemplified by *Rpg1*, a gene conferring resistance against barley stem rust (also consult Chapter 8 by Jahoor et al. in this book). This gene is located on barley chromosome 7H in a region that is collinear with the distal end of the short arm of rice chromosome 6. Therefore, efforts were made to identify rice orthologue to *Rpg1* gene (Kilian *et al.,* 1997). Detailed genetic mapping was performed utilizing new markers from barley cDNAs. Sequence analysis of a rice BAC clone that covered the corresponding syntenic barley genomic region, however, failed to identify in the rice genome, a homologue to disease resistance gene *Rpg1* (Han *et al.,* 1999). Finally, *Rpg1* gene was isolated as a receptor kinase gene by map-based cloning, using markers from rice and barley genomes for tagging (Brueggeman *et al.,* 2002).

It is obvious from the above examples and discussion that synteny and collinearity across grass species do exist and reflect the gene clustering and

order that occurred in the ancestral plant, which diverged from other groups of plants more than 60 million years ago. However, orthologues of all kinds of genes are not available in all cereal species. The synteny and collinearity are generally conserved for genes that are essential for the survival of its descendents, so that such genes can be isolated utilizing collinearity approach. In contrast, often the genes that originate after speciation to adapt to the new environment or due to interaction with specific pathogens may not have orthologues, so that for the isolation of these genes, synteny or collinearity may be of little help and map based cloning may have to be used. Despite this, it is only logical to conclude that in majority of cases, confirmation of synteny by a detailed nucleotide sequence analysis of each species will allow identification and isolation of many agronomically important genes.

8. STANDARD SEQUENCE FOR GENOME ANALYSIS

Genome sequences can also be used for comparing genomes both at the intraspecific and interspecific levels. However, in either case, a standard genome is required, with which other genomes can be compared. For instance, collinearity of genes has been studied between the genomes of *japonica* rice and *indica* rice, among different species of *Oryza* (21 wild species are known in the genus *Oryza*; Lu and Jackson, 2002), and also among different cereals representing different genera. Rice genome provides for a standard genome for this purpose, as illustrated in the following section.

8.1. Standard Sequence for Intraspecific Genome Analysis

The association of genes with important agronomic traits in a rice cultivar requires that a complete, accurate, map-based genome sequence is available in the target cultivar/subspecies (*japonica* or *indica*). At present, the availability of target sequence to compare with high-quality Nipponbare genome sequence is limited. The genome sequence data of *indica* rice variety 93-11 released by BGI, though not sufficient to cover the entire genome, allow a direct comparison with the Nipponbare sequence to reveal basic information that underlies intraspecies variations in cultivated rice. In another study, conducted for an extensive comparison of *japonica* Nipponbare with *indica* rice cultivar, Feng *et al.,* (2002) sequenced 2.3 Mb of chromosome 4 of Guangluai 4 (an *indica* rice). Although, this comparison revealed extensive sequence collinearity between the genomes of *japonica* rice and *indica* rice,

deviations from collinearity were common and involved substitutions, insertions and deletions. For instance, within this 2.3 Mb sequences, 9,056 single nucleotide polymorphisms (SNPs) were found giving a frequency of one SNP per 268 bp. Similarly, insertions and deletions (InDels) accounted for 63 InDels in Guangluai 4 sequence and 138 InDels in Nipponbare sequence. Despite this, the results suggested an overall conservation of genomic sequences in the genomes of the two subspecies of rice.

In another study, using AcycloPrime FP SNP detection kit (PerkinElmer Life Sciences), primers were designed from Nipponbare genome sequence, for detection of SNPs, so that a total of 2800 SNPs including substitutions, deletions, and insertions were detected among six varieties of *japonica* and *indica* (Nasu *et al.,* 2002). The frequencies of polymorphism within *japonica* varieties and between *indica* and *japonica* varieties were estimated to be 0.03-0.05% and 0.68-0.70%, respectively. The latter value is higher than that obtained from the sequence comparison (0.37%), because this value also included InDels. As many as 64 SNPs that occurred among *japonica* varieties were also placed on the *japonica-indica* genetic map, thus demonstrating the utility of SNPs as genetic markers even within closely related varieties of a species. In addition, the SNP frequencies between *japonica* and wild rice, *O. rufipogon* and between *indica* and *O. rufipogon* were estimated to be 0.33% and 0.71-0.75%, respectively. These values were unexpectedly low, keeping in view the known level of genetic diversity between *O. sativa* and *O rufipogon*. Thus, further detailed sequence comparison is necessary using the Nipponbare genome sequence.

BAC-end sequencing followed by *in silico* mapping of the resulting BAC-end sequences on the accurate standard genome sequence is another approach for intraspecific genome analysis. This approach can be used both for the collection of SNPs and for construction of a physical map of the genomes of individual rice varieties that have not been used for genome sequencing. One such study was conducted using the *indica* variety, Kasalath, which is one of the parents of F2 population earlier produced for molecular genetic mapping (Kurata *et al.,* 1994a). A BAC library with 15 times genome coverage was constructed in this variety and the ends of all clones were sequenced. After collecting end-sequences with reliable quality, BLASTN search using the Nipponbare genome sequence for each chromosome was carried out. About 2,000 Kasalath BACs were thus mapped on chromosome 1 based on the criteria that the end-sequences should be mapped in pair and that the paired end sequences be separated by a reasonable BAC insert size of 150 kb. It was also ensured that if a BAC-end sequence spans any of the mapped Nipponbare markers, then the latter should be available in the corresponding Kasalath sequences. Following this approach, so far 22 contigs covering 38.7

Mb or 89% of chromosome 1 have been constructed and are presented in Figure 1 (Wu *et al.,* unpublished). Such a map constructed by *in silico* mapping is very useful for map-based cloning or for sequencing of genomic regions controlling an important trait. This is because the corresponding BAC clones can be directly utilized either for map-based cloning or for further study of the corresponding genomic region. This strategy may also be applied to wild relatives of cultivated *Oryza sativa.* However, the success rate of *in silico* mapping of the genome of wild species may be lower than that observed among cultivated varieties due to unusually low frequency of SNPs observed, while comparing *O. sativa* and *O. rufipogon* (Nasu *et al.,* 2002). If the intergenic region is rich in SNPs or InDels, sequences using Nipponbare as a standard can be a very efficient approach. Furthermore, *in silico* mapping may be difficult and less useful, when genome sequence are assembled from whole genome shot-gun strategy. There are two reasons for this; first, that the rice genome has many repetitive sequences, thus increasing the likelihood of assigning the sequences inaccurately, and second, that while using the shot-gun strategy, the physically mapped BAC clones containing important genes will not be available for any further research in the field of application genomics.

8.2 Standard Sequence for Interspecific Genome Analysis (Microcollinearity)

Due to non-availability of whole genome sequences in other cereals, sequence comparison between rice and other cereals is limited to specific regions of these other cereal genomes. However, several comparisons have been attempted in genomic regions containing important genes, thus elucidating, what is described as microcollinearity among cereal genomes. One such region containing the genes *shrunken2* (*sh2)/ anthocyaninless1 (a1)* has been examined in rice, sorghum and maize genomes, resolving homology among the corresponding regions of these three genomes (Chen *et al.,* 1997). In order to identify *sh2/a1* homologous regions in rice and sorghum, the maize DNA fragments containing *sh2* gene were used to screen rice and sorghum BAC libraries. Clones carrying the *sh2* gene were selected and the positive clones were screened for the presence of *a1* by hybridization with *a1* probe from maize (Civardi *et al.,* 1994; Chen *et al.,* 1998). Thus, since these clones were positive for both *sh2* and *a1*, it was inferred that both genes are close to each other in all the three species. Sequence analysis of about 40 kb of homologous regions from rice and sorghum genomes also resolved the same linear order of not only *sh2* and *a1*, but also that of two other genes, *X1*

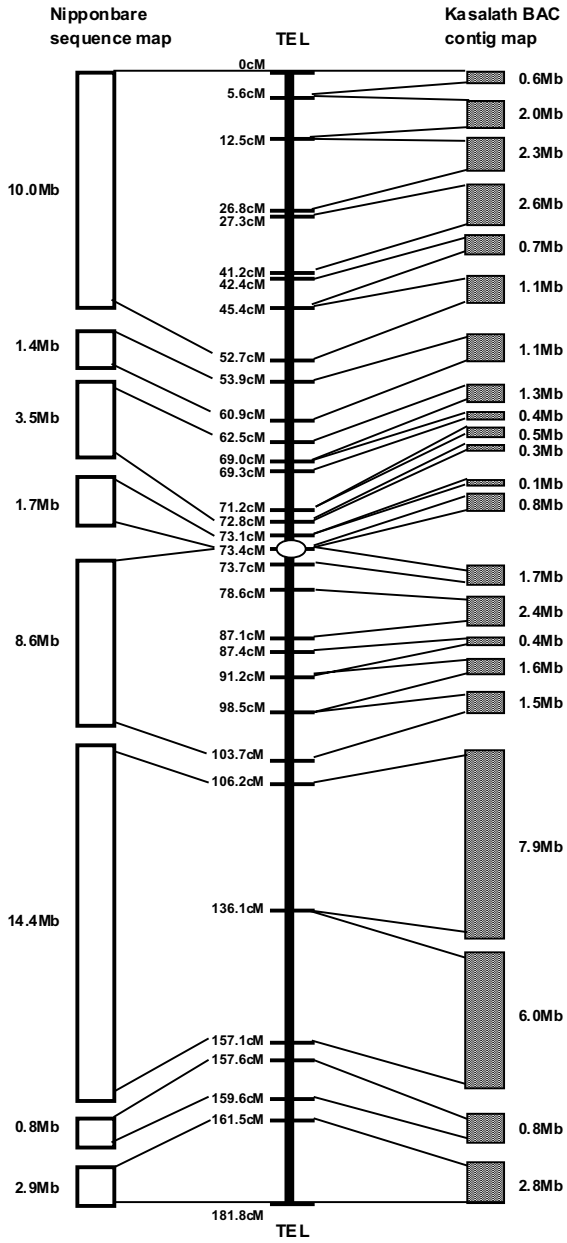


Figure 1. Results of *in silico* mapping of Kasalath chromosome 1 using the Nipponbare genome sequence as reference. The black vertical bar at the center indicates the genetic map of chromosome 1 (Harushima *et al.*, 1998) with the position of markers at the end of contigs for either Nipponbare or Kasalath. The contigs of Nipponbare BAC/PAC clones used for construction of the pseudomolecule are shown as open boxes on the left with the corresponding physical length (Mb). The black shaded boxes on the right indicate the contigs of the Kasalath BAC clones assigned by *in silico* mapping with their estimated physical length (Mb).

and *X2* found in the same region. In maize, the order of *sh2*, *X1, X2*, and *a1* was the same as in rice and sorghum, but the distance from *sh2* to *a1* was about 140 kb (Chen *et al.,* 1998). In sorghum, two homologous genes were found for *a1*, namely, *a1-a* and *a1-b* (Chen *et al.,* 1998), suggesting that duplication might have occurred after its divergence from the ancestral species. However, the intron size differed and the number of exons in *X1* is reduced. Except for this variation, the gene order of homologous genes in these three species is conserved. The wheat genome also carries *sh2/a1* genes, but in this genome, the linear order *sh2*, *X1*, *X2* and *a1* was found to be interrupted due to a break between *X1* and *X2* genes, resulting in a translocation of *sh2* and *X1* homologues to a nonhomoeologous chromosome (Li and Gill, 2002).

Microsynteny among grass species has also been detected in the chromosomal regions containing the *alcohol dehydrogenase 1 (Adh1)/ alcohol dehydrogenase 2 (Adh2)* genes. In case of rice, both are closely located on chromosome 11S, but in maize, *Adh1* and *Adh2* are located on chromosomes 1 and 4, respectively (Tarchini *et al.*, 2000). Judging from the collinearity of *Adh1/Adh2* region, it was inferred that *Adh1* must have been translocated after the evolutionary divergence of the two species. In addition, several predicted rice genes in rice could not be found in the corresponding genomic region of maize.

Microsynteny based on the rice genome sequence reveals how some genes are conserved although the intergenic regions diverged among different cereal genomes. This does not make the concept of 'synteny' obtained by genetic analysis of recombination invalid. The basic idea of 'synteny' is still useful in map-based cloning, if we pay attention to the characteristics of the target trait and confirm the presence of orthologous genes.

9. HIGH QUALITY SEQUENCES OF RICE CHROMOSOMES 1, 4, AND 10

The ultimate goal of the international rice genome sequencing consortium is to elucidate a finished-quality sequence of the entire genome. So far, the high quality sequences for three rice chromosomes have been completed and published (Sasaki *et al.,* 2002, Feng *et al.,* 2002, The Rice chromosome 10 Sequencing Consortium, 2003). Detailed analysis of the sequence has revealed characteristic features of the rice genome that were not detected in the draft sequence. Analysis of these rice chromosomes revealed 6,756 protein coding genes for chromosome 1, 4,658 for chromosome 4 and 3,471

genes (including transposable elements) for chromosome 10. The average gene density is almost similar in these chromosomes: 1 gene per 6.4 kb in chromosome 1, 1 gene per 7.4 kb in chromosome 4 and 1 gene per 6.46 kb in chromosome 10. However, these values are lower than the corresponding gene density for the Arabidopsis genome with 1 gene per 4.0 kb. The GC content of the rice genome is about 43% of the base pairs and GC-rich islands are randomly distributed along the chromosomes. Moreover, genes are not concentrated in GC-rich regions such that in chromosome 1, gene density is higher in distal region (18-19 genes per 100 kb) as compared to the proximal region near the centromere (10-12 genes per 100 kb). Rice chromosome 1 is characterized by several gene families that are dispersed or arranged in tandem repeats. For example, 15 receptor-like kinases occur as tandem repeats at the distal end of chromosome 1 (Sasaki *et al.,* 2002). Such duplication is also often observed in cereals other than rice, such as the case of storage protein kafirin of sorghum and zein in maize (Song *et al.,* 2002).

A comparison of the draft and finished sequence was made using a fraction of rice chromosome 10 covered by the *indica* sequence (The Chromosome 10 Sequencing Consortium, 2003). Although only about 4-9% of a given 1 Mb region of chromosome 10 was not covered in the draft sequences, analysis of the coding fraction revealed that most of the genes were interrupted in the draft sequence. This indicates that a large number of the predicted genes in the draft sequence represent gene fragments.

In whole shotgun sequencing, the repeat sequences are either masked or completely removed before assembly into contigs. An estimated 78 and 38 Mb of repeats were excluded in the BGI (Yu *et al.* 2002) and Syngenta (Goff *et al.* 2002) assembly, respectively. Analysis of the three rice chromosomes have also shown the presence of functional genes in repetitive sequences. Transposable elements embedded in these sequences can restructure the genome and control gene action, and some of the allelic variations selected in rice may actually be attributed to these elements. Therefore a high-quality finished sequence is indispensable in understanding gene regulation and complex biological processes.

10. SUMMARY AND OUTLOOK

Among cereals, the rice genome has been used and will continue to be used as a model genome both at the intraspecific and interspecific levels for a variety of purposes including evolutionary studies, gene discovery, map based cloning, etc. A number of dense genetic and physical maps (including BAC-based physical maps) of all the 12 rice chromosomes are now available

for comparative studies. Several draft sequences of the rice genome are also available and are being extensively used for study of other cereal genomes. Using this wealth of information, genes are being discovered and isolated in other cereals, although for rapidly evolving genes, like those for disease resistance, difficulties are encountered. A comparison of rice genome with other cereal genomes, however, suggested that duplication, and translocations of genomic segments occurred after the divergence of different grass species from a common ancestor. It has also been shown that transposable elements have been inserted into the cereal genomes especially to expand the size of the genomes of different members of Triticeae and Andropogoneae. These transposable elements complicate the identification of syntenous genes and comparisons at the sequence level.

Although straightforward and widespread, synteny among grass species is limited, when examined at the sequence level. In some cases, this will limit the utility of rice genome for gene discovery and cloning of agronomically important orthologous genes in other cereals. Nevertheless, the synteny/collinearity of DNA markers and that of DNA sequences in some regions provided a meaningful concept to elucidate evolution and divergence of grass species from a common ancestor. Detailed genome sequence is available only for cultivated species, *Oryza sativa*, subspecies *japonica*, variety Nipponbare and less extensively for *O. sativa*, subspecies *indica*, varieties 93-11 and Guangluai 4. Comparison of nucleotide sequence of 2.3 Mb stretch of chromosome 4 of *japonica* and *indica* genomes indicated that the predicted gene order was the same but gene number differed in these two subspecies (Feng *et al.,* 2002). Many insertions/deletions were observed, so that *japonica* had many more transposable elements than *indica* in this region. The sequence variation is much more than expected even within the most closely related rice subspecies, which are thought to diverge from wild *Oryza* species about 9000 years ago. Since almost all the cereals compared so far are cultivated species, analysis of a wild grass species may elucidate the effect of domestication on gene multiplication or segmental rearrangement.

The genome analysis will also help in our efforts for improvement of staple foods for yield and quality, which is a continuous process because neither the conditions of cultivation, nor the genomes of different cereals are stable. Crop improvement programmes have to be targeted to the need of adaptations to a variety of biotic and abiotic stresses. We must continue to strive towards understanding the evolution and divergence of cereal crops based on nucleotide sequence and to create viable strategies on how to incorporate or delete the target gene(s) to breed varieties of cereal crops that would be more productive. Rice genome will continue to serve as a model for this goal, although the target species itself has to be directly analyzed to extract

information that is unique to individual species and genotypes. With the completion of the rice genome sequence, analysis of the gene-rich regions of maize and wheat genomes by sequencing will undoubtedly enrich our knowledge of synteny among grass species.

11. REFERENCES

Ahn S., Tanksley S.D. Comparative linkage maps of the rice and maize genomes. Proc Nat Acad Sci USA 1993; 90: 7980-7984

Arumuganathan K., Earle E.D. Nuclear DNA content of some important plant species. Plant Mol Biol Rep 1991; 9: 208-218

Barker N.P, Linder H.P., Harley E. Phylogeny of Poaceae based on *rbcL* sequences. Syst Bot 1995; 20: 423-435

Barry G.F. The use of the Monsanto draft rice genome sequence in research. Plant Physiol 2001; 125: 1164-1165

Bent A.F., Kunkel B.N., Dahlbeck D., Brown K.L., Schmidt R., Giraudat J., Leung J., Staskawicz B.L. *RPS2* of *Arabidopsis thaliana*: a leucine-rich repeat class of plant disease resistance genes. Science 1994; 265: 1856-1860

Brueggeman R., Rostoks N., Kudrna D., Killian A., Han F., Chen J., Druka A., Steffenson B., Kleinhofs A. The barley stem rust-resistance gene *Rpg1* is a novel disease-resistance gene with homology to receptor kinases. Proc Nat Acad Sci USA 2002; 99: 9328-9333

Chen M., San Miguel P., de Oliveira A.C., Woo S.S., Zhang H., Wing R.A., Bennetzen J.L. Microcolinearity in *sh2*-homologous regions of the maize, rice, and sorghum genomes. Proc Nat Acad Sci USA 1997; 94: 3431-3435

Chen M., Presting G., Barbazuk W.B. , Goicoechea J.L., Blackmon B., Gang G., Kim H., Frish D., Yu Y., Sun S., *et al.* An integrated physical and genetic map of the rice genome. Plant Cell 2002; 14: 537-545.

Chen M., SanMiguel P., Bennetzen J.L. Sequence organization and conservation in *sh2/a1*-homologous regions of sorghum and rice. Genetics 1998; 148: 435-443

Civardi L., Xia Y., Edwards K.J., Schnable P.S., Nikolau B.J. The relationship between genetic and physical distances in the cloned *a1-sh2* interval of the *Zea mays* L. genome. Proc Nat Acad Sci USA 1994; 91: 8268-8272

Clayton W.D., Renvoize S.A. *Genera Graminum-Grasses of the World (Kew Bulletin Additional Series XIII),* Her Majesty's Stationery Office, London, United Kingdom, 1986

Devos K.M., Gale M.D. Comparative genetics in the grasses. Plant Mol Biol 1997; 35:3-15

Devos K.M., Gale M.D. Genome relationships: The grass model in current research. Plant Cell 2000; 12 : 637-646

Feng Q., Zhang Y., Hao P., Wang S., Fu G., Huang Y., Li Y., Zhu J., Liu Y., Hu X., *et al.,* Sequence and analysis of rice chromosome 4. Nature 2002; 420: 316-320

Gale M.D., Devos K.M. Plant comparative genetics after 10 years. Science 1998; 282: 656-659

Gale M.D., Youssefian S. Dwarfing genes in wheat. In: G.E. Russell (ed) *Progress in Plant Breeding*. Butterworths, London, United Kingdom, 1985, pp. 1-35

Goff S.A., Ricke D., Lan T.-H., Presting G., Wang R., Dunn M., Glazebrook J., Sessions A., Oeller P., Varma H., *et al.,* A draft sequence of the rice genome (*Oryza sativa* L.ssp. *japonica*). Science 2002; 296: 92-100

Grant M.R., Godiard L., Straube E., Ashfield T., Lewald J., Sattler A., Innes R.W., Dangl J.L. Structure of Arabidopsis RPM1 gene enabling dual specificity disease resistance. Science 1995; 269: 843-846

Han F., Kilian A., Chen J.P., Kudrna D., Steffenson B., Yamamoto K., Matsumoto T., Sasaki T., Kleinhofs A. Sequence analysis of a rice BAC covering the syntenous barley *Rpg1* region. Genome 1999; 42: 1071-1076

Harushima Y., Yano M., Shomura A., Sato M., Shimano T., Kuboki Y., Yamamoto T., Lin S.Y., Antonio B.A., Parco A., *et al.* A high-density rice genetic linkage map with 2275 markers using a single F2 population. Genetics 1998; 148: 479-494

Helentjaris T. Implications for conserved genomic structure among plant species. Proc Nat Acad Sci USA 1993; 8308-8309

Ikeda A., Ueguchi-Tanaka M., Sonoda Y., Kitano H., Koshioka M., Futsuhara Y., Matsuoka M., Yamaguchi J. *slender* rice, a constitutive gibberellin response mutant, is caused by a null mutation of the *SLR1* gene, an ortholog of the height-regulating gene *GAI/RGA/RHT/D8*. Plant Cell 2001; 13: 999-1010

Jacobs B.F., Kingston J.D., Jacobs L.L. The origin of grass-dominated ecosystems. Ann Missouri Bot Gard 1999; 86: 590-643

Jena K.K. and Khush G.S. production of monosomic alien addition lines of *Oryza sativa* having a single chromosome of *O. officinalis*. In: Rice Genetics, IRRI, Manila, Philippines, 1986, pp. 199-208

Khush G.S., Singh R.J., Sur S.C., Librojo A.L. Primary trisomics of rice: origin, morphology, cytology and use in linkage mapping. Genetics 1984; 107: 141-163

Kilian A., Chen J., Han F., Steffenson B., Kleinhofs A. Towards map-based cloning of the barley stem rust resistance genes *Rpg1* and *rpg4* using rice as an intergenomic cloning vehicle. Plant Mol Biol 1997; 35:187-195

Kurata N., Nagamura Y., Yamamoto K., Harushima Y., Sue N., Wu J., Antonio B.A., Shomura A., Shimizu T., Lin S.Y., *et al.* A 300 kilobase interval genetic map of rice including 883 expressed sequences. Nature Genet 1994a; 8: 365-372

Kurata N., Moore G., Nagamura Y., Foote T., Yano M., Minobe Y., Gale M.D. Conservation of genome structure between rice and wheat. Bio/Technology 1994b; 12: 276-278

Leister D., Kurth J., Laurie D.A., Yano M., Sasaki,T., Devos K., Graner A., Li W., Gill B.S. The colinearity of the *Sh2/A1* orthologous region in rice, sorghum and maize is interrupted and accompanied by genome expansion in the Triticeae. Genetics 2002; 160: 1153-1162

Leister D, Kurth J, Laurie DA, Yano M, Sasaki T, Devos K, Graner A, Schulze-Lefert P. Rapid reorganization of resistance gene homologues in cereal genomes. Proc Nat Acad Sci USA 1998; 95: 370-375

Linder H.P., Ferguson I.K. On the pollen morphology and phylogeny of the Restionales and Poales. Grana 1985; 24: 65-76

Lu B.R., Jackson M. *Wild Rice Taxonomy,* http://www.knowledgebank.irri.org/wildRiceTaxonomy/WildRiceTax.htm), International Rice Research Institute, 2002

Martin G.B., Brommonschenkel S.H., Chunwongse J., Frary A., Ganal M.W., Spivey R., Wu T. Earle E.D. Tanksley S.D. Map-based cloning of a protein kinase gene conferring disease resistance in tomato Science 1993, 262: 1432-1436

McCouch S.R., Kochert G., Yu Z.H., Wang Z.Y., Khush G.S., Koffman W.R., Tanksley S.D. molecular mapping of rice chromosomes. Theor Appl Genet 1988, 76: 815-829.

McCouch S.R., Tanksley S.D. "Development and use of restriction fragment length polymorphism in rice breeding and genetics." In: G.S. Khush, G.H. Toenniessen (eds.) *Rice Biotechnology*, CAB International., Wallingford, U.K. 1991

Mindrinos M., Katagiri F., Yu G.L., Ausubel F.M. The *A.thaliana* disease resistance gene *RPS2* encodes a protein containing a nucleotide-binding site and leucine-rich repeats. Cell 1994; 78: 1089-1099

Misra R.N., Jena K.K., Sen P. Cytogenetics of trisomics in indica rices. In *Rice Genetics*, 1986, IRRI, Manila, Philippines, pp. 173-183.

Moore G. Cereal genome evolution: pastoral pursuits with 'Lego' genomes. Curr Op Genetics Dev 1995; 5: 717-724

Moore G., Gale M.D., Kurata N., Flavell R. Molecular analysis of small grain cereal genomes-current status and prospects. Bio/Technology 1993; 11: 584-589

Moore G., Aragon-Alcaide L., Roberts M., Reader S., Miller T., Foote T. Are rice chromosomes components of a holocentric chromosome ancestor? Plant Mol Biol 1997; 35: 17-23

Motto M., Marsan P.A. Construction and use of molecular maps in cereals. In Jain, S.M., Brar, D.S., Ahluwalia, B.S. "*Molecular techniques in Crop Improvement*". Kluwer Academic Publishers, Netherlands, 2002, pp. 347-370

Nasu S., Suzuki J., Ohta R., Hasegawa K., Yui R., Kitazawa N., Monna L., Minobe Y. Search for and analysis of single nucleotide polymorphisms (SNPs) in rice (*Oryza sativa*, *Oryza rufipogon*) and establishment of SNP markers. DNA Research 2002; 9: 163-171

Pan Q., Liu Y.-S., Budai-Hadrian O., Sela M., Carmel-Goren L., Zamir D., Fluhr R. Comparative genetics of nucleotide binding site-leucine rich repeat resistance gene homologues in the genomes of two dicotyledons: tomato and Arabidopsis. Genetics 2000; 155: 309-322

Peng J., Carol P., Richards D.E., King K.E., Cowling R.J., Murphy G.P., Harberd N.P. The Arabidopsis *GAI* gene defines a signaling pathway that negatively regulates gibberellin responses. Gene Dev 1997; 11: 3194-3205

Peng J., Richards D.E., Hartley N.M., Murphy G.P., Devos K.M., Flintham J.E., Beales J., Fish L.J., Worland A.J., Pelica F., *et al.,* 'Green revolution' genes encode mutant gibberellin response modulators. Nature 1999; 400: 256-261

Renwick J.H. The mapping of human chromosomes. Ann Rev Genet 1971; 5: 81-120

Sakata K., Antonio B.A., Mukai Y., Nagasaki H., Sakai Y., Makino K., Sasaki T. INE: a rice genome database with an integrated map view. Nucl Acid Res 2000; 28: 97-101

Sakata K., Nagamura Y., Numa H., Antonio B.A., Nagasaki H., Idonuma A., Watanabe W., Shimizu Y., Horiuchi I., Matsumoto T., *et al,* RiceGAAS: an automated annotation system and database for rice genome sequence. Nucl Acid Res 2002; 30: 98-102

Sasaki T., Yano M., Kurata N., Yamamoto K. The Japanese rice genome research program. Genome Res 1996; 6: 661-666

Sasaki T., Burr B. International rice genome sequencing project: The effort to completely sequence the rice genome. Curr Opin Plant Biol 2000; 3: 138-141

Sasaki T. Matsumoto T., Yamamoto K., Sakata K., Baba T., Katayose Y., Wu J., Niimura Y., Cheng Z., Nagamura Y., *et al.* The genome sequence and structure of rice chromosome 1. Nature 2002; 420: 312-316

Shantz H.L. The place of grasslands in the earth's cover of vegetation. Ecology 1954; 35: 143-145

Shure M., Wessler S., Fedoroff N. Molecular identification and isolation of the *Waxy* locus in maize. Cell 1983; 35: 225-233

Song R., Llaca V., Messing J. Mosaic organization of orthologous sequences in grass genomes. Genome Res 2002; 12: 1549-1555

Soreng R.J., Davis J.I. Phylogenetics and character evolution in the grass family (Poaceae): simultaneous analysis of morphological and chloroplast DNA restriction site character sets. Bot Rev 1998; 64: 1-85

TAGI, The Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 2000; 408: 796-815

Tarchini R., Biddle P., Wineland R., Tingey S., Rafalski A. The complete sequence of 340 kb of DNA around the rice *Adh1-Adh2* region reveals interrupted colinearity with maize chromosome 4. Plant Cell 2000; 12: 381-391

The Grass Phylogeny Working Group "A phylogeny of the grass family (Poaceae), as inferred from eight character sets." In: S.W.L. Jacobs, J. Everett (eds.) *Grasses: Systematics and*

Evolution, Commonwealth Scientific and Industrial Research Organization, Collingwood, Victoria, Australia, 2000, pp. 3-7

The Rice Chromosome 10 Sequencing Consortium. In-depth view of structure, activity, and evolution of rice chromosome 10. Science 2003; 300: 1566-1569

The Rice Full-Length cDNA Consortium. Collection, mapping, and annotation of over 28,000 cDNA clones from *japonica* rice. Science 2003; 301: 376-379

Wu T., Earle E.D., Tanksley S.D. Map-based cloning of a protein kinase gene conferring disease resistance in tomato. Science 1993; 262: 1432-1436

Wu J., Maehara T., Shimokawa T., Yamamoto S., Harada C., Takazaki Y., Ono N., Mukai Y., Koike K., Yazaki J., *et al.* A comprehensive rice transcript map containing 6591 expressed sequence tag sites. Plant Cell 2002; 14: 525-535

Yan L., Loukoianov A., Tranquilli G., Helquera M., Fahima T., Dubcovsky J. Positional cloning of wheat vernalization gene VRN1. Proc Nat Acad Sci USA 2003; 100: 6263-6268

Yoshimura S., Yamanouchi U., Katayose Y., Toki S., Wang Z.-X., Kono I., Kurata N., Yano M., Iwata N., Sasaki T. Expression of *Xa1*, a bacterial blight-resistance gene in rice, is induced by bacterial inoculation. Proc Nat Acad Sci USA 1998; 95: 1663-1668

Yu J., Hu,S., Wang J., Wong G. K.-S., Li S., Liu B., Deng Y., Dai L., Zhou Y., *et al.* A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science 2002; 296: 79-91

Yu Y.G., Buss G.R., Saghai Maroof M.A. Isolation of a superfamily of candidate disease-resistance genes in soybean based on a conserved nucleotide-binding site. Proc Nat Acad Sci USA 1996; 93: 11751-11756

Chapter 19

CEREAL GENOMICS RESEARCH IN THE POST-GENOMIC ERA

Mark E. Sorrells

Department of Plant Breeding, Cornell University, Ithaca, New York 14853, USA
E-mail: mes12@cornell.edu

1. INTRODUCTION

The release of the genome sequences for *Arabidopsis thaliana* and rice (*Oryza sativa* L.) has substantially altered the genomics landscape and it is likely that several other plant genomes will be sequenced over the next 5 to 7 years. Consequently, plant research is now strongly model-organism oriented and is increasingly driven by questions that can be addressed by whole genome sequence analyses and related technologies such as large-scale reverse and forward genetics, whole genome transcript/proteomics analysis, and high-throughput genotyping. Plant species for which there are fewer genomic sequences available are likely to be anchored to such model species using comparative genomics methodologies. Elucidation of structure-function relationships of genes and genomes is more efficient in model species and efficient methods of transferring that information to other species are vitally important for crop species with large, complex genomes or less research support. However, approaches unique to individual plant species may also need to be developed, because each species has its own unique attributes, and may differ from other species in the available level of genetic diversity, its uses, and information base.

In the post-genomic era, the molecular maps and the genomic sequences will be extensively utilized for a much broader field of research, described as comparative genomics, which involves estimation of structural and functional similarity among living organisms at some level of organization. Integration of information across multiple levels of organization will facilitate the

P.K. Gupta and R.K. Varshney (eds.), Cereal Genomics, 559-584.

identification of genes controlling traits of interest. Recent advances in this field of research have already dramatically changed research strategies. The complementarity of information available in different species lends further support and power to comparative analyses. This is because different species have evolved different alleles for same genes, new genes with different functions, and differential gene expression. In addition, humans have shaped the evolution of some species to their benefit and in the process focused on traits that may differ between species. However there are also a number of traits in common across species that are fundamental to domestication. Comparative analysis of the variation in these genes can reveal much about structural and functional relationships.

Comparative genomics research has several goals, which include the following: (i) to compare the organization of related genomes and to infer the basic processes of genome evolution, (ii) to transfer information from model species to other related species, and (iii) to integrate information on gene location and expression across species. Comparative maps using anchor probes mapped in multiple species have been a critical tool for information transfer among species, although this is now being extended through the use of whole genome or partial genome sequences. Consensus maps have also proven useful for increasing the number of markers available for comparison, especially for species with low polymorphism (see Chapter 3 by Varshney *et al.* in this book). Genetic dissection of individual traits, chromosome location of genes for these traits and their expression, and integration of information about metabolic pathways can facilitate the rational selection of candidate genes. Molecular maps also allow assessment of allelic diversity and relative value of individual alleles, so that superior alleles for genes of economic importance may be identified. Plant breeders can use this information to assemble the best alleles into superior crop varieties. Cereal genomics is a very broad field of research, much of which is comprehensively reviewed in other chapters of this volume. Because of the increasing importance of the model systems approach to cereal genomics this review will cover the results of comparative mapping, the principles of comparative genomics using examples of applications to cereal crops, and the impact of this research on our concepts of genome evolution.

2. CREATING LINKS BETWEEN GENOMES

A comparative study of the locations of genes within genomes of different species establishes orthology. While relationship by descent implies conservation of gene function, rapidly evolving genes may be exceptions to this genera rule. In the past, cDNA clones have been used as anchor probes

to identify orthologous loci across multiple species or across genomes within a polyploid species. Generally, cDNA clones represent genomic sequences that are sufficiently conserved and hybridize well to genomic DNA from species that belong to different genera or families. Clones of single- or low-copy genes are preferred because they reduce the likelihood of mapping a paralogue.

2.1. Garden Blots for Anchor Probes

Clones suitable for use as anchor probes are identified by screening anonymous cDNA clones on "garden blots" containing DNA of the species to be used for comparative mapping. Those clones that hybridize well to DNA of the species on the garden blot are then screened for DNA polymorphism between parents of mapping populations. Because of attrition in the screening and surveys, hundreds of clones must be screened to identify those that can be mapped in multiple species and give good genome coverage. In an important study, Van Deynze *et al.* (1998) screened over 1,800 probes on garden blots containing DNA of rice, maize, sorghum, sugarcane, wheat, barley and oat, so that 153 of these were selected as anchor probes. Out of the selected probes, the number mapped in each species will then depend on the polymorphism between the parents of the populations,. The cDNA's derived from etiolated leaf libraries used by Van Deynze *et al.* (1998) mostly coded for proteins indicative of heterotrophic activity, involved in the TCA cycle or the glycolytic pathway. A comparison of the frequency of cross-hybridisation on DNA from five species for cDNAs derived from barley, maize, oat, and rice showed that oat cDNA's hybridized successfully to multiple species much more frequently than that of the other species (Van Deynze *et al.,* 1998). The rice cDNAs hybridized to the fewest species even though those clones had been previously evaluated and mapped in rice.

2.2. Southern Blots and PCR

Southern hybridization using anchor probes has been a reliable method for evaluation of relationships among species and genera and for comparative mapping (Van Deynze *et al.,* 1998). This is because PCR-based fragment amplification may be an all-or-none reaction (dominant), may amplify non-orthologous loci, or inadequately sample sequence variation because of the specificity of the primers. To be useful for comparative mapping, a molecular marker must identify orthologous loci in two or more species and exhibit sufficient level of polymorphism within a species to facilitate determination

of map location. It is apparent that for PCR-based markers, these criteria are in direct conflict because DNA sequence variation is essential for polymorphism, whereas conservation of DNA sequence is essential for designing primers that function within and across species. However, recently ESTs containing simple sequence repeats (SSRs) have been recognized as a valuable source of molecular markers (EST-SSRs) (Kantety *et al.,* 2002; Morgante *et al*., 2002; Varshney *et al*., 2002). The EST-SSRs can be used to identify orthologous loci across species (Scott *et al.,* 2000; Thiel *et al*., 2003). DNA sequences containing conserved regions of a gene that flank a hypervariable region are most useful for designing PCR-based markers that can amplify orthologous gene fragments across species. Although microsatellite (SSR) markers derived from genomic libraries are more polymorphic than those from expressed genes (Cho *et al.,* 2000; Scott *et al.,* 2000; Eujayl *et al.,* 2001; Eujayl *et al.,* 2002; Gupta *et al*., 2003), genomic microsatellite markers generally will not amplify loci in species other than the one from which they originated. Yu *et al.* (2004) used sequence similarity analysis to identify 156 cross-species superclusters and 138 singletons for developing primer pairs that were then tested on the genomic DNAs of four grass species; barley, maize, rice and wheat. Primer pairs for 141 superclusters and 128 singletons produced PCR amplicons and 228 primers amplified DNA from two or more species. Like anchor probes, EST-SSRs can also be useful for identifying orthologous loci in the different genomes of polyploids. Mapping multiple loci with a probe or EST-SSR aids in the identification of homoeologous chromosomes, which is very useful for map construction. Over 13,000 untested EST-SSR primer sets have been designed for wheat (5425), barley (3036) and rice (4726) and published in GrainGenes (http://wheat.pw.usda.gov/ggpages/ITMI/EST-SSR) for anyone to evaluate on their materials. Researchers are requested to send any results from using those primers sets back to GrainGenes so that others can benefit from a community effort.

2.3. Consensus Maps

Low polymorphism is a major limitation for mapping in cereal crops such as wheat (Nelson *et al.,* 1995a, b), barley (Kleinhofs *et al.,* 1993), and oat (O'Donoughue *et al.,* 1995). This problem can be largely overcome by using hypervariable markers, various gel and non-gel based detection systems, and consensus maps. Consensus maps are important components of a comparative mapping strategy, especially for polyploids such as wheat, where map information from multiple genomes can be merged. Consensus maps bring together information on locus order from multiple maps for a particular species or from multiple genomes within a polyploid into a single

comprehensive map for each similar chromosome group (Nelson *et al.,* 1995 a, 1995b). This greatly amplifies the number of loci for comparison with maps of other species, thus increasing the resolution of the comparative map. Construction of consensus maps can present challenges and their reliability is directly proportional to the number of loci in common among the maps included. Computer programmes have been developed to create consensus maps (e.g. Qi *et al.,* 1996); however, if there are sufficient loci for alignment, two maps can be merged manually. Consensus maps are particularly useful for species with low polymorphism because parents of different populations are, to some extent, complementary as to which probes are polymorphic. Probes that identify multiple loci within a genome complicate construction of consensus maps because different loci may be polymorphic in different populations. The order of loci in a consensus map that are situated between common loci is, at best, an estimate. However, for most applications using low-resolution comparative maps, small differences in locus order on a consensus map are inconsequential, as the markers will always need to be mapped in the target population.

3. COMPARATIVE MAPS OF CEREAL CROPS

For the domesticated grasses, molecular genetic maps have been used for comparative genomics. While comparing the maps of different grasses, the recognition of conserved linkage blocks and their relationships with rice linkage groups have led to hypotheses about the basic organization of the ancestral grass genome (Moore *et al.,* 1995; Devos and Gale, 1997; Gale and Devos, 1998; Wilson *et al.,* 1999). This has also provided impetus for subsequent investigations to examine conservation in more detail. Comparative mapping, like many biological phenomena, becomes more complex as we acquire new knowledge and delve further using more sophisticated tools and techniques.

3.1. Resolution of Comparative Maps

Comparisons of genetic linkage maps are severely limited in their resolution by the number of orthologous loci detected and by population size. Early comparative maps (e.g. Hulbert *et al.,* 1990; Ahn and Tanksley, 1993; Ahn *et al.,* 1993; Kurata *et al.,* 1994; Moore *et al.,* 1995; Van Deynze *et al.,* 1995abc; Devos and Gale, 1997; Gale and Devos, 1998) greatly underestimated the complexity of genome relationships. Those low resolution

comparative maps are biased by the use of predominately single copy probes that do not sample multicopy regions, simplifying assumptions about collinearity, and overemphasizing gene rich regions (Bennetzen, 2000; Gaut, 2001; 2002). Later studies using higher density maps (Wilson *et al.,* 1999) and large-scale genome sequencing (e.g. Chen *et al.,* 1997; Tikhunov *et al.,* 1999) resolved more rearrangements. Wilson *et al.* (1999) described a higher resolution rice/maize (*Zea mays* L.) comparative map based on Southern mapping that detailed more than 20 rearrangements including chromosome duplications, inversions, and translocations. Those maps were based almost entirely on RFLP analyses but their resolution was well below what is required for accurate assessment of collinearity. An in-depth study of the patterns of genome duplication in maize revealed far more homology among maize chromosomes than was previously known (Gaut, 2001), thus underscoring the oversimplification of interspecies genome relationships in the early investigations and also emphasizing the need for reassessment of the evolutionary paradigm for grasses.

3.2. Reassessment of Collinearity using Genome Sequences

With the availability of the Arabidopsis and rice genome sequences, it has become feasible to utilize whole genome or partial genome sequences for comparative genomics research. The information generated will then have increased utility for evolutionary studies and for transferring information from model species to related large-genome species for enhancing crop improvement strategies. The use of comparative sequence analysis methods to cross reference genes between species makes it possible to greatly enhance the resolution of comparative maps, to study gene evolution patterns, to identify conserved regions between the genomes, and to facilitate interspecies gene cloning. An obvious question that arises is how transferable is map location between monocots and dicots? Devos *et al.* (1999) compared rice ESTs to the Arabidopsis genome sequence and assessed the collinearity between these model representatives of the monocot and dicot subclasses of flowering plants. Their comparisons of two regions of up to 3cM on Arabidopsis chromosome 1 and rice ESTs with homology to Arabidopsis genes from 10 BAC clones revealed little conservation even from regions containing closely linked genes in one of the species. Liu *et al.* (2001) also reported minimal collinearity between 20 Mb of rice sequence and the complete Arabidopsis genome sequence. The longest collinear sub-BAC level region with three or more rice hits and one or two interruptions of non-collinear hits, was 63kb for rice with a mean of 25kb.

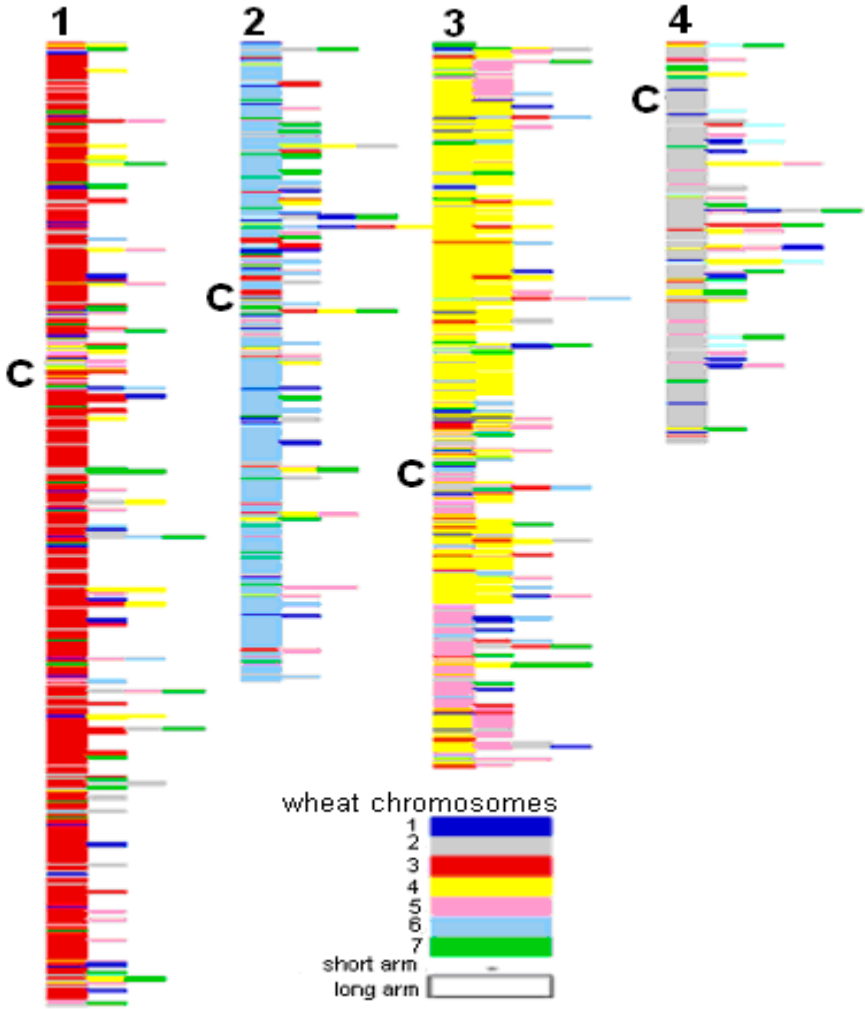


Figure 1. Rice-wheat genome relationships. Rice chromosomes showing the location on wheat chromosome arms for the most similar wheat gene sequences. Adjacent coloured rectangles indicate wheat genes mapped to multiple wheat genome locations. Wheat chromosome locations were based on homoeologous groups. Rice BAC/PAC sequences that did not match any wheat sequence as well as redundant matches were omitted. 'C' indicates the rice centromere location.

....continued

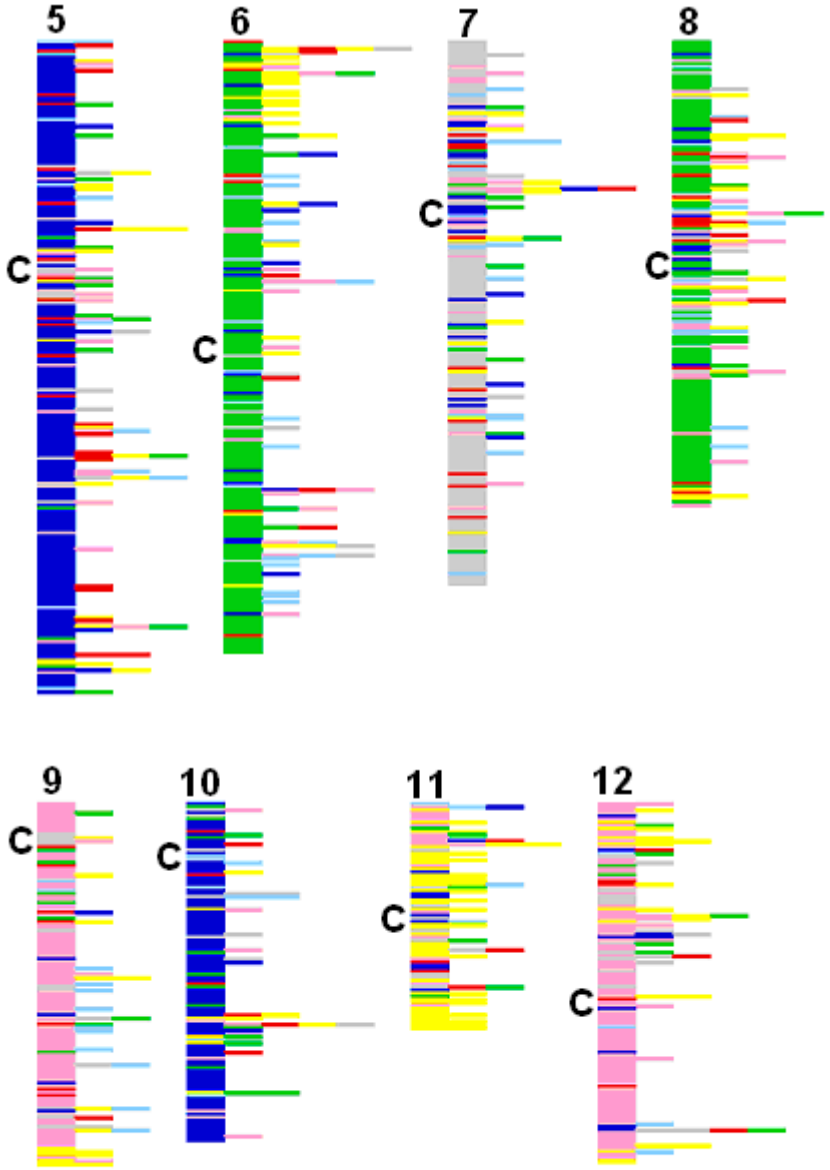


Figure 1. Continued

Recently, among the cereals, sequence-based comparative maps have been reported for sorghum chromosome 3 versus rice chromosome 1 (Klein *et al.,* 2003) and for the wheat/rice genomes (Sorrells *et al.,* 2003). Sequences from sorghum BAC contigs were subjected to BLASTX analysis to identify coding regions and these DNA sequences were compared with rice genome sequences. Although extensive conservation of gene content and order were observed, several small-scale changes in collinearity and one large rearrangement were identified. Sorrells *et al.,* (2003) compared 4485 ESTs that were physically mapped in wheat chromosome bins, to the public rice genome sequence data from 2251 ordered BAC/PAC clones using BLAST. A rice genome view of homoeologous wheat genome locations based on comparative sequence analysis revealed numerous apparent chromosomal rearrangements that complicate the use of rice as a model for cross-species transfer of information in non-conserved regions (Fig. 1). For most individual rice chromosomes, there was a strong relationship to one or two wheat homoeologous groups. Most of these genome relationships were apparent from earlier RFLP-based comparative maps (Kurata *et al.,* 1994; Van Deynze *et al.,* 1995abc; Sarma *et al.,* 2000), but the effective resolution was increased 25-30 fold. Compared to the RFLP-based maps, the rice-wheat genome relationships showed much more detail of the localized homoeology as well as a higher frequency of breakdown in collinearity throughout the genomes than previously reported (Sorrells *et al.,* 2003).

3.3. Complex Relationships Resolved by Comparison of Deletion Maps of the Wheat Genome with the Rice Whole Genome Sequence

La Rota and Sorrells (2004) presented a view of the wheat deletion map and its relationship to the rice genomic sequence (Table 1), which revealed partial conservation of gene content and order at the resolution conferred by the chromosome deletion breakpoints in the wheat genome. Wheat chromosome 3 was found to be the most conserved and wheat chromosome 5 was the least conserved, when compared with the rice genome sequence. Wheat chromosome 3 and rice chromosome 1 have many genes in common. However, using only single copy genes, even deletion bins in the most conserved regions often contained sequences homoeologous to more than one rice chromosome (Table 1). These analyses indicate that there has been an abundance of rearrangements, insertions, deletions, and duplications that, in many cases, will complicate the use of rice as a model for cross-species transfer of information in non-conserved regions. An analysis involving comparison of sequences of the maize genome with those of other cereals

Table 1. Wheat-rice genome relationships. Ordered wheat deletion bins for wheat chromosome groups 3 and 5 showing the number of gene matches to each rice chromosome. The number of matches utilizes only putative single copy genes in wheat. The number of matches in each table cell is bold and underlined when containing significant associations between wheat bins and rice chromosomes (Binomial distribution with $p<0.01$). Only wheat chromosome groups 3 and 5 are shown for brevity. Modified from La Rota and Sorrells (2004)

Wheat	Rice chromosome												Number of	EST
bin	1	2	3	4	5	6	7	8	9	10	11	12	Signif. Hits	total
3AS4-0.45-1.00	**46**	8	2	2	3		3	4	1	1			37	**107**
3AS2-0.23-0.45	**14**		2					1				1	6	**24**
C-3AS2-0.23	**12**		1					1			1		2	**17**
C3A	1	1											1	**3**
C-3AL3-0.42	**44**				2		2						6	**54**
3AL3-0.42-0.78	**54**			1	3	2	5	2	2	2		1	15	**87**
3AL5-0.78-1.00	**38**	4	3	1	3	1	2	3	1	2	1		45	**104**
3BS8-0.78-1.00	**12**	2		1		3						1	31	**50**
3BS9-0.57-0.78	**16**	2		2	3		1	1	1				22	**48**
3BS1-0.33-0.57	**25**	3	1		1	1	1			1		2	16	**51**
C-3BS1-0.33	**26**	1	1					3			1		17	**49**
C3B	2													**2**
C-3BL2-0.22	**43**			2	2		2	1				1	11	**62**
3BL2-0.22-0.50	**42**	1	1	1	1	2	3	2	1	3			17	**74**
3BL10-0.50-0.63	**16**	1	1		3						1	1	8	**31**
3BL7-0.63-1.00	**53**	4	4		4	1	2	2	1	2	2		55	**130**
3DS6-0.55-1.00	**30**	4	1	1	2			1	1				39	**79**
3DS3-0.24-0.55	**37**	3	4		1		1	1		1		2	27	**77**
C-3DS3-0.24	**11**	1			1		1	1			1		4	**20**
C3D	3												1	**4**
C-3DL2-0.27	**49**				3								12	**64**
3DL2-0.27-0.81	**85**	2	3	2	3	2	2	2	3	2	2	3	30	**141**
3DL3-0.81-1.00	**19**	2	2		2	1	1	1		2	1	1	39	**71**
5AS7/10-0.98-1.00												2	1	**3**
5AS3-0.75-0.98	4	2		3		1	1	1				**23**	19	**54**
5AS1-0.40-0.75	2	1									3	**16**	2	**24**
C-5AS1-0.40	1			1					1	1	2	**7**	8	**21**
C5A									1				1	**2**
C-5AL12-0.35					1		1		**7**		1	2	9	**21**
5AL12-0.35-0.57		1	3	1	1	1		1	**14**		1	3	6	**32**
5AL10-0.57-0.78	3	2	**23**			2	2		**12**		3	1	24	**72**
5AL17-0.78-0.87			**17**			1	3		1			1	17	**40**
5AL23-0.87-1.00	5		**18**	2	2	1	1	5			1	2	5	**42**
5BS6-0.81-1.00	4	1										**8**	10	**23**
5BS5-0.71-0.81		1	2							1		**6**	7	**17**
5BS8-0.56-0.71				1	1							**14**	5	**21**
5BS4-0.43-0.56	2								1		2	**21**	6	**32**
C-5BS4-0.43								1	1	1	2	2	5	**12**
C5B									1			1	1	**3**
C-5BL6-0.29	1					1	2		**7**			2	5	**18**
5BL6-0.29-0.55		1		1	1	1		1	**8**	1		1	5	**20**
5BL1-0.55-0.75		1						1	**13**		3	1	5	**24**
5BL14-0.75-0.76		1	**6**		1				4				16	**28**
5BL9-0.76-0.79	2		**30**			1	1				2	2	18	**56**
5BL16-0.79-1.00	5	1	**30**		2	1	6	1	6			1	38	**91**
5DS2-0.78-1.00	4	2				1	1					**7**	12	**27**
5DS5-0.67-0.78		1	1									3		**5**
5DS1-0.63-0.67												**6**	5	**11**
C-5DS1-0.63	1			1			1	2	1	1	3	**29**	11	**50**
C5D		1							1			1		**3**
C-5DL1-0.60	2	1	3	1	2	2	3	3	**36**	1	3	**12**	31	**100**
5DL1-0.60-0.74	2	3	2		1		2	2	**22**				18	**52**
5DL9-0.74-0.76			3					1	2		1		4	**11**
5DL5-0.76-1.00	5	1	**62**		2	2	4	2	3	1	1	2	49	**134**

will also be challenging, in view of the complexity of duplications in maize genome (Gaut, 2001). Comparative DNA sequence analysis can be complicated by several factors. For example, DNA sequence comparisons among species whose genomes have not been fully sequenced may not give a

correct picture, because a best sequence match may be actually a secondary match due to lack of sequence belonging to the best match. Only a complete genome sequence of both species would facilitate the identification of paralogues and evolutionary hot spots, where one of the genomes may be diverging more rapidly. Consequently, DNA sequence-based comparative maps dramatically increase the resolution of the RFLP-based maps and provide a much better estimate of the shortest conserved evolutionary unit sequence (O'Brien *et al.,* 1993).

3.4. Comparison of Wheat ESTs and Rice Genome Sequences

Different regions of the grass genomes are known to evolve at different rates (e.g. Dvorak *et al.,* 1998; Akhunov *et al.,* 2003; Gaut *et al.,* 2000; Draye *et al.,* 2001), thus complicating further the estimates of the average number of structural changes per million years of divergence; however, earlier estimates (0.14) appear to be too low (Paterson *et al.,* 1996). While occasional artifacts may arise from using "the best hits" between wheat ESTs and rice genomic sequence, the high stringency used in the wheat/rice comparisons may reduce such errors except where the orthologue diverges more rapidly or is deleted after duplication (Sorrells *et al.,* 2003; La Rota and Sorrells, 2004). In the event, where orthologues diverge more rapidly than prologues, a strategy incorporating a weighting factor for chromosome location together with sequence similarity will give a more robust estimate of conservation. However, the utility of the comparison will be compromised if gene function is not retained. Only a comparison of two completely sequenced genomes can generate a complete picture of genome evolution that elucidates relationships among all genes. Nevertheless, the enhanced resolution afforded by comparative DNA sequence analysis for rice and other cereal crops, especially in conserved regions, will facilitate the selection of markers for saturation mapping of a wheat chromosome region and for selecting candidate genes, both of which are important for developing functional molecular markers and for understanding Triticeae gene evolution.

3.5. Sequence-Based Comparisons Involving Wheat, Rice and Maize Genome Sequences

The sequence-based comparisons between wheat and rice genomes described above as well as recent studies of the *indica* and *japonica* rice subspecies (Feng *et al.,* 2002) and maize inbreds (Fu and Dooner 2002; Song and

Messing 2003) indicate that grass genomes may be evolving more rapidly than previously thought. Gaut (2002) recently reexamined the evolution of grass genomes with respect to their phylogeny. He reanalyzed previously published comparative map data as well as comparative sequence analyses and demonstrated that earlier reports overestimated genome conservation among species, especially those involving comparisons with the maize genome (Gaut, 2001). Gaut (2002) also utilized a phylogenetic analysis of the grasses (Kellogg, 2000) to illustrate conceptual problems that would arise, while using rice as an ancestral genome despite its small size and simple structure. He concluded that grass genomes are evolutionarily labile for many characteristics, including genome size and chromosome number, and that the current collinearity paradigm is in need of reassessment. Feng *et al.* (2002) analyzed DNA sequence alignments between 2.3Mb of three contiguous segments of chromosome 4 from the two rice subspecies, *indica* and *japonica*. Although there was extensive sequence collinearity, they identified 9,056 single-nucleotide polymorphisms (1 per 268bp) as well as 63 and 138 InDels (many in coding regions) for the *indica* and *japonica* sequences, respectively.

Fu and Dooner (2003) sequenced over 100 kb from the *bz1* genomic region of two different maize lines and found substantial differences between them. Retrotransposon clusters and genetic complement differed markedly between the two inbred lines, demonstrating that genetic microcollinearity can be violated even within the same species. Their results relate to the underlying genetic basis of hybrid vigour in maize, the meaning of "allelism", and the assessment of genetic distances. The implications are that lines lacking different genes would complement one another and exhibit hybrid vigour, but lines lacking most of the same genes would not complement and thus, would be placed in the same heterotic group. The breakdown in gene collinearity among maize inbreds was substantiated by Song and Messing (2003), but in addition, they showed that gene expression patterns in hybrids deviated significantly from the complementation expected. These studies suggest that there are genome-wide mechanisms effecting frequent sequence rearrangements that characterize these genomes at the megabase level and dramatically affect gene expression. Gene that are differentially expressed in wheat hybrids and their parents have also been identified through differential display technique (Wu *et al*., 2003).

4. MARKER ASSISTED SELECTION

The use of MAS in plant breeding has been the subject of a large number of publications in recent years (e.g. Lande and Thompson 1990; Hospital *et al.,*

1992; Dudley 1993; Langridge *et al.*, 2001; Koebner and Summers 2003; Morgante and Salamini 2003), however there have been few reports of successful variety or germplasm releases that specified the use of molecular markers for selection. This may be due to the paucity of public plant breeding programmes, the high cost of MAS, intellectual property restrictions, or other limitations (see Chapter 10 by Koebner in this book). Recent reports describing the use of MAS for the development of isolines or special genetic stocks include bacterial blight resistance in rice (Singh *et al.,* 2001), barley yellow dwarf resistance (Jefferies *et al.*, 2003) and malting quality (Han *et al.,* 1997) in barley and heading date in rice (Lin *et al.,* 2000). For several cereal crops, there are two major limitations to overcome for successful application of MAS; a lack of closely-linked, flanking markers that are polymorphic in elite germplasm and small QTL effects that are subject to genotype × environment interaction. PCR-based markers from related species are rarely functional unless they derive from highly conserved sequences. Wheat is the classic example of a species where markers are frequently the limiting factor; however, major efforts to alleviate this problem are having an effect (Langridge *et al.*, 2001; Chalmers *et al.*, 2001).

Implementation of MAS requires polymorphic markers that are tightly linked to the allele of interest. To be cost-effective, the markers should be adaptable to high-throughput detection systems. Because the molecular marker maps of most crop species are low resolution, the number of markers available is usually quite limited. This combined with the low polymorphism typical of elite germplasm used in breeding programmes, and the lack of tightly linked, polymorphic markers is the major limitation in using MAS. Low-resolution comparative maps can be used to identify homoeologous regions in the model species and DNA sequences from that species can then be used with BLAST analysis to identify within species sequences for marker development. With relatively high-resolution maps, it may be possible to identify candidate genes that are responsible for the trait of interest. In any case, the low polymorphism is a limitation that is difficult to overcome without considerable time and expense. One approach to circumventing the lack of polymorphism is to clone genomic DNA sequences close to or in the gene of interest. The sequences from the parents are compared and primers specific to DNA sequence differences between the parents are designed. In the case of polyploids, this can be somewhat challenging due to the multiple copies of homoeologous regions. This is because the polymorphism has to be unique not only between the parents but also genome specific. The ideal marker is a PCR-based, allele-specific assay where the primers are designed to amplify only those sequences that are responsible for the desired phenotype.

Another limitation of MAS is related to the fact that several genes with small effects control most traits of economic importance. These QTL are often influenced by the genetic background of the parent and interact with the environment. Comparative analysis of QTL can be useful for discovery and validation (Tuberosa *et al.,* 2002). In general, using today's technology, if one to three genes can explain half of the variation for a trait, then MAS has some value. If the trait is controlled by more than three genes or they explain less than half of the variation in the trait, then conventional breeding and selection techniques will likely be at least as efficient. One strategy for countering the limitation of selecting QTL with small effects is to screen germplasm accessions for alleles that have larger effects. Essentially, this is what transformation attempts to do; i.e., introduce a gene with a large beneficial effect. If the gene(s) controlling the trait are known then accessions can be screened for variation and classified according to their alleles using association analysis. A successful search for an allele with a major effect on a quantitative trait can have a large impact on a breeding programme and could facilitate the use of the marker or even direct phenotypic screening.

5. DOMESTICATION GENES

Genes that control traits of importance for a plant to survive in nature but limit the value of plants to humankind are often referred to as domestication genes. Such traits include seed shattering, inflorescence or fruit size, free threshing, lodging, specific colours of the plant, seed, or fruit, perenniality, etc. On the basis of correspondence of map position across related genera, Paterson *et al.* (1995) proposed that QTLs affecting plant height, seed size, non-shattering of grain, and photoperiod sensitivity are likely to be orthologous in sorghum, rice, and maize. In another report, Lin *et al.* (1995) presented evidence that quantitative trait loci (QTLs) with major effects on height and flowering in sorghum have counterparts in homoeologous segments of rice, wheat, barley and maize genomes. The capacity of a plant to produce rhizomes is a trait that confers perenniality and therefore led to domestication. For example, Hu *et al.,* (2003) reported that two genes in rice controlling rhizomatousness corresponded to QTL controlling this trait in *Sorghum propinquum*. Tillering in maize is another example of an important trait that differentiates it from its ancestor teosinte and thus contributed to domestication. The trait has been found to be controlled by several QTL, one of which was *tb1* that was later found to be involved in the control of lateral branch length and floral development (Doebley *et al.,* 1995; Hubbard *et al.,* 2002). Tillering genes have also been reported on chromosome 6H of barley (Babb and Muehlbauer, 2003) and chromosome 2 of rice (Jiabin *et al.,* 2001), but none of these appears to be related to *tb1* on maize chromosome 1.

Because barley chromosome 6H shares considerable homoeology to rice chromosomes 2, tillering genes in these two crops may be orthologous. However, there are examples of domestication traits apparently controlled by genes that are not orthologous across species. Paterson *et al.* (1995) detected an occasional lack of correspondence for domestication related QTLs across taxa. For example, in wheat and barley, genes affecting grain shattering due to fragile rachis have been mapped on chromosomes 2 and 3 of barley (Kandemir *et al.,* 2000) and 5A of wheat (Luo *et al.,* 2000). In contrast, shattering in rice involved loss of grain from the florets rather than fragile rachis (Shinya *et al.,* 1995), so that the shattering genes in rice are not orthologous to those in wheat and barley. Gale and Devos (1998) hypothesized that different species could have unique genes for the same domestication related trait to allow adaptation to the specific environmental conditions to which they are exposed.

6. POSITIONAL CLONING

One of the anticipated benefits of the complete rice genome sequence is its use for generating markers that can be used for high resolution comparative mapping to facilitate positional gene cloning in related grass species (see Chapter 11 by Stein and Graner in this book). Positional cloning depends on identifying markers whose genetic distance to the gene is within a large-insert clone, so that a library can be screened with that marker to identify a clone carrying the gene (Tanksley *et al.,* 1995). Thus, conservation of gene content and order at the megabase level (as well as a large-insert library in the target or closely related species) is essential for efficient use of a model species for this purpose. However, assessments of micro-collinearity between rice and members of Triticeae have revealed both conservation (Dubcovsky *et al.* 2001; 2003, Dunford *et al.,* 1995; Yan *et al.,* 2003) and intergenic breakages and segmental translocations (Kilian *et al.,* 1995; Han *et al.,* 1999; Li and Gill, 2002; Ramakrishna *et al.,* 2002; Bennetzen and Ramakrishna, 2002), the latter sometimes complicating the use of collinearity for positional cloning. In one of the more successful uses of comparative genetics, Yan *et al.,* (2003) used accessions of *T. monococcum* differing in vernalization response to fine map and clone the *Vrn1* gene. They reported almost perfect collinearity for rice, sorghum, and wheat for genes in the *Vrn1* region spanning approximately 0.1cM. Gene composition and order were also found to be conserved in the *adh1* region of maize and sorghum, but not in rice (Tikhonov *et al.,* 1999; Tarchini *et al.,* 2000). Duplications of loci separated by large genetic distances in different regions of the same chromosome can complicate comparative mapping, especially when polymorphism levels limit the number of fragments mapped in a given population (Chen *et al.,* 1997).

Gene duplication followed by sequence divergence and small translocations of single genes (Tarchini *et al.,* 2000), multigene families (Dubcovsky and Dvorak 1995) and the rapidly-evolving nature of certain genes, such as disease resistance genes (Leister *et al.,* 1998; Keller and Feuillet 2000) can all lead to rapid rearrangement of resistance-like genes and their nonsyntenic distribution in cereal genomes (Leister *et al.,* 1998). Although macro-colinearity does not always predict micro-collinearity, the level of conservation can be assessed simultaneously with fine mapping. However, the recent recognition of need to evaluate micro-collinearity for most situations has made the use of model species for positional cloning more complicated, due to the time and labor required for phenotyping and mapping of a large population for fine-scale analysis.

7. ALLELIC DIVERSITY

The use of molecular tools in a breeding programme is often limited by a lack of both molecular and phenotypic information for adapted germplasm. Comprehensive knowledge of the genetic diversity for agronomic traits, as well as for biotic and abiotic stress tolerance and grain quality traits is vital for identifying germplasm that will enhance crop improvement. Because of genotype × environment interaction, many of these traits must be evaluated in the target production region. In addition, marker assisted selection depends on detailed knowledge of the relationships between marker and trait alleles and the sources of the superior alleles. Genome-wide assessment of germplasm collections for allelic relationships is cost-prohibitive using today's technology; however, a representative collection of the most useful germplasm accessions that has been characterized for traits of interest and for variation in candidate genes can be quite valuable to a breeding programme. When combined with knowledge of gene or QTL location, diversity maps that show the distribution of allelic diversity (Dvorak *et al.,* 1998; Gaut *et al.,* 2000; Draye *et al.,* 2001; Tenaillon *et al.,* 2001) over a genome may become an effective tool for sampling variation.

A core collection of germplasm is useful for a variety of purposes including the following: (i) estimation of genetic diversity of targeted genes for specific traits, (ii) characterization of marker(s)-trait allele relationships for use in MAS, (iii) cross- referencing both gene and the genome locations with model species, and (iv) association analysis. Accurate ranking of the desirability of alleles is critical for selecting the source of a gene that will be used in cloning, transformation, or MAS because of the large investment that follows. Methods such as QTL analyses in bi-parental mapping populations are limited to comparisons of two alleles at each locus and genotype ×

environment interaction complicates the ranking of alleles across segregating populations. However, using either very closely linked markers or the DNA sequence of the gene responsible for the phenotype, association analysis can identify alleles in a collection of germplasm accessions and assess their diversity and relative value (Remington *et al.,* 2001; Thornsberry *et al.,* 2001) (see next section).

8. ASSOCIATION ANALYSES AND LINKAGE DISEQUILIBRIUM (LD)

Association analysis does not require a segregating population and may in some cases be more powerful than linkage analysis for identifying the genes responsible for the variation in a quantitative trait (Buckler and Thornsberry, 2002; for review see Flint-Garcia *et al.,* 2003). Combined with a correction for population structure (Prichard *et al.,* 2000), this methodology allows for large-scale assessment of allele/trait relationships. Comprehensive genome-wide scans for polymorphism are generally not practical for cereal genomes using today's technology, thus the alternative approach of focusing on variation in candidate genes or DNA markers closely linked to previously identified QTL has been widely used (e.g. Lander and Schork 1994; Remington *et al.,* 2001; Thornsberry *et al.,* 2001). The extent of LD around a locus determines the resolution of association analyses and the number of markers that would be required to scan the entire genome. Because genetic recombination is not evenly distributed over the genomes of most species, the linkage distance between markers and candidate genes varies widely. Simulations estimating the power of detecting the association of variation in a candidate gene with the phenotype indicate that population size is important (Long and Langley, 1999). For a population size of 500, there is a high probability of detecting the association, even when the gene accounts for as little as 8% of the variation. For a population size of 100, gene effects accounting for 15% of the variation can be detected.

A high degree of LD facilitates association analysis of markers linked to a QTL but genome-wide scans hinder the identification of candidate genes (for review see Flint-Garcia et al., 2003). In maize, the rapid decay of LD provides a means of mapping candidate genes with high precision and at the same time allows one to associate alleles with phenotypic values (Thornsberry *et al.,* 2001). For those species with high LD, comparative mapping and transcript profiling are necessary for narrowing the list of candidate genes. Various kinds of populations can be designed with the appropriate resolution. For example, segmental introgression lines will have

high LD while long-term breeding populations that have been intermated for many generations would have low LD.

9. EMERGING TECHNOLOGIES

High-throughput DNA sequencing and related technologies continue to improve at a rapid rate; however, the genome sequencing of most cereal species is still technically challenging. The maize genome sequencing is well underway and new methods are being developed to deal with the large amount of low-complexity DNA sequence in this large genome (http://www.maizegenome.org/). In November 2003, an international meeting sponsored by the National Science Foundation and the United States Department of Agriculture was held to discuss the need and a strategy for sequencing the wheat genome consisting of 16,000 Mbp (Gill and Appels, 2004). It was argued that the wheat genome sequence would provide a model for structural and functional changes that accompany polyploidy and that model species cannot be used to study the unique traits in wheat. Other genome sequencing candidates include barley (*Hordeum vulgare* L.), *Aegilops tauschii* L., and *Brachypodium distachyon*, a model grass with a genome size similar to Arabidopsis (Draper *et al.,* 2001).

Protein, metabolite and transcript profiling are critical technologies for understanding the genetic control of agronomic traits (see Chapters 15 and 16 of this book), especially in response to biotic and abiotic stresses. Recently, expression profiling has been used to study responses to drought in barley (Ozturk *et al.,* 2002), maize (Yu and Setter, 2003), and rice (Dubouzet *et al.,* 2003), as well as responses to hormones such as brassinosteroids (Yang et al. 2003) and abscisic acid (Lin *et al.,* 2003). Currently, the photolithographic oligonucleotide arrays are commercially available only for barley (http://www.affymetrix.com). High density oligo-arrays have been designed for Arabidopsis that can interrogate gene expression over the entire genome (http://signal.salk.edu), thus allowing the identification of all transcribed genes. Other applications include mapping “Single Feature Polymorphisms” in segregating populations and genome-wide association mapping using a set of divergent genotypes. Eventually, a haplotype map will be developed for different cereals that would greatly facilitate mapping. Most of the recent proteomic studies have involved characterization of proteins in specific tissues or organelles such as rice anthers (Kerim *et al.,* 2003), mitochondria (Heazlewood *et al.,*) and leaf sheaths (Shen-Shihua *et al.,* 2002) and wheat amyloplasts (Andon *et al.,* 2002). However, a recent study examined leaf proteins that were differentially expressed in tolerant and susceptible rice varieties under drought stress (Salekdeh *et al.,* 2002). Particularly interesting

are studies that have analyzed the segregation for transcriptional profiles (Schadt et al., 2003) or proteome expression and mapped them as quantitative trait loci (de Vienne et al., 2001; Consoli et al., 2002). This kind of analysis can be used to identify chromosomal regions that control the expression of suites of genes and when linked to phenotypes could greatly enhance our understanding of their genetic control. Of course the power of these technologies will be amplified when information is combined from multiple sources. Linking the results of association mapping with those from linkage mapping and expression profiling can lead to a better understanding of how allelic variation affects traits of interest as well as the distribution of genetic variation across the genome.

10. INTEGRATION OF INFORMATION FOR GENE DISCOVERY

Improved tools for cross-species, analyses are facilitating the integration of information from different sources as well as the transfer of information among species. Online databases (e.g. Graingenes: http://wheat.pw.usda.gov; ZMDB: http://www.zmdb.iastate.edu; Gramene: http://gramene.org) are a wealth of information for germplasm, genes, and maps, and they are presenting the information in novel ways that facilitate interpretation and utilization (see Chapter 14 by Mathews *et al*. in this book). One of the most exciting prospects is the visualization of integrated information about genes, their expression, metabolic pathways, genome location, and agronomic phenotypes. Many metabolic pathways have been elucidated using microorganisms and model species; however, much of the information is applicable to a broader range of species. The Kyoto Encyclopedia of Genes and Genomes (KEGG: http://www.kegg.com/) is one example of an impressive array of genomic information linked to metabolic pathways, regulatory pathways, and molecular assemblies. Ultimately, linking gene to phenotype is our goal.

Figure 2 illustrates the integration of information from various sources that contribute to the identification of the genes controlling a trait of interest and eventually to the understanding of their functions. For any trait of interest, we first need to know how many important genes control the trait and where they are located in the genome. QTL mapping is still the most common approach for acquiring that information, although analysis of various kinds of mutants and association analysis also contribute to that knowledge base. Once the location of the genes is known, we next want to learn their function. Knowledge of the metabolic pathway that might be involved may be possible to select a subset of candidate genes or ESTs that have been previously

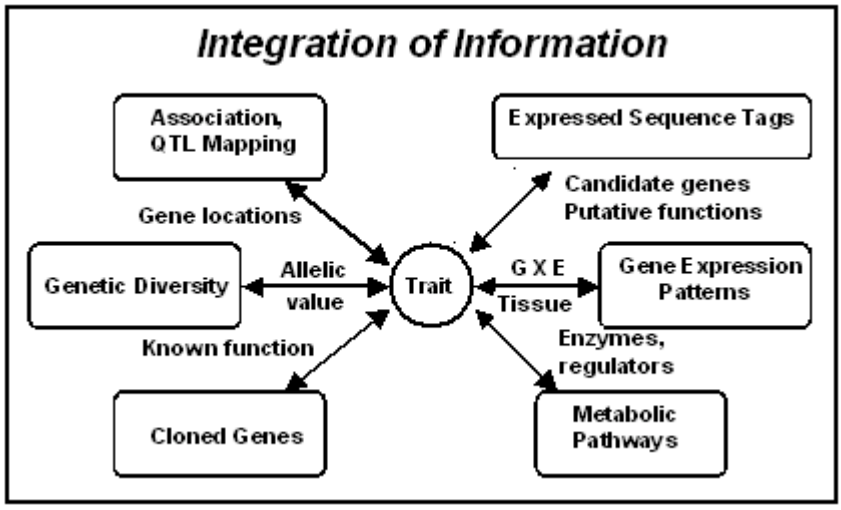


Figure 2. Integration of information from different sources is important for gene discovery and characterization that may lead to trait improvement.

located to that region of the genome. These may be genes that have been previously cloned and characterized or they might be ESTs that have been assigned a putative function. Supporting evidence for the candidate gene may be obtained from gene expression data that indicates tissue or developmental-stage specificity. Once there is ample evidence for the role of a particular gene, the final but most important step is to characterize allelic variation in the gene as described above.

11. SUMMARY AND OUTLOOK

As we move to the genomics model for systems biology, in which one begins with sequence data and then proceeds to determination of function, the key to understanding will be the ability to execute high throughput genotyping and precision phenotyping experiments. Genomics research has emphasized structural aspects in recent years; however, the focus is shifting to determining the functional role of genes and understanding the mechanisms of evolutionary change that have resulted in the diversity of living organisms we see today. Methods for genome-wide gene expression studies are also developing rapidly and are critical to our understanding of protein structure-

function relationships that are necessary for predicting gene function and for rationally engineering genes. Bioinformatics will play an increasingly important role in the integration of information from different species and sources through the use of novel approaches to analysis and visualization of complex data (for details consult Chapter 14 by Matthews *et al*. in this book). Structural genomics research linking genes and genomes across species benefits all species but is especially important for large-genome species as well as those that receive less funding. We have already gained a great deal of knowledge about biological systems and their never-ending complexity. Sorting out the components that can be easily and reliably manipulated is the challenge.

Breeders and geneticists must not lose site of our long-term goal, which is crop improvement (see Chapter 20 by Li and Gill in this book). Breeding progress depends on (i) discovery and generation of genetic variation for agronomic traits, (ii) development of genotypes with new or improved attributes due to superior combinations of alleles at multiple loci, and (iii) accurate selection of rare genotypes that possess the new improved characteristics. Consequently, efficient methods are needed for identifying and evaluating allelic effects on a large scale so that desirable alleles can be assembled in superior varieties. This can be facilitated by integration of genetic information across species, identification of superior alleles, and by focusing on the most important genes and traits for the species of interest. As we expand our knowledge of how genes evolve and interact to produce the nearly infinite range of phenotypes, new opportunities to manipulate genetic variation to the benefit of humankind will arise.

12. REFERENCES

Ahn S. Anderson J.A., Sorrells M.E., Tanksley S.D. Homoeologous relationships of rice, wheat and maize chromosomes. Mol Gen Genet 1993; 241: 483-490

Ahn S.N., Tanksley S.D. Comparative linkage maps of the rice and maize genomes. Proc Natl Acad Sci (USA) 1993; 90: 7980-7984

Akhunov E.D., Goodyear J.A., Geng S., Qi L.L., Echalier B., Gill B.S., Lazo G., Chao S., Anderson O.D., Linkiewicz A.M., *et al.* The organization and rate of evolution of the wheat transcriptome are correlated with recombination rates along chromosome arms. Genome Res 2003; 5: 753-763.

Andon N.L., Hollingworth S., Koller A., Greenland A.J., Yates J.R.I. Haynes P.A. Proteomic characterization of wheat amyloplasts using identification of proteins by tandem mass spectrometry. Proteomics 2002; 2: 1156-1168

Babb S., Muehlbauer G. Genetic and morphological characterization of the barley uniculm2 (*cul2*) mutant. Theor Appl Genet 2003; 106:846-857

Bennetzen J.L. Comparative sequence analysis of plant nuclear genomes: Microcolinearity and its many exceptions. Plant Cell 2000; 12: 1021-1029

Bennetzen J.L., Ramakrishna W. Numerous small rearrangements of gene content, order, and

orientation differentiate grass genomes. Plant Mol Biol 2002; 48: 821-827
Buckler E.S., Thornsberry J.M. Plant molecular diversity and applications to genomics. Curr Opin Plant Biol 2002; 5: 107-111
Chalmers K.J., Campbell A.W., Kretschmer J., Karakousis A., Henschke P.H., Pierens S., Harker N., Pallotta M., Cornish G.B., Shariflou M.R*., et al.* Construction of three linkage maps in bread wheat, *Triticum aestivum*. Aust J Agric Res 2001; 52: 1089-1119
Chen M., San Miguel P., de Oliveira A.C., Woo S.S., Zhang H., Wing R.A., Bennetzen J.L. Microcollinearity in *sh2*-homologous regions of the maize, rice, and sorghum genomes. Proc Natl Acad Sci USA 1997; 94: 3431–3435.
Cho Y.G., Ishii T., Temnkh S., Chen X., Lipovich L., McCouch S.R., Park W.D., Ayer N., Cartinhour S. Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (Oryza sativa). Theor Appl Genet 2000; 100: 713-722
Consoli L., Lefevre A., Zivy M., de Vienne D. Damerval C. QTL analysis of proteome and transcriptome variations for dissecting the genetic architecture of complex traits in maize. Plant Mol Biol 2002; 48: 575-581
de Vienne D., Bost B., Fievet J., Zivy M. Dillmann C. Genetic variability of proteome expression and metabolic control. Plant Physiol Biochem 2001; 39: 271-283
Devos K.M., Beales J., Nagamura Y., and Sasaki T. Arabidopsis–rice: Will collinearity allow gene prediction across the eudicot–monocot divide? Genome Res 1999; 9: 825–829
Devos K.M., Gale M.D. Comparative genetics in the grasses. Plant Mol Biol 1997; **35:** 3–15
Draper J., Mur L.A.J., Jenkins G., Ghosh-Biswas G.C., Bablak P., Hasterok R., Routledge P.M. Brachpodium distachyon. A new model system for functional genomics in grasses. Plant Physiol. 2001; 127: 1539-1555
Draye X., Lin Y.R., Qian X.Y., Bowers J.E., Burow G.B., Morrell P.L., Peterson D.G., Presting G.G., Ren S.X., Wing R.A., *et al.* Toward integration of comparative genetic, physical, diversity, and cytomolecular maps for grasses and grains, using the sorghum genome as a foundation. Plant Physiol 2001; 105:1225-1241
Doebley J., Stec A., Gustus C. Teosinte branched 1 and the origin of maize: evidence for epistasis and the evolution of dominance. Genetics 1995; 141: 333-346
Dubcovsky J., Dvorak J. Ribosomal RNA loci: Nomads in the Triticeae genomes. Genetics 1995; 140: 1367-1377
Dubcovsky J., Ramakrishna W., SanMiguel P., Busso C.S., Yan L., Shiloff B.A., Bennetzen J.L. Comparative sequence analysis of colinear barley and rice BACs. Plant Physiol 2001; 125: 1342-1353
Dubouzet J.G., Sakuma Y., Ito Y., Kasuga M., Dubouzet E., Miura S., Seki M., Shinozaki K. Yamaguchi S.K. OsDREB genes in rice, *Oryza sativa* L., encode transcription activators that function in drought-, high-salt- and cold-responsive gene expression. Plant Jour 2003; 33: 751-763
Dudley J.W. Molecular markers in plant improvement: Manipulation of genes affecting quantitative traits. Crop Sci 1993; 33: 660-668
Dunford R.P., Kurata N., Laurie D.A., Money T.A., Minobe Y., Moore G. Conservation of fine-scale DNA marker order in the genomes of rice and the Triticeae. Nucl Acid Res 1995; 23: 2724-2728
Dvorak J., Cheng L.M., Li Y.Z. Restriction fragment length polymorphism and divergence in the genomic regions of high and low recombination in self-fertilizing and cross-fertilizing Aegilops species. Genetics 1998; 148: 423-434
Eujayl I., Sorrells M.E., Baum M., Wolters P., Powell W. Assessment of genotypic variation among cultivated durum wheat based on EST-SSRS and genomic SSRS. Euphytica 2001; 119: 39-43
Eujayl I., Sorrells M.E., Baum M., Wolters P., Powell W. Isolation of EST-derived microsatellite markers for genotyping the A and B genomes of wheat. Theor Appl Genet 2002; 104: 399-407
Feng Q., Zhang Y.J., Hao P., Wang S.Y., Fu G., Huang Y.C., Li Y., Zhu J.J., Liu Y.L., Hu X.,

et al. Sequence and analysis of rice chromosome 4. Nature 2002; 420: 316-320
Flint-Garcia S.A., Thornsberry J.M. Buckler IV E.S. Structure of linkage disequilibrium in plants. Ann. Rev. Plant Biol. 2003; 54: 357-374
Fu H., Dooner H.K. Intraspecific violation of colinearity and its implications in maize. Proc Natl Acad Sci USA 2002; 99: 9573-9578
Gale M.D., Devos K.M. Comparative genetics in the grasses. Proc Natl Acad Sci USA 1998; 95: 1971-1974
Gaut B.S., Le Thierry d'Ennequin M., Peek A.S., and Sawkins M.C. Maize as a model for the evolution of plant nuclear genomes. Proc Natl Acad Sci USA 2000; 97: 7008–7015
Gaut B.S. Patterns of chromosomal duplication in maize and their implications for comparative maps of the grasses. Genome Res 2001; 11: 55-66
Gaut B.S. Evolutionary dynamics of grass genomes. New Phytol 2002; 154: 15-28
Gill B.S., Appels R. NSF/USDA/IGROW workshop on wheat genome sequencing. Genetics 2004; In review.
Gupta P.K., Rustgi S., Sharma S., Singh R., Kumar N., Balyan H.S. Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. Mol Gen Genom 2003; 270:315-323
Han F., Romagosa I., Ullrich S.E., Jones B.L., Hayes P.M., Wesenberg D.M. Molecular marker-assisted selection for malting quality traits in barley. Mol Breed 1997; 3: 427-437
Heazlewood J.L., Howell K.A., Whelan J. Millar A.H. Towards an analysis of the rice mitochondrial proteome. Plant. Physiol 2003; 132: 230-242
Hospital F., Chevalet C., Mulsant P. Using markers in gene introgression breeding programs. Genetics 1992; 132: 1199-1210
Hu F.Y., Tao D.Y., Sacks E., Fu B.Y., Xu P., Li J., Yang Y., McNally K., Khush G.S., Paterson A.H. Convergent evolution of perenniality in rice and soghum. Proc Natl Acad Sci USA 2003; 100:4050-4054
Hubbard L., McSteen P., Doebley J., Hake S. Expression patterns and mutant phenotype of teosinte branched1 correlate with growth suppression in maize and teosinte. Genetics 2002; 162: 1927-1935
Hulbert S.H., Richter T.E., Axtell J.D., Bennetzen J.L. Genetic mapping and characterization of sorghum and related crops by means of maize DNA probes. Proc Natl Acad Sci USA 1990; 87: 4251-4255
Jefferies S.P., King B.J., Barr A.R., Warner P., Logue S.J., Langridge P. Marker assisted backcross introgression of the Yd2 gene conferring resistance to barley yellow dwarf virus in barley. Plant Breed 2003; 122: 52
Jiabin T., Wanyong Z., Wenming W., Bingtian M., Yong L., Haojie L., Hongai X., Ping L., Lihuang Z. Genetic analysis and gene mapping of a rice few-tillering mutant in early backcross populations (*Oryza sativa* L.). Science in China Series C Life-Sciences 2001; 44: 570-575
Kandemir N., Kudrna D.A., Ullrich S.E., Kleinhofs A. Molecular marker assisted genetic analysis of head shattering in six-rowed barley. Theor Appl Genet 2000; 101: 203-210
Kantety R.V., La Rota M., Matthews D.E., Sorrells M.E. Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. Plant Mol Biol 2002; 48: 501-510
Keller B., Feuillet C. Colinearity and gene density in grass genomes. Trends Plant Sci 2000; 5: 246-251
Kellogg E.A. The grasses: A case study in macroevolution. Annu Rev Ecol Syst 2000; 31: 217-238
Kerim T., Imin N., Weinman J.J. Rolfe B.G. Proteome analysis of male gametophyte development in rice anthers. Proteomics 2003; 3: 738-751
Kilian A., Kudrna D.A., Kleinhofs A., Yano M., Kurata N., Steffenson B., Sasaki T. Rice-barley synteny and its application to saturation mapping of the barley Rpg1 region.

Nucl Acid Res 1995; 23: 2729-2733
Klein P.E., Klein R.R., Vrebalov J., Mullet J.E. Sequence based alignment of sorghum chromosome 3 and rice chromosome 1 reveals extensive conservation of gene order and one major chromosomal rearrangement. Plant Jour 2003; 605-621
Kleinhofs A., Kilian A., SaghaiMaroof M.A., Biyashev R.M., Hayes P., Chen F.Q., Lapitan N., Fenwick A., Blake T.K., Kanazin V., *et al*. A molecular, isozyme and morphological map of the barley (Hordeum vulgare) genome. Theor Appl Genet 1993; 86: 705-712
Koebner R.M.D., Summers R.W. 21st century wheat breeding: Plot selection or plate detection? Trends Biotech 2003; 21: 59-63
Kurata N., Moore G., Nagamura Y., Foote T., Yano M., Minobe Y., Gale M. Conservation of genomic structure between rice and wheat. Bio/Technol 1994; 12: 276-278
La Rota C.M., Sorrells M.E. Comparative DNA sequence analysis of mapped wheat ESTs reveals complexity of genome relationships between rice and wheat. Funct Intreg Genom 2004; 4: 34-46
Lande R., Thompson R. Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 1990; 124: 743-756
Lander E.S., Schork N.J. Genetic dissection of complex traits. Science 1994; 265: 2037-2048
Langridge P., Lagudah E.S., Holton T.A., Appels R., Sharp P.J., Chalmers K.J. Trends in genetic and genome analysis in wheat: a review. Aust J Agric Res 2001; 52: 1043-1077
Leister D., Kurth J., Laurie D.A., Yano M., Sasaki T., Devos K., Graner A., Schulze-Lefert P. Rapid reorganization of resistance gene homologues in cereal genomes. Proc Natl Acad Sci USA 1998; 95: 370-375
Li W., Gill B.S. The colinearity of the Sh2/A1 orthologous region in rice sorghum and maize is interrupted and accompanied by genome expansion in the Triticeae. Genetics. 2002; 160: 1153-1162
Lin Y.R., Schertz K.F., Paterson A.H. Comparative analysis of QTLs affecting plant height and maturity across the Poaceae, in reference to an interspecific sorghum population. Genetics 1995; 141: 391-412
Lin F., Xu S.L., Ni W.M., Chu Z.Q., Z.H. X. Xue H.W. Identification of ABA-responsive genes in rice shoots via cDNA macroarray. Cell Research 2003; 13: 59-68
Lin H.X., Yamamoto T., Sasaki T., Yano M. Characterization and detection of epistatic interactions of 3 QTLs, Hd1, Hd2, and Hd3, controlling heading date in rice using nearly isogenic lines. Theor Appl Genet 2000; 101: 1021-1028
Liu H., Sachidanandam R., Stein L. Comparative genomics between rice and Arabidopsis shows scant collinearity in gene order. Genome Res 2001; 11: 2020-2026
Long A.D., Langley C.H. The power of association studies to detect the contribution of candidate genetic loci to variation in complex traits. Genome Res 1999; 9: 720-731
Luo M.C., Yang Z.L., Dvorak J. The Q locus of Iranian and European spelt wheat. Theor Appl Genet 2000; 100: 602-606
Moore G., Devos K.M., Wang Z., Gale M.D. Cereal genome evolution-Grasses, line up and form a circle. Curr Biol 1995; 5: 737-739
Morgante M., Hanafey M. Powell W. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. Nature Genet 2002; 30: 194-200
Morgante M. Salamini F. From plant genomics to breeding practice. Curr Opin Biotech 2003; 14: 214-219
Nelson J.C., VanDeynze A.E., Autrique E., Sorrells M.E., Lu Y.H., Merlino M., Atkinson M., Leroy P. Molecular mapping of wheat. Homoeologous group 2. Genome 1995a; 38: 517-524
Nelson J.C., VanDeynze A.E., Autrique E., Sorrells M.E., Lu Y.H., Negre S., Bernard M., Leroy P. Molecular mapping of wheat. Homoeologous group 3. Genome 1995b; 38: 525-533
O'Brien S.J., Womack J.E., Lyons L.A., Moore K.J., Jenkins N.A., Copeland N.G. Anchored

reference loci for comparative genome mapping in mammals. Nat Genet 1993; 3: 103-112

O'Donoughue L.S., Kianian S.F., Ryapati P.J., Penner G.A., Sorrells M.E., Tanksley S.D., Phillips R.L., Rines H.W., Lee M., Fedak G., *et al.* A molecular linkage map of cultivated oat. (*Avena byzantina* cv. Kanota x *A. sativa* cv. Ogle). Genome 1995; 38: 368-380

Ozturk Z.N., Talame V., Deyholos M., Michalowski C.B., Galbraith D.W., Gozukirmizi N., Tuberosa R. Bohnert H.J. Monitoring large-scale changes in transcript abundance in drought- and salt-stressed barley. Plant Mol Biol 2002; 48: 551-553

Paterson A.H., Lan T.H., Reischmann K.P., Chang C., Lin Y.R., Liu S.C., Burow M.D., Kowalski S.P., Katsar C.S., DelMonte T.A. et al. Toward a unified map of higher plant chromosomes, transcending the monocot–dicot divergence. Nat Genet 1996; 14: 380–382

Paterson A.H., Lin Y.R., Li Z., Schertz K.F., Doebley J.F., Pinson S.R.M., Liu S.C., Stansel J.W., Irvine J.E. Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. Science 1995; 269: 1714-1718

Pritchard J.K., Stephens M., Rosenberg N.A., Donnelly P. Association mapping in structured populations. Am Jour Hum Genet 2000; 67: 170-181

Qi X, Stam P., Lindhout P. Comparison and integration of four barley genetic maps. Genome 1996; 39: 379-394

Ramakrishna W., Dubcovsky J., Park Y.-J., Busso C.S., Emberton J., SanMiguel P., Bennetzen J.L. Different types and rates of genome evolution detected by comparative sequence analysis of orthologous segments from four cereal genomes. Genetics 2002; 162: 1389-1400

Remington D.L., Thornsberry J.M., Matsuoka Y., Wilson L.M., Whitt S.R., Doebley J., Kresovich S.K., Goodman M.M., Buckler E.S. Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc Natl Acad Sci USA 2001; 98: 11479-11484

Salekdeh G.H., Siopongco J., Wade L.J., Ghareyazie B., Bennett J. Proteomic analysis of rice leaves during drought stress and recovery. Proteomics 2002; 2: 1131-1145

Sarma R.N., Fish L., Gill B.S., Snape J.W. Physical characterization of the homoeologous Group 5 chromosomes of wheat in terms of rice linkage blocks, and physical mapping of some important genes. Genome 2000; 43: 191-198

Schadt E.E., Monks S.A., Drake T.A., Lusis A.J., Che N., Colinayo V., Ruff T.G., Milligan S.B., Lamb J.R., Cavet G., *et al.* Genetics of gene expression surveyed in maize, mouse and man. Nature London 2003; 422: 297-302

Scott K.D., Eggler P., Seaton G., Rossetto M., Ablett E.M., Lee L.S., Henry R.J. Analysis of SSRs derived from grape ESTs. Theor Appl Genet 2000; 100: 723-726

Shen S., Matsubae M., Takao T., Tanaka N., Komatsu S. A proteomic analysis of leaf sheaths from rice. J Biochem 2002; 132: 613-620

Shinya O., Noriko S., Fumihiro F., Tasuke Y. Association between grain shattering habit and formation of abscission layer controlled by grain shattering gene sh-2 in rice (Oryza sativa L.). Jpn Jour Crop Sci 1995; 64: 607-615

Singh S., Sidhu J.S., Huang N., Vikal Y., Li Z., Brar D.S., Dhaliwal H.S., Khush G.S. Pyramiding three bacterial blight resistance genes (*xa5*, *xa13* and *Xa21*) using marker-assisted selection into indica rice cultivar PR106. Theor Appl Genet 2001; 102: 1011-1015

Song R., Messing J. Gene expression of a gene family in maize based on noncollinear haplotypes. Proc Natl Acad Sci USA 2003; 100: 9055-9060

Sorrells M.E., La Rota M., Bermudez-Kandianis C.E., Greene R.A., Kantety R., Munkvold J.D., Miftahudin M., Mahmoud A., Ma X., Gustafson P.J., *et al.* Comparative DNA sequence analysis of wheat and rice genomes. Genome Res 2003; 13: 1818-1827

Tanksley S.D., Ganal M.W., Martin G.B. Chromosome landing: a paradigm for map-based

gene cloning in plants with large genomes. Trends Genet 1995; 11: 63-68

Tarchini R., Biddle P., Wineland R., Tingey S., Rafalski A. The complete sequence of 340 kb of DNA around the rice Adh1-Adh2 region reveals interrupted colinearity with maize chromosome 4. Plant Cell 2000; 12: 381-391

Tenaillon M.I., Sawkins M.C., Long A.D., Gaut R.L., Doebley J.F., Gaut B.S. Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. Mays L.). Proc Natl Acad Sci USA 2001; 98: 9161-9166

Thiel T., Michalek W., Varshney R.K., Graner A. Exploiting EST databases for the developement and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.) Theor Appl Genet 2003; 106: 411-422

Thornsberry J.M., Goodman M.M., Doebley J., Kresovich S., Nielsen D., Buckler E.S. Dwarf8 polymorphisms associate with variation in flowering time. Nat Genet. 2001; 28: 286-289

Tikhonov A.P., SanMiguel P.J., Nakajima Y., Gorenstein N.M., Bennetzen J.L., Avramova Z. Colinearity and its exceptions in orthologous adh regions of maize and sorghum. Proc Natl Acad Sci USA 1999; 96: 7409-7414

Tuberosa R., Salvi S., Sanguineti M.C., Landi P., Maccaferri M. Conti S. Mapping QTLs regulating morpho physiological traits and yield: Case studies, shortcomings and perspectives in drought stressed maize. Ann Bot 2002; 89: 941-963

Van Deynze A.E., Dubcovsky J., Gill K.S., Nelson J.C., Sorrells M.E., Dvorak J., Gill B.S., Lagudah E.S., McCouch S.R., Appels R. Molecular-genetic maps for group 1 chromosomes of Triticeae species and their relation to chromosomes in rice and oat. Genome 1995a; 38: 45-59

Van Deynze A.E., Nelson J.C., O'Donoughue L.S., Ahn S.N., Siripoonwiwat W., Harrington S.E., Yglesias E.S., Braga D.P., McCouch S.R., Sorrells M.E. Comparative mapping in grasses. Oat relationships. Mol Gen Genet 1995b; 249: 349-356

Van Deynze A.E., Nelson J.C., Yglesias E.S., Harrington S.E., Braga D.P., McCouch S.R., Sorrells M.E. Comparative mapping in grasses. Wheat relationships. Mol Gen Genet 1995c; 248: 744-754

Van Deynze A.E., Sorrells M.E., Park W.D., Ayres N.M., Fu H., Cartinhour S.W., Paul E., McCouch S.R. Anchor probes for comparative mapping of grass genera. Theor Appl Genet 1998; 97: 356-369

Varshney R.K., Thiel T., Stein N., Langridge P., Graner A. *In silico* analysis on frequency and distribution of microsatellites in ESTs of some cereal species. Cell Mol Biol Lett 2002; 7: 537-546

Wilson W.A., Harrington S.E., Woodman W.L., Lee M., Sorrells M.E., McCouch S.R. Can we infer the genome structure of progenitor maize through comparative analysis of rice, maize, and the domesticated panicoids? Genetics 1999; 153: 453-473

Yan L., Loukoianov A., Tranquilli G., Helguera M., Fahima T., Dubcovsky J. Positional cloning of the wheat vernalization gene VRN1. Proc Natl Acad Sci USA 2003; 100: 6263-6268

Yang G., Matsuoka M., Iwasaki Y., Komatsu S. A novel brassinolide-enhanced gene identified by cDNA microarray is involved in the growth of rice. Plant Mol Biol 2003; 52: 843-854

Yu J.K., Singh S., Dake T.M., Benscher D., Gill B.S., Sorrells M.E. Development and mapping of EST-derived simple sequence repeat (SSR) markers for hexaploid wheat. Genome 2004: in press

Yu L.X., Setter T.L. Comparative transcriptional profiling of placenta and endosperm in developing maize kernels in response to water deficit. Plant Physiol 2003; 131:568-582

Chapter 20

GENOMICS FOR CEREAL IMPROVEMENT

Wanlong Li and Bikram S. Gill*

WGRC, Department of Plant Pathology, Throckmorton Plant Science Center, Kansas State University, Manhattan, KS 66506-5502, USA
*Author for correspondence: bsg@ksu.edu

1. INTRODUCTION

Cereals, including wheat, rice, maize, sorghum, barley, rye, oats and millets constitute the staple food of the world since their domestication approximately 10,000 years ago. They are the most important cultivated plants for food production and acreage, providing over 60% of the calories and proteins in our daily diet. Most likely, they will remain as a major food source in the foreseeable future. Therefore, any constraints on cereal production directly impact world food security.

Cereal production must keep pace with an ever-increasing world population. During the past four decades, world population doubled without a significant shortage of food. Thanks to the "green revolution", the cereal production system was intensified, leading to a significant increase in yield per unit land area and time through adoption of new genotypes of cereals (semi-dwarf varieties of wheat and rice and improved maize hybrids), through increased application of N fertilizer and massive investment in irrigation infrastructure.

In the 21st century, the challenge in cereal production is further exacerbated by shrinking resources and climate changes. World population will increase by 50% and reach 9.3 billion by the year 2050 (www.fao.org). Therefore, the challenge for the next 50 years is not only feeding 3 billion more people, but

P.K. Gupta and R.K. Varshney (eds.), Cereal Genomics, 585-634.

also that of coping with the limited, even reduced resource base. Although the yield increase during the last few decades was mainly due to an intensified cereal-production system, the acreage for wheat, rice and maize also expanded by 5%, 17% and 15%, respectively. There is little room for further expansion of agriculture into the natural ecosystem. On the contrary, arable land is shrinking at an astonishing rate in the developing countries because of soil erosion caused by overploughing, overgrazing and deforesting, and rapid industrialization. A second limitation is non renewable resources, such as water, phosphorus, and potassium. The lack of water, a result of population growth and increase in demand for food production, will be a very serious problem that severely threatens modern irrigation agriculture. Another factor challenging cereal production is global warming, which is dramatically changing the world climate. It is estimated that the average temperature will increase by 2°C in 50 years. Together with the erosion of soil and other ecological factors, climate change would cause degradation of biodiversity. The question therefore, is how to meet the increasing demand for cereal production without sacrificing our living environment.

Science and technology are an active component of productivity. Fortunately, we are living during an explosion of information and knowledge in the field of life sciences. Genomics has been the new frontier and has profoundly changed many aspects of life science during the last two decades. It also promises to play an important role in cereal improvement. First, genomics deepens our understanding about gene networks for cereal development and agronomy, through the available molecular maps, genomic and EST sequences and pathways of genes and information about QTLs. Comparative genomics studies have transformed grasses into a single genetic system. Information, such as collinearity and gene function, gained from one cereal crop also benefits the improvement of other cereals. Second, genomics is revolutionizing breeding methodology through marker-assisted selection and directed mutagenesis, which significantly enhance the efficiency of breeding for improvement of agronomical traits. Third, genomics accelerates cereal biotechnology by providing more native target genes. Many agronomic traits are under control of genes with unknown functions, which can be mapped and cloned based on their position on genetic maps (map-based or positional cloning). The cloned genes, containing their own exons, introns and regulatory elements, are ready to be transformed into other varieties of the same crop or into other cereals without additional modification.

In the different chapters of this book, the progress of cereal genomics has been reviewed in its many aspects: genome organization, marker technology, map-based cloning, genome sequencing, functional genomics of the response of cereals to biotic and abiotic stresses and bioinformatics. This chapter will focus on application of genomics to several important goals involved in cereal improvement.

2. GENOMICS FOR GERMPLASM ENHANCEMENT

Germplasm is the material basis for the continuous improvement of crops. Traditionally, germplasm identification and assessment was based on phenotyping. Because the number of morphological traits available is limited and the expression of agronomic traits, most of them under control of multiple loci, is liable to fluctuations in the environment or restricted by quarantine, identifying and quantifying the genetic variation contained in germplasm collections was difficult. This led to under-representation or redundancy of collections in gene banks and hindered exploitation of the full potential of germplasm in crop improvement. Recent years have seen a transition in germplasm studies from a traditional phenotype-based approach to marker-based and gene-based approaches (Tanksley and McCouch, 1997). Various molecular markers gave comparable results when used for germplasm evaluation and identification. Because of the relative ease and rapidity in experimental manipulation, PCR-based markers are used more than RFLP markers (see Chapter 2 by Somers and Chapter 3 by Varshney *et al.* in this book). SSRs are more popular than the other PCR-based markers for their hypervariability, frequent codominance and locus-specificity (for a review see Gupta and Varshney, 2000). Several researchers have suggested that a relatively small number of SSR markers could provide enough information for genotyping (Ni *et al.,* 2002; Pestsova *et al.,* 2000; Senior *et al*., 1998; Prasad *et al.,* 2000; Zhang *et al.,* 2002). Multiplexing of fluorescence-labeled PCR products in combination with an automatic sequencing system have increased greatly the throughput, efficiency and accuracy of germplasm genotyping with SSR markers (Coburn *et al.,* 2002). Microarray- or capillary-based SNP haplotyping will maximize germplasm potential via allele mining.

2.1. Diversity in Landraces vs Cultivars

The scientific community and public were concerned about the genetic vulnerability of cereal crops and the genetic erosion of landraces during the past 30 years. Molecular marker evaluations indicated that genetic diversity among varieties or inbred lines is higher than expected, although it is lower than that among the landraces. In maize, an open pollinated crop, data from RFLP (Dubreuil and Charcoset, 1999), RAPD (Moeller and Scaal, 1999), AFLP (Lübberstadt *et al.,* 2000) and SSR (Senior *et al.,* 1998; Gethi *et al.,* 2002; Warberton *et al.,* 2002; Labate *et al.,* 2003) analyses demonstrated a high level of variation among and within populations. Genetic diversity within populations was found to be higher than that between populations (Warburton *et al.,* 2002). For maize inbred lines, genetic variation is significantly higher between heterotic groups than that within each group (Lübberstadt *et al.,* 2002). A comparative study of RFLPs showed that a large number of alleles present in maize landraces were absent in inbred lines (Dubreuil and Charcoset, 1999). Analysis of 21 SNP loci along chromosome 1 suggested that maize inbred lines retained 77% of genetic diversity of the landraces (Tenaillon *et al.,* 2001). Compared to historical inbred lines, diversity was reduced at the gene level in current inbred lines but not at population level (Lu and Bernardo, 2001).

RFLP and STS analyses revealed relatively narrow genetic diversity in modern breeding lines and modern varieties of common and durum wheats when compared with landraces (Autrique *et al.,* 1996; Chen et al., 1994; Röder *et al.,* 2003; Zhang *et al.,* 2002). Using 19 SSR markers, Röder *et al.* (2003) found 198 alleles in 502 European varieties, 280 alleles in 450 European landraces and 323 alleles in 544 non-European landraces. Of the 339 alleles found in 994 landraces, 147 are only present in landraces but not in varieties, suggesting a genetic similarity of 57% (Röder *et al.,* 2003). Analyses of wheat varieties released at different times revealed trends of genetic diversity in a country or a region. For instance, genetic diversity is decreasing in Iranian wheats (Sayed-Tabatabaei and Shohnejat-Bushehri, 2003), but remains constant in the wheat varieties of Argentina (Manifeto *et al.,* 2001), UK (Donini *et al.,* 2000) and the Yaqui Valley of Mexico (Souza *et al.,* 1994), which benefited from the green revolution.

Characterization of Chinese rice germplasm with SSRs showed that landraces contained the highest genetic diversity, 70% of which was maintained in

cultivars, and 28% of alleles were lost during genetic improvement (Yang *et al.,* 1994). On the contrary, genetic variation found in Indian traditional Basmati rice cultivars was much lower than in improved varieties (Nagaraju *et al.,* 2002). A similar situation was found in Chinese sorghum germplasm because of the single origin of the landraces and the introduction of new germplasms into the breeding program (Yang *et al.,* 1996). Similar levels of genetic diversity were found within the landraces and within breeding lines of sorghum from South Africa (Uptmoor *et al.,* 2003).

2.2. Diversity for Classification of Germplasm

In all cases, genomic analyses of cereal germplasm confirmed the botanical classification, ecological and heterotic groups, ecogeographical populations and pedigrees. In turn, marker technologies are used for parentage determination (Berry *et al.,* 2003), denomination (Yang *et al.,* 1996) and variety ownership protection (Troyer and Rocheford, 2002). In those cases where genes or QTLs of interest have been identified, tagged or cloned, germplasm can be classified unambiguously based on haplotyping. For example, a major QTL for resistance to Fusarium head blight disease was mapped to the short arm of chromosome 3B of a Chinese wheat variety Sumai 3. When five SSR markers closely linked with this QTL were used to genotype 71 resistant germplasm lines, eight haplotypes were identified and only seven lines were found to have the haplotype of Sumai 3 (Liu and Anderson, 2003).

2.3. Diversity for Study of Domestication and Collection Sites

Comparative genomics of cereal crops along with their wild ancestors have provided insight into the domestication and suggested guidelines for collecting, evaluating and utilizing germplasm. For instance, SSR genotyping demonstrated that maize (*Zea mays* subsp. *mays* L.) originated from a single domestication from teosinte (*Z. mays* subsp. *parviglumis*) in the central Balsas River drainage of Mexico ~9,000 years ago, and spread into North and South Americas after diversification in Mexican Highland (Matsuoka *et al.,* 2002). Similarly, AFLP fingerprinting of einkorn and emmer wheats and

barley, along with their wild progenitors, indicated that einkorn (Heun *et al.,* 1997) and emmer wheat (Özkan *et al.,* 2002) were domesticated in a very small area of southeastern Turkey near the Tigris and Euphrates rivers, and barley (Badr *et al.,* 2000) was domesticated in the Jordan Valley on the western edge of the Fertile Crescent >10,000 years ago. Although STS analysis of the wheat D genome suggested that multiple *Aegilops tauschii* parents were involved in the origin of common wheat (Talbert *et al.,* 1998), all wheat shares a single D-genome gene pool, *i.e*., the form *strangulata* of *Ae. tauschii* (Dvorák *et al.,* 1998a). Molecular characterization of *Ae. tauschii* accessions revealed that the *strangulata* gene pool is larger than expected because of the gene flow from the form *tauschii* of *Ae. tauschii* (Lubbers *et al.,* 1991; Dvorák *et al.,* 1998a). It was also shown that for *Ae. tauschii*, the highest diversity occurred in the Caucasian countries and the lowest diversity was found in Central Asian countries (Pestsova *et al.,* 2000). In the case of sorghum, a exceptionally high level of diversity was found in the sorghum landraces from Eritrean (Ghebru *et al.,* 2002), confirming that Abyssinia is the center of sorghum diversity. Two subspecies of Asian rice, *Oryza sativa* subsp. *indica* and subsp. *japonica* were domesticated separately from the annual and perennial strains of *O. rufipogon*, respectively (Cheng *et al.,* 2003).

2.4. Diversity within a Genome

Diversity within a genome is largely shaped by recombination and selection and is not homogenous. For example, in rice, Coburn *et al.* (2002) observed a higher level of polymorphism on chromosomes 10 and 11 and a lower level on the long arm of chromosome 4. Compared to *indica* rice, *japonica* rice showed a much higher level of diversity on chromosomes 6 and 7 (Ni *et al.,* 2002) and the short arm of chromosome 12 (Coburn *et al.,* 2002), but a considerable lower level of diversity on chromosomes 2 (Ni *et al.,* 2002) and 3 (Coburn *et al.,* 2002). In *Aegilops*, the polymorphism level of a locus was correlated with recombination rate along the centromere to telomere axis (Dvorák *et al.,* 1998b). Screening 470 SSR markers that were derived from maize ESTs and monomorphic in maize against a large number of maize landraces and subsp. *parviglumis*, Vigouroux *et al.* (2002) found that 15 SSRs were polymorphic in wild subspecies but monomorphic in landraces, indicating that the host genes are agronomically important and were under stringent selection during domestication. Comparison between landraces and

improved varieties of Chinese wheat revealed a significant difference in the level of diversity within the D genome (Zhang *et al.,* 2003), indicating that a high selection pressure was applied to the D genome during the breeding process.

2.4. Introgression of Alien Genes

For the long-term sustainability of cereal production, introducing alien or exotic genes from wild species is imperative. Many useful genes have been transferred from wild relatives into cereal crops, most of which are single genes conferring resistance to various diseases. Fifty-seven genes for resistance to diseases and pests were introduced into wheat from other genera of the Triticeae via alien translocations; the size of alien fragments and the translocation breakpoints were precisely determined by genomic in situ hybridization (for review see Friebe *et al.*, 1996). Alien genes introduced by homologous recombination can be mapped genetically with markers. Molecular characterization of the introgression lines led to the map-based cloning of disease resistance genes *Xa21* from rice (Song *et al.,* 1995) and *Lr21* from wheat (Huang L. *et al.,* 2003). However, most agronomically important traits, including yield, nutritional quality and stress tolerance, are controlled by multiple loci and inherited quantitatively. Although wild species are usually inferior to crops agronomically, transgressive variation was observed in crosses involving elite lines and wild species (see Tanksley and Nelson, 1996), suggesting that wild species contain genes useful for improvement of agronomic traits. Tanksley and Nelson (1996) proposed a strategy referred to "Advanced Backcross QTL (AB-QTL) analysis" integrating the QTL discovery with variety development by transferring useful QTLs from unadapted donors to elite lines. AB-QTL analysis has been widely adopted in cereals to transfer exotic or alien QTLs, and a large number of QTLs have been identified mainly for yield and yield components; 34-60% of the QTLs with favorable effect were derived from exotic germplasm. In rice, Xiao *et al.* (1998), Moncada *et al.* (2001) and Thomson *et al.* (2003) identified 68, 25 and 76 QTLs for 12, 4 and 13 agronomic traits, respectively, in advanced backcross populations between wild rice (*O. rufipogon*) and elite rice lines. Through genotyping 72 preselected BC_2F_2 plants derived from a cross between a German variety and a synthetic wheat, Huang X.Q. *et al.* (2003) identified 40 putative QTLs, including 11 for yield, 16 for yield components, eight for ear emergence and five for plant height.

Similarly, Pillen *et al.* (2003) identified 86 putative QTLs by genotyping a population of 136 BC_2F_2 plants derived from a cross between a barley variety and wild barley (*Hordeum vulgare* subsp. *spontaneum*). The AB-QTL strategy was also applied to crosses between two elite lines of maize (Ho *et al.* 2002) and also used to discriminate the closely linked QTLs in rice (Monna *et al.,* 2002b; Saito *et al.,* 2003; Takeuchi *et al.,* 2003).

3. GENOMICS FOR INCREASE IN YIELD

Grain yield is the most important and most complex agronomic trait. Physiologically, yield is determined by the source (photosynthetic organs, mainly the upper leaves), sink (mainly endosperm) and translocation capacity of assimilates. Mathematically, yield is a product of a series of components, plants per unit area, spikes or panicles per plant, grains per spike and grain weight. Genetically, yield is controlled by numerous genes, additive × additive (AA) (epistasis) and genotype × environment (GE) interactions. In maize, rice and sorghum, hybrids have given increased grain yield on a large scale due to heterosis. The genetic basis of heterosis has been an active area of research and has been debated for nearly a century. With the availability of high-density linkage maps and QTL technology, significant advances have been made in the field of genetic dissection of yield, yield components and heterosis, thus, greatly increasing our knowledge of factors related to yield improvement.

3.1. Yield and Yield Components

In cereals crops, a large number of studies have been conducted involving QTL analyses for yield or yield components. The major QTLs conditioning >10% of phenotypic variation (PV) detected with a LOD >3.0 and confirmed in multiple locations are listed in Table 1. Most QTLs vary considerably in different environments, but QTLs making a significant proportion of all the QTLs were detected only in a single trial, indicating GE interaction. Among the yield components, 1,000-grain weight and grain weight/plant were least affected by environment, whereas spikes/plant, spikelets/spike and yield were most affected. Usually, QTLs with larger additive effects are less affected by environment.

Table 1. Major QTLs responsible for yield and yield components

Crop	Chromo-some	Marker interval	% of P.V.[1]	Reference
I. Yield				
Barley	3H	denso	15.1 - 64.5	Yin *et al*. (1999)
Maize	6	npi280	35.0	Veldboom and Lee (1994)
	10	UMC1506 - HbrMT320	12.1	Ho *et al*. (2002)
Rice	1	RM1	23.8 -39.1	Brondani *et al*. (2002)
	2	RG374 - RG394	10.7	Zhuang *et al*. (1997)
	2	CDO718	12.3	Thomson *et al*. (2003)
	3	RM16	18.7 -19.7	Brondani *et al*. (2002)
	4	OS15	23.6 -38.3	Brondani *et al*. (2002)
	4	RG91-RG449	11.4 -11.9	Hittalmani *et al*. (2003)
	5	RG573	11.0	Zhuang *et al*. (1997)
Sorghum	A	UMC58 - UMC23	14.9	Rami *et al*. (1998)
	G	UMC121a	30.8	Rami *et al*. (1998)
	G	BNL9.11	25.6	Rami *et al*. (1998)
	K	UMC115	18.0	Rami *et al*. (1998)
Wheat	2D	gwm702	11.5	Huang XQ *et al*. (2003)
	3B	gwm685	21.6	Huang XQ *et al*. (2003)
	3B	gwm493	9.6	Huang XQ *et al*. (2003)
	4A	BCD1738	17.0 -27.0	Araki *et al*. (1999)
	4D	gdm1163	12.3	Huang XQ *et al*. (2003)
	4D	gdm129	10.1	Huang XQ *et al*. (2003)
	5A	*q*	23.0 - 27.0	Kato *et al*. (2000)
II. Spikes/ plant				
Maize	3	UMC165A	19.0	Veldboom and Lee (1994)
	6	BNL5.47	24.0	Veldboom and Lee (1994)
Millet	2	PSM856 - PSM176	15.2	Poncet *et al*. (2002)
	6	Est - E713	37.0	Poncet *et al*. (2002)
	7	PSM655	12.9	Poncet *et al*. (2002)
Rice	1	RG374 - RG394	10.7	Zhuang *et al*. (1997)
	1	RZ730 - RG801	15.1 - 27.7	Hittalmani *et al*. (2003)
	4	R514	10.6	Ishimaru *et al*. (2001)
	4	RG143 - RG214	26.1	Zhuang *et al*. (1997)
	4	G163 - RG214	18.0 - 25.7	Hittalmani *et al*. (2003)
	8	RZ562 - RZ66	10.9	Xiao *et al*. (1996b)
	8	RM223	11.0 -30.8	Brondani *et al*. (2002)
	11	RM4B	9.9 -25.92	Brondani *et al*. (2002)

Table 1. Continued

Rice	1	RM1 - RM220	40.4 - 40. 8	Brondani *et al*. (2002)
	1	RM1 - RG532	12.4	Zhuang *et al*. (2002)
	1	RG374 - RG394	12.1	Zhuang *et al*. (1997)
	2	RG25 - RG437	13.3	Zhuang *et al*. (1997)
	2	CDO718	10	Thomson *et al*. (2003)
	3	RZ993 - CDO1081	12.3	Xiao *et al*. (1996b)
	4	OS15	30.0 - 35.9	Brondani *et al*. (2002)
	4	RG214 - G177	19.4 - 24.5	Lu *et al*. (1996)
	4	RG788 - RZ565	15.0 - 19.9	Hittalmani *et al*. (2003)
	4	RG163 - RG214	10.6 - 41.1	Hittalmani *et al*. (2003)
	4	RZ590 - CDO539	18.6	Xiao *et al*. (1996b)
	4	RG143 - RG214	11.5	Zhuang *et al*. (1997)
	6	G122 - G1314B	13.4 - 25.1	Lu *et al*. (1996), He *et al*. (2001)
	7	RM70 - Est_9	8.4 - 10.7	Hittalmani *et al*. (2003)
	8	RG562 - RG978	15.7	Zhuang *et al*. (1997)
	12	RM235 - RG181	15.0 - 19.1	Hittalmani *et al*. (2003)
Wheat	2D	BCD661/Ppd-D1	16.0 - 22.0	Li *et al*. (2002)
	4A	BCD1738	46.0 -52.0	Araki *et al*. (1999)
	5A	CDO457 - PSR164	26.0 - 39.0	Kato *et al*. (2000)
	7A	FBA97	16.0 - 22.0	Li *et al*. (2002)
III. Grains/ spike				
Barley	4H	MctgEaccC	11.7 -12.8	Teulat *et al*. (2001)
Maize	2	UMC78	41.0*	Veldboom and Lee (1994)
	4	UMC15	14.0*	Veldboom and Lee (1994)
	3	UMC165A	11.0**	Veldboom and Lee (1994)
	5	UMC27	14.0**	Veldboom and Lee (1994)
	6	npi280	35.0**	Veldboom and Lee (1994)
	8	BNL19.08	10.0**	Veldboom and Lee (1994)
Millet	2	PSM176 -PSM 592	60.9***	Poncet *et al*. (2002)
	5	PSM651 - PSM735	22.5***	Poncet *et al*. (2002)
Rice	1	RZ730	13.4	Xiao *et al*. (1998)
	3	RG910a - RG418	11.8	Li *et al*. (1997)
	3	C1087 - RZ403	15.4 -17.6	Xing *et al*. (2002)
	3	C136	20.6	Ishimaru *et al*. (2001)
	3	RZ993 - CDO1081	12.1	Xiao *et al*. (1996b)
	4	RG214 - C513	12.9 -28.0	Lu *et al*. (1996), He *et al*. (2001)
	4	CDO244 - RG864	22.3	Xiao *et al*. (1996b)

Table 1. Continued

Rice	5	RG360 - RZ296	18.1	Xiao *et al*. (1996b)
	5	RG9 - RG182	12.6	Zhuang *et al*. (1997)
	6	G294 - G329	13.0 - 25.9	Lu *et al*. (1996)
	8	RZ66 - RG598	14.1	Zhuang *et al*. (1997)
	8	C10122	11.4	Ishimaru *et al*. (2001)
Sorghum	B	UMC135 - SSSCIR92	12.3	Rami *et al*. (1998)
	C	BNL7.25 - BNL6.25	13.8	Rami *et al*. (1998)
Sorghum	F	BNL14.07 - BNL16.06	16.1	Rami *et al*. (1998)
Wheat	3A	Eps	18.3	Shah *et al*. (1999)
	3A	BCD1555	12.3	Shah *et al*. (1999)
	4A	BCD1738	12.0 - 27.0***	Araki *et al*. (1999)
	5A	q	17.0 - 42.0***	Kato *et al*. (2000)
	5A	CDO188 - BCD9	16.0 -18.0***	Kato *et al*. (2000)
	5A	PSR326 - CDO412	10.0 -16.0***	Kato *et al*. (2000)
Wild Rice	9	RZ2b	15.3	Kennard *et al*. (2002)
IV. Grain weight				
Barley	2H	Hvbkasi - vrs1	12.0 - 46.0	Marquez-Cedillo *et al*. (2001)
	6H	BCD1	10.8 - 14.1	Teulat *et al*. (2001)
Maize	1	BNL5.59 - UMC 23	28.6	Doebley *et al*. (1994)
	1	UMC157 - UMC37B	11.6	Doebley *et al*. (1994)
	2	UMC34 - UMC131	23.6	Doebley *et al*. (1994)
	3	UMC18A - UMC16A	31.1 - 34.4	Doebley *et al*. (1994)
	3	UMC175	22.0	Veldboom and Lee (1994)
	4	BNL5.46 - UMC42A - UMC66	12.8 - 15.7	Doebley *et al*. (1994)
	4	npi410	12.0	Veldboom and Lee (1994)
	5	BNL5.71 - UMC126	11.1 - 15.0	Schon *et al*. (1994)
	5	UMC166	11.0	Veldboom and Lee (1994)
	6	npi280	19.0	Veldboom and Lee (1994)
	7	BNL14.07 - UMC151	12.1 - 14.4	Schon *et al*. (1994)
	8	BNL9.11 - UMC328	18.1 - 19.9	Schon *et al*. (1994)
Rice	1	C949 - G370	13.2 - 19.0	Lu *et al*. (1996)
	1	RZ801 - RG331	12.2 - 20.9	Hittalmani *et al*. (2003)
	1	RG690 - RM212	12.1 - 13.5	Hittalmani *et al*. (2003)
	2	G1314A - G243A	12.1 - 18.6	Lu *et al*. (1996)
	3	RG455a - CDO109a	11.7	Li *et al*. (1997)
	3	RZ574 - RZ284	9.8 - 16.8	Hittalmani *et al*. (2003)
	3	RM16	30.3 - 32.8	Brondani *et al*. (2002)

Table 1. Continued

Rice	3	CDO1081	15.3	Xiao *et al*. (1996b)
	3	C1087 -RZ403	13.1 - 20.8	Xing *et al*. (2002)
	3	RZ672	10.2	Xiao *et al*. (1998)
	3	RM49 - CDO337	14.2	Hittalmani *et al*. (2003)
	4	RG143 - RG214	8.5 - 14.6	Zhuang *et al*. (1997)
	4	RM252	24.1 - 25.2	Brondani *et al*. (2002)
	4	OS15	19.7 - 23.3	Brondani *et al*. (2002)
	4	CDO244 - RG864	10	Xiao *et al*. (1996b)
	5	RG9 - RG182	11.5 - 14.8	Zhuang *et al*. (1997)
	5	RG573	10.9 - 111.4	Zhuang *et al*. (1997)
	5	RZ296	10.1	Xiao *et al*. (1996b)
	6	RG433 - Cat_1	7.6 - 10.4	Hittalmani *et al*. (2003)
	8	RG855 - G192	12.3 - 16.1	Lu *et al*. (1996)
	10	G134 - RZ500	12.5 - 25.1	Hittalmani *et al*. (2003)
	10	RG561 - RM228	13.4	Zhuang *et al*. (2002)
	10	RG241 -RG561	10.1	Zhuang *et al*. (1997)
	11	RM4B	13.3 - 15.5	Brondani *et al*. (2002)
Sorghum	B	UMC135 - SSSCIR92	35.2	Rami *et al*. (1998)
	B	BNL6.25-UMC84	10.7	Rami *et al*. (1998)
	G	UMC162 - UMC10	11.6	Rami *et al*. (1998)
Wheat	1A	CDO92	11.8	Campbell *et al*. (1999)
	1A	wmc333	15.1	Varshney *et al.* (2000)
	1B	ksuG9	11.1	Campbell *et al*. (1999)
	2A	gwm636	17.2	Huang XQ *et al*. (2003)
	2D	gdm6	15.4	Huang XQ *et al*. (2003)
	3A	CDO638	12.2	Shah *et al*. (1999)
	3A or 3B	BCD361	10.9	Campbell *et al*. (1999)
	3B	CDO718	12.2	Campbell *et al*. (1999)
	4D	gdm61	14.3	Huang XQ *et al*. (2003)
	5A	CDO475 - PSR164	11.0 - 19.0	Kato *et al*. (2000)
	5B	gwm544	16	Huang XQ *et al*. (2003)
	7A	gwm573b	14.5	Huang XQ *et al*. (2003)
	7B	gwm983	25.9	Huang XQ *et al*. (2003)
	7B	gwm46	20.6	Huang XQ *et al*. (2003)
	7D	gwm1002	17.3	Huang XQ *et al*. (2003)
V. Grain length				
Maize	1	UMC157	10.0	Veldboom and Lee (1994)
	3	UMC165A	19.0	Veldboom and Lee (1994)

Table 1. Continued

Maize	6	npi280	30.0	Veldboom and Lee (1994)
Wheat	3B	CDO718	21.9	Campbell *et al*. (1999)
Wild Rice	9	RZ206	14.4	Kennard *et al*. (2002)

% of P.V.[1] = percent of phenotypic variation; * indicates kernel row; **indicates ear length; *** indicate ear grain weight

In addition to additive effects, epistasis also was well documented in rice (Li *et al.*, 1997; Yu *et al.*, 1997; Xing *et al.*, 2002; Zhuang *et al.*, 2002). Three types of AA interactions were recognized: between QTLs with their own additive effects, between QTLs only showing epistatic effects and between additive QTLs and epistatic QTLs (Zhuang *et al.*, 2002), although the molecular mechanisms underlying the epistasis remain unknown.

A significant phenomenon encountered is the clustering of QTLs for different traits in the same genomic regions. In most cases, the same allele phases were involved in the clustered QTLs. For example, in maize, marker npi280 on the long arm of chromosome 6 explained 19% PV in 100-grain weight variation, 35% PV in ear length and 35% PV in yield; marker UMC165A on chromosome 6 explained 19% PV in ears/plant, 11% PV in ear length and 19% in PV in kernel length (Veldboom and Lee, 1994). Most likely, these colocalized QTLs resulted from pleiotropism. In a few exceptions, where the allele phases were different for the clustered QTLs, closely linked loci might exist. Furthermore, genes for several morphological traits were pleiotropic on yield components. The *Q* locus on the long arm of the wheat chromosome 5A ontributed significantly to the number of grains/spike, because recessives and heterozygotes cause partial sterility in the upper part of the spike (Kato *et al.,* 2000). The *vrs1* locus on the long arm of barley chromosome 2H, which determines fluorescence row type, coincided with the largest QTLs for yield, grain plumpness, grain weight and plant height (Marquez-Cedillo *et al.,* 2001). Although molecular genetics demonstrated that the genes responsible for starch synthesis, *e.g.* ADP-glucose pyrophosphorylase (Giroux *et al.,* 1996; Smidansky *et al.,* 2002) and cell wall-bound invertase-2 (Cheng *et al.,* 1996) played an important role in increasing grain yield, the QTLs for yield and yield components did not overlap with the mapped genes involving carbon metabolism (Ishimaru *et al.,* 2001). Although some putative orthologous QTLs for grain weight were identified in rice, maize and barley (Thomson *et al.,* 2003; Xiao *et al.* 1996b), a relatively small effort has been

made in the comparative mapping of QTLs for yield and yield component traits in cereals in general.

3.2. Heterosis

Genotyping was widely used to predict heterosis based on an essential hypothesis that hybrid performance is strongly correlated with general marker heterozygosity (Boppenmaier *et al.,* 1992; Dudley *et al.,* 1991; Jordan *et al.,* 2003; Melchinger *et al.,* 1990; Xiao *et al.,* 1996a; Zhang *et al.,* 1996). However, the correlation was very weak in maize and sorghum germplasm (Boppenmaier *et al.*, 1992; Dudley *et al.*, 1991; Jordan *et al.*, 2003; Melchinger *et al.*, 1990) and in rice mapping populations (Hua *et al.*, 2002; Xiao *et al.*, 1996a; Yu *et al.*, 1997). Higher correlation was observed in rice germplasm but not consistently, depending on the traits scored and materials used (Xiao *et al.,* 1996a; Zhang *et al.*, 1996), mainly because QTLs conditioning yield and yield components reside only in particular genomic regions and are not randomly distributed in the whole genome (Yu *et al.*, 1997; Jordan *et al.*, 2003).

3.2.1. QTLs for Heterosis in Rice

Studies on heterosis in rice provoked more controversy. Xiao *et al.* (1995), who genotyped a population of 194 F_8 recombinant inbred lines (RILs) derived from a subspecific cross and phenotyped the RILs and two F_7BC_1 populations derived from backcross between the RILs with respective parents, observed that 73% of the QTLs were detected in only one F_7BC_1 population. They suggested that dominance might be the major genetic basis of heterosis. Using similar types of populations (F_{10} RILs and $F_{10}BC_1$) plus two testcross populations, Li *et al.* (2001) and Luo *et al.* (2001) identified a large number of epistatic QTL pairs and a few main-effect QTLs, which explained >65% PV for yield and yield components. They found that about 90% of QTLs appeared to be overdominant and that most epistasis arose from complementary loci and proposed that overdominant epistatic loci are the primary genetic basis of inbreeding depression and heterosis in rice (Li *et al.,* 2001; Luo *et al.,* 2001). Genetically dissecting yield and yield components using an F_3 population derived from a elite rice hybrid, Yu *et al.* (1997) observed overdominance for most QTLs for yield, a few for yield

components and extensive digenic interactions including AA, additive × dominance and dominance × dominance. For the same hybrid, Hua *et al.* (2002) developed an F_1"immobilized F_2" population containing 360 F_1 from crosses between 240 F_9 RILs, and confirmed various digenic interactions, the AA being the prevalent type. Analyzing mid-parent heterosis QTLs based on the immobilized F_2 population, Hua *et al.* (2003) proposed a synthesis: heterotic effects (partial-, full- and overdominance) at a single-locus level and AA interactions as the genetic basis of heterosis.

3.2.2. QTLs for Heterosis in Maize- Overdominance

Using F_4 and F_3BC_1 maize populations, Stuber *et al.* (1992) detected more QTLs for yield than for yield components and found that heterozygotes had higher values for the yield phenotype than their homozygotes for most QTLs. They also concluded that overdominance (or pseudo-overdominance) for single QTLs played an important role in heterosis. Using a random-mated maize population, Lu *et al.* (2003) found that 86% of the QTLs for grain yield showed overdominance. In both studies, little epistasis was detected among the QTLs. Comparative sequence analysis of the *bz* genomic region of the two maize lines, McC and B73, provided a molecular basis for the dominance model of heterosis and inbreeding depression (Fu and Dooner, 2002). Compared with McC, four genes were missing from the *bz* region of B73. Four haplotypes were identified when *bz* region of different maize lines were genotyped. Lines lacking different genes may complement one another in an F_1 hybrid and show hybrid vigor. Fixing heterosis has not been possible through assembly of all the genes by crossing over at that location of the chromosome. The loss of functional genes may be the basis of inbreeding depression. However, a comparative study of *z1C* genomic region of two maize lines supported overdominance model of heterosis (Song and Messing, 2003). Comparison of the *z1C* region of B73 and BSSS53 revealed significant loss of gene collinearity. No dosage effect for the non-allelic genes was observed in the hybrid. The expression of the remaining *z1C* members exhibited overdominance in one of the reciprocal hybrids. Thus, Song and Messing (2003) believed that trans-acting mechanisms that regulate the expression of *z1C* genes diverged in the maize lines, and that multiple interactions between alternative regulatory factors, unique to each line, may underlie the overdominance.

3.2.3. Heterosis and Epigenetic Changes

In addition to genetic variation, heterosis may also be associated with epigenetic changes. The *Pl-Rh* allele of the *purple plant locus* (*pl*) is subject to epigenetic changes in gene expression. Another allele *Pl'-mah* with low expression is dominant over the allele *Pl-Rh* with high expression. *Pl'-mah* exhibits a higher expression level when it is heterozygous with other *pl* alleles or hemizygous, indicating that the interaction between *Pl-Rh* and *Pl'-mah* is necessary to maintain the low-expression state of *Pl'-mah.* The *Pl-Rh* sequence for this interaction was absent in other alleles. The high-expression state can be transmitted to next generation only from the heterozygote of *Pl'-mah* with other *pl* alleles and not from the hemizygote, implying that synapsis may be required to fix the epigenetic changes (Hollick and Chandler, 1998).

4. GENOMICS TO MANIPULATE PLANT TYPE

Grain yield is realized by transporting and storing photosynthetic product in kernels at the population level. Plant type, including plant height, tiller number and position of tillers and leaves, contribute significantly to population structure, photosynthetic efficiency and the grain yield. Before the 1960s, traditional varieties of rice and wheat were tall and leafy, so that lodging before maturity resulted in low utilization of solar energy, N fertilizer and water, leading to low harvest index and low yield. In contrast, improved rice and wheat varieties released in mid 1960s had shorter stature, more tillers, erect leaves, higher harvest index, tolerance to high stand density and resistance to lodging, which laid the foundation of the "green revolution" (Khush, 2001). Tropical maize was also benefited from reduced height.

4.1. Plant Height

Plant height is a quantitative trait with high heritability, least influenced by environmental factors and controlled by either a single gene or a few QTLs. Plant height is mainly reduced by the use of the semidwarfing genes: *sd-1* in rice and *Rht-1* in wheat. The reduction of plant height in maize, however,

was obtained through manipulating polygenes (see Khush, 2001). Single genes and major QTLs mapped for plant height are listed in Table 2.

Molecular cloning and characterization of the semi-dwarfing genes indicated that plant hormones, mainly gibberellin (GA) and auxin, play an important role in regulating plant height in cereals. *Rht-1* is well known not only for its role in semidwarf wheat varieties but also for its semidominant inheritance and GA insensitivity (GAI). A GAI mutant was also identified in the model plant Arabidopsis. This gene encodes a transcription factor that negatively regulates the GA response (Peng *et al.,* 1997) and belongs to GRAS superfamily that has a putative NLS, a Ser/Thr-rich repeat and the LXXLL motif. The GAI protein has a DELLA motif near the N terminus. In the wild type, GAI negatively regulates GA signal transduction, and GA can release the repression by phosphorylating and finally degrading GAI via the ubiquitin/26S proteasome pathway mediated by the SCF^{GID2} complex (Sasaki *et al.,* 2003). In the mutant, the deletion of a 17-amino-acid fragment including the DELLA motif caused resistance to the GA effect and led to the dwarf phenotype (Peng *et al.,* 1997). Database searches found a rice EST with a strong similarity to the 17 amino-acid peptide sequence. A wheat cDNA clone was isolated using the rice EST as a probe and was mapped to an orthologous region of rice, maize and wheat, where the wheat semidwarfing gene *Rht-1* and maize semidwarfing gene *d8* reside. Genomic clones from wild type wheat and maize showed an amino-acid identity of >62% to the Arabidopsis GAI protein. Genomic sequences from three *d8* mutants revealed deletions of different sizes in N-terminal region, which included the DELLA motif or a region near it. Truncated proteins were deduced from the DNA sequences of wheat *Rht-B1b* and *Rht-D1b* mutants due to the nucleotide substitutions that created stop codons near the DELLA motif. Most probably, ribosomal scanning following translational termination at the mutant stop codons in *Rht-B1b* and *Rht-D1b* permits translation reinitiation at one of the several Met codons that closely follow these stop codons, and the resultant N-terminally truncated product confers the mutant phenotype (Peng *et al.,* 1999). Transformation of Basmati rice with the Arabidopsis GAI mutant allele significantly reduced the plant height (Peng *et al.,* 1999). All these results suggested that the "Green Revolution" gene *Rht-1b* and maize semidwarf *d8* are gain-of-function mutant alleles of the orthologous locus for GAI (Peng *et al.,* 1999). Alterations in the NLS or C-terminal region of rice orthologue for GAI locus resulted into the loss-of-function, constitutive-GA response mutants *slender* (*slr1*), in which

Table 2. Major genes/ QTLs for plant type in cereals

Crops	Chromo-some	Marker interval	% P.V.[1]	Candi-date gene	Reference
I. Plant height					
Barley	2H	MWG858 - MWG889a	16.7		Kicherer *et al*. (2000)
	3H	Bmag013	13.8		Teulat *et al*. (2001)
	3H	MWG847 - MWG977	14.5		Kicherer *et al*. (2000)
Millet	7	PSM655	16.1		Poncet *et al*. (2002)
Maize	1	m5	13.3		Sari-Gorla *et al*. (1999)
	2	m27	11.7		Sari-Gorla *et al*. (1999)
	8	m127	14.4		Sari-Gorla *et al*. (1999)
Oats	18	UMN145B	~100	*Dw6*	Milach *et al*. (1997)
	19	CDO1437B	~100	*Dw7*	Milach *et al*. (1997)
	unknown	CDO1319	~100	*Dw8*	Milach *et al*. (1997)
Rice	1	RZ730 - RG810	25.2 - 63.3	*sd-1*	Hittalmani *et al*. (2003)
	1	RZ730	42	*sd-1*	Xiao *et al*. (1998)
	1	C949 -C86	31.5		He *et al*. (2001)
	1	RG532	20.2		Xiao *et al*. (1998)
	1	R2414	19.2		Ishimaru *et al*. (2001)
	1	C112	12.9		Ishimaru *et al*. (2001)
	2	RG256 -RG324B	15.1 -21.3		Zhuang *et al*. (1997)
	3	RZ574 - RZ284	11.4 - 18.0		Hittalmani *et al*. (2003)
	3	RG348 - RG944	21.1		Li *et al*. (1995)
	4	C513	16.2 - 16.8		Lu *et al*. (1996)
	4	Y1065	12.6		Xiao *et al*. (1998)
	4	C975 - RG508	11.9		He *et al*. (2001)
	7	RG769 - RG511	7.9 - 10.4		Hittalmani *et al*. (2003)
	7	R2401	12.3		Ishimaru *et al*. (2001)
	7	R565	12.1		Ishimaru *et al*. (2001)
	8	RG20 - RG1034	25.1		Li *et al*. (1995)
	8	RM210	23.8		Xiao *et al*. (1998)
	9	RM219	18.4		Xiao *et al*. (1998)
	12	CDO459	29.4		Xiao *et al*. (1998)
Rye	4R	PSR392	~100	*np*	Malyshev *et al*. (2001)
	5R	WG199	~100	*Dw1*	Korzun *et al*. (1996)
Sorghum	A	UMC23 -SSCIR88	29.0 - 60.2		Rami *et al*. (1998)
	A	ISU165 - CDO20	11.0 - 16.2		Hart *et al*. (2001)
	D	TXA12	17.5 - 17.8		Klein *et al*. (2001)
	D	SB428a	54.8	*Dw2*	Lin *et al*. (1995)

Table 2. Continued

Sorghum	E	TXS1579	24.3 - 40.1		Hart *et al*. (2001)
	E	TXA10062	11.8 - 17.5		Klein *et al*. (2001)
	F	UMC149 - UMC122	14		Rami *et al*. (1998)
	J	SB815	23.1		Lin *et al*. (1995)
Wheat	1B	CDO118	15.0 - 30.0		Cadalen *et al*. (1998)
	1B	PSR949 - gwm18	13.3		Keller *et al*. (1999)
	2A	PSR958 - PSR566c	29.3		Keller *et al*. (1999)
	2B	gdm87a	17.4		Huang *et al*. (2003)
	2D	gwm261	~100	*Rht8*	Korzun *et al*. (1998)
	3A	Eps	42.4		Shah *et al*. (1999)
	3A	bcd1555	10.4		Shah *et al*. (1999)
	4A	PSR119/Wx-B1	27.0 - 29.0		Araki *et al*. (1999)
	4A	BCD1738/Hd	20.0 - 27.0		Araki *et al*. (1999)
	4A	gwm397 - GLK315	22.5		Keller *et al*. (1999)
	4B	GLK566	13.0 -20.0	*Rht-B1*	Cadalen *et al*. (1998)
	4B	FBA1	10.0 -15.0		Cadalen *et al*. (1998)
	4B	gwm149	11.8	*Rht-B1*	Huang *et al*. (2003)
	4D	FBA211	9.0 -15.0	*Rht-D1*	Cadalen *et al*. (1998)
	4D	gdm61	29.5	*Rht-D1*	Huang *et al*. (2003)
	5A	104*	31.4		Keller *et al*. (1999)
	5A	PSR918b - PSR1201	23.1		Keller *et al*. (1999)
	5B	4*	13.9		Keller *et al*. (1999)
	6A	gwm570	16.5		Huang *et al*. (2003)
	6A	6*	11.7		Keller *et al*. (1999)
	7A	GLK478	10.3 -11.7		Cadalen *et al*. (1998)
	7B	ksuD2	7.7 -16.5		Cadalen *et al*. (1998)
	7B	PSR964 - PSR142	18.3		Keller *et al*. (1999)
Wild Rice	1	CDO328	13		Kennard *et al*. (2002)
	1	RZ730b	12.1		Kennard *et al*. (2002)
	2	RG139b	15.4		Kennard *et al*. (2002)
	10	CD)1380	14.7		Kennard *et al*. (2002)
II. Tiller number					
Millet	7	PSM655	16.4		Poncet *et al*. (2002)
Rice	4	OG60	12.1		Brondani *et al*. (2002)
	5	RM163	15.8		Brondani *et al*. (2002)
	7	RM2	16		Brondani *et al*. (2002)
	8	RM223	32.6		Brondani *et al*. (2002)
	11	RM4B	27.3		Brondani *et al*. (2002)

Table 2. Continued

Rice	11	STSG34	14.8	Brondani *et al*. (2002)
Sorghum	A	GAP36	31.8 - 34.0	Hart *et al*. (2001)
	I	TXS1868 - CDO244.2	39	Hart *et al*. (2001)
	I	TXS1030 - TXP97	23	Hart *et al*. (2001)
Wheat	1D	MWG837	14.0 - 15.0	Li *et al*. (2002)
	2D	BCD611	11.0 -15.0	Li *et al*. (2002)
	4A	PSR163	10.0 - 16.0	Kato *et al*. (2000)
	5A	Vrn1	7.0 - 37.0	Kato *et al*. (2000)
	5A	PSR370	10.0 - 19.0	Kato *et al*. (2000)
	6A	PSR10	12.0 - 31.0	Li *et al*. (2002)
Wild Rice	5	RZ912a	13.1	Kennard *et al*. (2002)
	unknown	RZ399	10.8	Kennard *et al*. (2002)
III. Tiller angle				
Wheat	2A	gwm497	12.0 - 14.0	Li *et al*. (2002)
	3A	BCD1812 - CDO638	14.0 -19.0	Li *et al*. (2002)
IV. Culm thickness				
Wheat	2A	PSR958 -PSR566	31.2	Keller *et al*. (1999)
	2B	88*	13.2	Keller *et al*. (1999)
	3A	PSR598 - GLK529	14.7	Keller *et al*. (1999)
	3B	54*	11.3	Keller *et al*. (1999)
	4A	194*	16	Keller *et al*. (1999)
	4B		12.9	Keller *et al*. (1999)
	5A	PSR918 - PSR1201a	37.6	Keller *et al*. (1999)
	5B	40*	11.1	Keller *et al*. (1999)
V. Leaf angle				
Sorghum	A	UMC128 - ISU76.2	19.2	Hart *et al*. (2001)
	E	TXP92 - TXS1579	28.4 - 45.3	Hart *et al*. (2001)
Wheat	1A	86*	12.1	Keller *et al*. (1999)
	3B	54*	11.1	Keller *et al*. (1999)
	4A	gwm397 - GLK315	16.4	Keller *et al*. (1999)
	5A	42*	11.2	Keller *et al*. (1999)
	7D	64*	16.4	Keller *et al*. (1999)
Wild Rice	3	RZ448	19.2	Kennard *et al*. (2002)
	5	RZ576	25.5	Kennard *et al*. (2002)
	11	CDO365	14.4	Kennard *et al*. (2002)

Table 2. Continued

VI. Leaf width				
Wheat	1B	PSR949 - gwm18	14	Keller *et al*. (1999)
	3B	54*	19.7	Keller *et al*. (1999)
	5A	192*	14.9	Keller *et al*. (1999)
	5B	120*	11.2	Keller *et al*. (1999)
VII. Root length				
Rice	9	RZ12 - RM201	12.9	Venuprasad *et al*. (2002)
VIII. Root dry weight				
Rice	3	RM231 - RZ329	30.7	Venuprasad *et al*. (2002)
IX. Root volume				
Rice	3	RM231 - RZ329	29.1	Venuprasad *et al*. (2002)
X. Root thickness				
Rice	2	RG157 - RZ318	26.9	Venuprasad *et al*. (2002)

% of P.V.[1] = percent of phenotypic variation ; * The map position (cM) of highest LOD score

elongation was unaffected by the inhibitor of GA biosynthesis, although endogenous GA was maintained at a lower level as compared with wild-type plants (Ikeda *et al.,* 2001). This indicated that the C- and N-terminal parts of GAI protein play opposite roles in GA response. Characterization of the allelic variation at the *slender 1* (*Sln1*) locus of barley showed that premature translation termination caused constitutive-GA response and a slender phenotype, and a Glu-to-Gly substitution in the vicinity of the DELLA motif led to GA insensitivity and dwarf plants (Chandler *et al.,* 2002).

Another "green revolution" gene, *sd-1* of rice, involves the GA biosynthetic pathway and encodes GA_{20} oxidase (Sasaki *et al.,* 2002a; Monna *et al.,* 2002a; Spielmeyer *et al.,* 2002). *sd-1* was first identified in the Chinese variety Dee-Geo-Woo-Gen and was widely used to breed semidwarf rice. *sd-1* from Dee-Geo-Woo-Gen and its derivatives contained a 383-bp deletion, spanning the first and second exons, which resulted in a frame-shift and created a stop codon after the deletion (Sasaki *et al.,* 2002a; Monna *et al.,* 2002a; Spielmeyer *et al.,* 2002).

In the radiation-induced *sd-1* mutant Calrose 76, a single-base substitution in exon 2 caused an amino acid substitution Leu (CTC) to Phe (TTC) (Monna *et al.,* 2002a). A high level of GA_{53} accumulated in s*d-1* mutants, suggesting that GA_{20} oxidase catalyzes the following three steps: $GA_{53} \rightarrow GA_{44} \rightarrow GA_{19} \rightarrow GA_{20}$ (Sasaki *et al.,* 2002; Spielmeyer *et al.,* 2002). The genetic mechanisms by which GA regulates plant height are summarized in Fig. 1.

The maize dwarfing mutant *br2* significantly shortens the lower stalk internodes and reduces cell length and diameter by ~50%, but it does not affect other plant parts. Molecular cloning of *br2* by transposon tagging revealed a new dwarfing mechanism: blocking the polar transport of auxin (Multani *et al.,* 2003). *Br2* encodes an ATP-binding cassette transporter of the multidrug resistant (MDR) class of P-glycoproteins that modulate the polar movement of auxin. Characterization of isogenic lines and revertants of the sorghum dwarf *dw3* mutant indicated that sorghum *Dw3* is homoeologous to maize *Br2*. Mutation of *dw3* was caused by a direct duplication of an 882-bp segment in exon 5, and all the revertants lacked this duplication, indicating that the duplication is responsible for the dwarfing phenotype, and that unequal crossing over between the duplicated fragments led to the phenotypic instability. Furthermore, a stable *dw3* without direct duplication was recovered from the progeny through unequal recombination (Multani *et al.,* 2003).

4.2. Tiller Number

Tillering is of great importance in the grain yield in barley, rice and wheat and also played a critical role in the domestication of maize. Tiller number is inherited quantitatively in most cases (see Table 2 for major QTLs) and is affected by soil fertility and environmental factors, especially daylength and temperature. In wheat, photoperiod response genes (*Ppd*) (Li W. *et al.,* 2002) and vernalization genes (*Vrn*) show a strong pleiotropic effect on tiller number. The identification of monoculm or uniculm mutants in barley, rice and wheat and branched mutant in maize indicated that tillering ability, like plant height, could be under the control of a single gene. Molecular cloning of the gene, *teosinte branched 1* (*tb1*) from maize and *Monoculm 1* (*MOC1*) from rice initiated the effort to elucidate the pathway conditioning tillering ability.

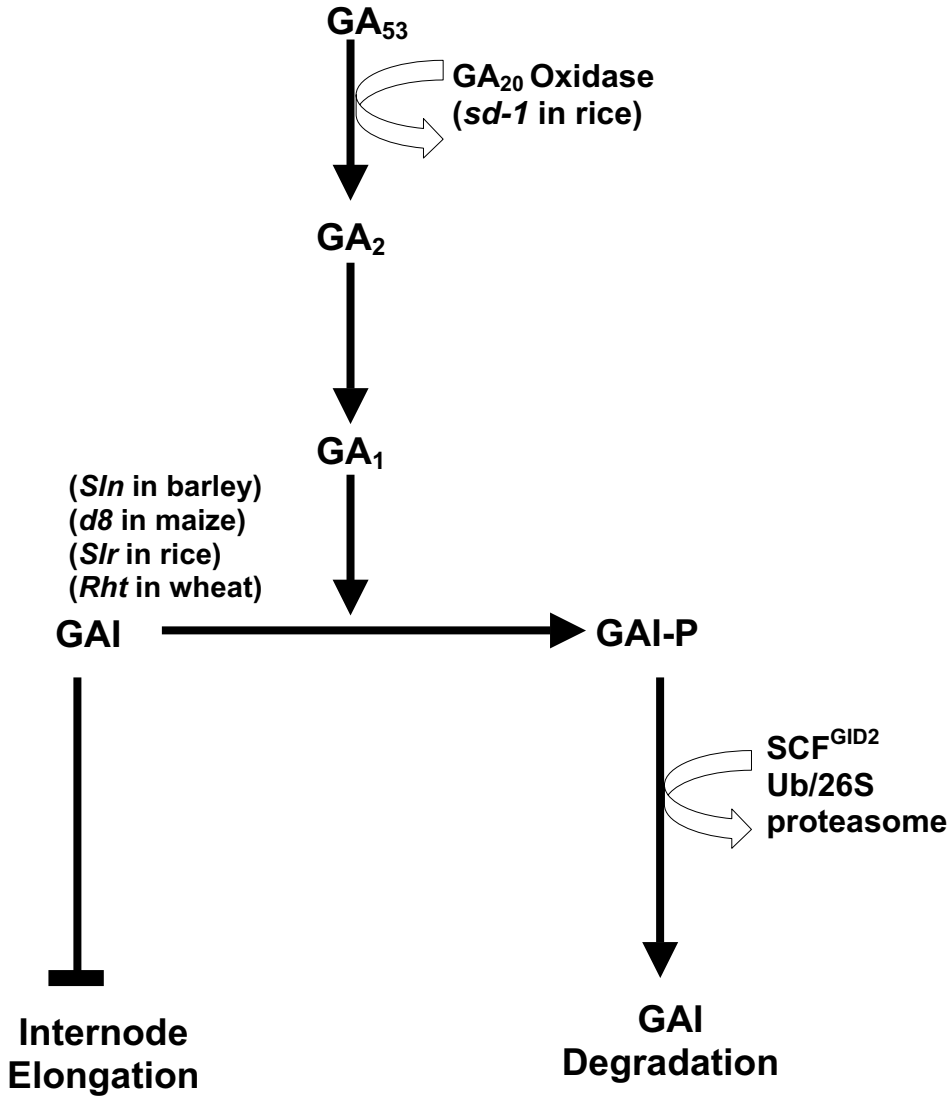


Figure 1. GA signaling and plant height. In the wild type, GAI represses growth and GA promotes GAI phosporylation and the phosporylated GAI (GAI-P) is degraded through the ubiquitin/26S proteasome pathway mediated by the SCF^{GID2} complex. The mutation in GA_{20} oxidase produced the semidwarfing *sd-1* rice. Small deletions or point mutation at or around the DELLA domain prevent a GAI response to the GA signal and caused GA-insensitive semidwarfs *Sln* of barley, *d8* of maize and *Rht* of wheat. Loss of the C-terminal portion of GAI caused the *slender* phenotype in rice and barley.

Morphologically, maize differs from its wild ancestor teosinte dramatically in many features of plant type, of which non-tillering and female fluorescence in maize are the most eye-catching. The maize *tb1* locus is largely responsible for these differences (Doebley *et al.,* 1995). Maize *tb1* mutant resembles teosinte because of a complete loss of apical dominance in secondary and tertiary axillary branching and elongation of internodes, indicating that *tb1* is a repressor to lateral growth, including development of tillers and spikelets (Hubbard *et al.,* 2002). Molecular cloning by transposon-tagging showed that *tb1* encodes a transcription factor containing a noncanonical basic-Helix-Loop-Helix (bHLH) structure (Doebley *et al.,* 1997), a domain termed as TCP because it was first found in TB1, CYC and PCF proteins (Cubas *et al.,* 1999). The TCP domain is required for DNA binding and dimerization (Cubas *et al.,* 1999). TB1 belongs to the second class of the TCP family and binds specifically to DNA sequence GTGGNCCC (Kosugi and Ohashi, 2002). Dramatic polymorphisms were found in the 5' regulatory region of maize *tb1* compared with teosinte, suggesting a prolonged domestication period (Wang *et al.,* 1999). Overexpression of rice orthologue, *OsTB1* in transgenic rice plants significantly reduced the lateral branching, increased culm thickness but did not affect axillary budding. The rice *fine culm* (*fc1*) mutant exhibits a threefold-higher tiller number than the wild type and carries a defective *OsTB1* that contains a single-base (C) deletion at position 327 and causes a frameshift leading to premature translation termination of the gene product (Takeda *et al.,* 2003). Khush (2001) proposed a new rice plant type model with fewer and stronger tillers and more grains in each panicle; *tb1* may be a good target for engineering for this purpose.

Map-based cloning of the rice *Monoculm1* (*MOC1*) locus led to the identification of a new axillary growth regulator, which is a transcription factor of the GRAS superfamily with homology to *LATERAL SUPPRESSIOR* of tomato. The *moc1* mutation was caused by insertion of a 1.9-kb retrotransposon in the ORF, which created a premature stop codon and caused nearly complete loss of lateral branching and axillary budding, including tillers and spikelets. Spatial and temporal expression patterns confirmed that *MOC1* is required for meristem initiation and axillary bud formation and suggested that *MOC1* downregulates *OsTB1* (Li X. *et al.,* 2003).

Other plant type characters such as thickness and angle of tillers, width and angle of leaves, length, thickness, dry weight and volume of roots were also studied and major QTLs identified (Table 2). Molecular cloning of the QTLs will help elucidate the underlying mechanisms to efficiently manipulate plant types.

5. GENOMICS TO UNDERSTAND FLOWERING-TIME

In nature, plants evolved mechanisms to sense the environment, so that they germinate and flower at a favorable time. Flowering time largely determines the adaptability of a plant to different ecogeographical regions. For crops, flowering time is also of agricultural importance; earliness not only helps crops avoid adverse weather, but also leaves adequate time for subsequent crops in multi-cropping rotation systems. Therefore, flowering time has been an active research area of plant biology. In Arabidopsis, a large number of genes have been isolated that involve four signal-transduction pathways regulating the timing of flowering including photoperiod, vernalization, autonomous and GA pathways. All these pathways converge to regulate the same set of genes, i.e., *FT* and *SOC1*, and further to *AP1* and *LFY*, finally leading to flowering (Mouradov *et al.,* 2002).

Flowering time has been studied in detail in cereals by genetic mapping, leading to its dissection into discrete QTLs with high heritability (Table 3). In tropical cereals (maize, rice and sorghum) and spring-type temperate cereals (wheat, barley, rye and oats), photoperiod response is the major determinant of flowering time. Tropical and temperate cereals differ in photoperiod response, the former belonging to short-day (SD) and the later to long-day (LD) plants, so that autumn-sown (winter type) temperate cereals need vernalization and an optimum photoperiod for flowering. The "earliness *per se*" factors also contribute to flowering time as fine-tune but independently of the photoperiod and vernalization responses.

5.1. Photoperiod Response

Historically, cereals have played an important role in research of the photoperiod response. Major loci controlling photoperiod sensitivity have

Table 3. Major QTLs for flowering time

Crop	Chromo-some	Marker interval	% of P.V.[1]	Candi-date gene	Reference
Barley	2H	B15c - ABC162	19.0 - 55.0		Marquez-Cedillo *et al*. (2001)
Millet	5	PSM651 - PSM735	59.0		Poncet *et al*. (2002)
	7	PSM655	24.2		Poncet *et al*. (2002)
Rice	1	C949 - G370	14.1		Lu *et al*. (1996)
	1	RZ801 - RG331	13.8 - 20.5		Hittalmani *et al*. (2003)
	1	RG532	7.5 - 14.9		Thomson *et al*. (2003)
	3	RG348 - RG944	44.7		Li *et al*. (1995)
	3	RZ574 - RZ284	11.3 -20.9		Hittalmani *et al*. (2003)
	4	RG908 - RG91	10.2 - 13.1		Hittalmani *et al*. (2003)
	5	RM163	21.0 - 26.1		Brondani *et al*. (2002)
	6	RG264	15.2		Xiao *et al*. (1998)
	7	RM214	10.5 - 25.8		Thomson *et al*. (2003)
	7	RG769 - RG511	11.4 - 12.1		Hittalmani *et al*. (2003)
	8	RG885 - RM44	35.2		He *et al*. (2001)
	8	RG20 - RG1034	42.5		Li *et al*. (1995)
	8	RG885 - RZ617	33.3 - 35.4		Lu *et al*. (1996)
	8	RG978 - RZ617	8.5 - 12.2		Hittalmani *et al*. (2003)
	10	RG257 - G1084	22.4		Lu *et al*. (1996)
	10	C16 - C223	9.3 - 18.3		Lu *et al*. (1996)
	12	G2140 - G1391	10.5		He *et al*. (2001)
Sorghum	D	SB188x	85.7	*Ma1*	Lin *et al*. (1995)
Wheat	1Am	WG241	47.0	*Eps-Am1*	Bullrich *et al*. (2002)
	2A	gwm95	11.5		Huang *et al*. (2003)
	2A	PSR131c	14.1 - 16.6		Ahmed *et al*. (2000)
	2A	gwm497	10.0 -11.0	*Ppd-A1*	Li *et al*. (2002)
	2B	PSR135	13.5 - 13.7	*Esp-2BS*	Ahmed *et al*. (2000)
	2D	gwm484	15.0		Huang *et al*. (2003)
	2D	BCD660	29.0 -31.0	*Ppd-D1*	Li *et al*. (2002)
	6A	gwm263	13.7		Huang *et al*. (2003)
	6A	gwm427	16.9		Huang *et al*. (2003)
	7A	PSR103	14.5 - 20.9	*Esp-7A*	Ahmed *et al*. (2000)

% of P.V.[1] = percent of phenotypic variation

been identified using traditional genetic analysis of spontaneous or induced mutants. Molecular mapping and cloning of these genes has shed light on the genetic basis of photoperiod response. For instance, the photoperiod-insensitive mutant *se5* of rice is completely deficient in photoperiod response and phytochromes and has a very early heading date. A rice EST showed high similarity to Arabidopsis *HY1* gene and was designated as *OsHY1* and

mapped to *se5*. Subsequent sequence analysis revealed that *OsHY1* from *se5* contained a 1-bp deletion that generated a premature stop codon and truncated the predicted protein. Transforming the *se5* mutant with wild type *OsHY1* from its progenitor Norin8 restored photoperiod sensitivity (Izawa *et al.,* 2000). The HY1 and HY2 proteins catalyze the synthesis of the chromophore of all the phytochromes (see Mouradov *et al.,* 2002). In sorghum, the photoperiod sensitivity gene *Ma3* encodes a phytochrome B. A 1-bp deletion was found 30 bp upstream of the translation termination in the *ma3* mutant compared to the wild-type plant (Childs *et al.,* 1997).

Fifteen heading-date QTLs (*Hd1*, *Hd2*, *Hd3a*, *Hd3b*, *Hd4* to *Hd14*) were identified with high-density maps based on F_2 and AB populations derived from a cross between *japonica* rice Nipponbare and *indica* rice Kasalath (Yano *et al.,* 2001). *Hd1* (Yano *et al.,* 2000), *Hd3a* (Kojima *et al.,* 2002) and *Hd6* (Takahashi *et al.,* 2001) were isolated by map-based cloning. A major QTL *Hd1* encodes a transcription factor showing a high level of similarity to *CONSTANS* (*CO*), a key player in the photoperiod pathway in Arabidopsis. Like *CO*, *Hd1* protein contains a zinc finger near the N terminus for protein-protein interaction and a CCT domain near the C terminus for nuclear localization. Compared with the functional allele from Nipponbare, extensive sequence variation was revealed in the Kasalath allele, including four SNPs, one two-base substitution, a 36-bp insertion and a 33-bp deletion in the first exon and two SNPs and a two-base deletion in the second exon. The *Hd1* locus was also isolated from *Se1* mutants HS66 and HS110 and their progenitor Ginbouzu. A 43-bp deletion and a 433-bp insertion occurred in the first exon of HS66 and HS110, respectively compared to that of Ginbouzu. Furthermore, Ginbouzu *Hd1* differs from Nipponbare *Hd1* for one SNP and a 36-bp insertion in the first exon. These results showed that *Hd1* was allelic to *Se1* (Yano *et al.,* 2000).

The *Hd6* QTL has a relatively small effect. The Kasalath allele is semidominant and encodes an alpha subunit of the protein kinase CK2 (Mouradov *et al.*, 2002). As expected, Kasalath carries a functional allele, and the predicted protein contains all the amino-acid positions conserved among the eukaryotes. A single SNP in the Nipponbare allele created a premature stop codon that truncated the predicted protein, causing loss of kinase activity (Takahashi *et al.*, 2001).

Fine mapping using AB population and NILs further dissected the original QTL *Hd3* into two distinct genes, *Hd3a* and *Hd3b* (Monna *et al.* 2002b). Hd3a showed a high sequence similarity to the *FLOWERING LOCUS T* (*FT*) gene, immediately downstream of *CO* in the photoperiod pathway and promoting flowering in Arabidopsis. The Kasalath *Hd3a* differs from Nipponbare *Hd3a* by three SNPs, one of them giving a synonymous codon leading to no change in amino acids, the other two causing a change in an amino acids (Asn in Kasalath and Pro in Nipponbare). Genetic transformation of Nipponbare demonstrated that both alleles are functional, although the Kasalath transgene tended to head earlier than the Nipponbare transgene. The copy number of the transgene positively correlated with the earliness (Kojima *et al.,* 2002). Molecular characterization of *Hd1* and *Hd3a* confirmed the interaction of these two QTLs under SD (short-day) conditions (Lin *et al.,* 2000; Monna *et al.,* 2002).

A rice homologue of the Arabidopsis *GIGANTEA* (*GI*) gene was isolated during characterization of the *se5* mutant and was designated as *OsGI* (Hamaya *et al.,* 2003). Overexpression of *OsGI* in transgenic rice caused late flowering under both LD and SD conditions. In this transgenic rice, expression of *Hd1* increased but that of *Hd3a* was suppressed (Hayama *et al.,* 2003). In Arabidopsis, *GI* upregulates *CO* and *CO* upregulates *FT* (Mouradov *et al.,* 2002). These results indicated that key regulatory genes for photoperiod pathway are conserved between the LD plant Arabidopsis and the SD plant rice but the regulation of *Hd3a* (*FT*) by *Hd1* (*CO*) was reversed in rice (Hayama *et al.,* 2003).

For the LD cereals barley, rye and wheat, a major photoperiod-response determinant *Ppd-1* was mapped to the short arm of homoeologous group 2. Comparative mapping of the barley *Ppd-H1* region was complicated by collinearity breakdown and sequence duplication in rice (Dunford *et al.* 2002), which caused difficulty for map-based cloning of *Ppd-1* using rice as a surrogate. CO-like sequences were isolated and mapped to barley chromosomes 1H, 2H, 5H, 6H and 7H, but none corresponded to *Ppd-H1* (Griffiths *et al.,* 2003). Wheat homologues of the Arabidopsis *CO* gene functioned similar to *Hd1* in transgenic rice and were mapped to the long arm of homoeologous group 6 (Nemoto *et al.,* 2003). These results imply that domestication and subsequent selections for variation in photoperiod response must have involved different genes in LD and SD cereals.

5.2. Vernalization Response

Autumn-sown cereals have to experience a long period of low temperature before flowering. This mechanism is called vernalization and evolved largely to prevent the floral meristem from cold damage in the winter. Major vernalization-response loci have been localized to the long arms of homoeologous group-4 (*Sh* of barley and *Vrn2* of einkorn wheat) and group-5 chromosomes (*Sh2* of barley, *Sp1* of rye and *Vrn1* of wheat). Map-based cloning and gene expression studies in einkorn wheat indicated that the wheat *Vrn1* gene is a homologue of the Arabidopsis transcription factor *APETALA 1* (*AP1*) of the MADS-box superfamily. Compared with the winter-type allele (*vrn1*), spring-type alleles (*Vrn1)* contain deletions of various sizes in the same region of promotor, adjacent to CG-box, although no difference was found in the coding region (Yan *et al.*, 2003). Based on the statistics (Tranquilli and Dubcovsky, 2000) and sequence analyses, a genetic model was proposed to explain the evolution of vernalization requirement in the Triticeae (Yan *et al.*, 2003). According to this model, *Vrn1* and *Vrn2* proteins act in same pathway with reverse effects on flowering. *Vrn1* promotes flowering but *Vrn2* represses expression of *Vrn1* and further represses flowering. Vernalization gradually reduces the abundance of *Vrn2* protein and restores the expression of *Vrn1* and leads to flowering. Therefore, only the genotype *vrn1vrn1Vrn2-* needs vernalization (Yan *et al.*, 2003). At nearly the same time, Trevaskis *et al.* (2003) reached the similar conclusion in common wheat and barley.

5.3. GA Pathway

In Arabidopsis, GA promotes flowering either through SOC1→LFY or through FPF1 or other components (Mouradov *et al.*, 2002). As described above, the maize dwarfing gene *D8* encodes a GAI protein that plays an important role in GA signaling (Peng *et al.*, 1999). Sequencing of 92 alleles at the *D8* locus identified 41 distinct haplotypes (Thornsberry *et al.*, 2001). Linkage disequilibrium (LD) analysis detected a significant association between the sequence polymorphisms and variation in flowering time, especially days to silking, but not with plant height because the DELLA domain was not altered in all the maize inbreds used. Nine polymorphisms were significantly associated with flowering time in all the field plots. These

polymorphisms include deletion, insertion and substitutions from the promotor to the C terminus and may have impact on the expression and function of *D8* protein. Deletion of two codons flanking the SH2-like domain had an estimated effect of reducing flowering time by 7 to 11 days (Thornsberry *et al.*, 2003). Because GAI is very similar to RGA at the DNA and protein-sequence levels and RGA negatively regulates flowering in Arabidopsis (see Mouradov *et al.*, 2002), the polymorphism found in the maize *D8* gene may suggest that this gene has a pleiotropic effect on flowering time through the GA pathway.

5.4. Earliness *per se*

Variation of flowering time was observed independently of vernalization and photoperiod response in wheat and barley; the genes were named "*earliness per se*" (*Esp*). Eight *Esp* loci were also mapped to all barley chromosomes except 1H (Laurie *et al*., 1995), an *Eps* locus homoeologous to *eps5L* was mapped on the long arm of wheat chromosome 5A (Kato *et al.*, 1999) and *Eps-A^{m}1* was mapped to the distal region of einkorn wheat chromosome 1A^m (Bullrich *et al.*, 2002). Compared to other pathways, earliness *per se* has a smaller effect on flowering. Similar to the autonomous pathway in Arabidopsis, the *Eps-A^{m}1* locus showed a strong interaction with temperature (Bullrich *et al.*, 2002). The autonomous pathway promotes flowering in Arabidopsis by suppressing the expression of *FLC* (see Mouradov *et al.,* 2002).

6. GENOMICS FOR BETTER GRAIN QUALITY

Grain quality usually includes nutritional and processing quality. While nutritional quality mainly deals with the content of essential amino acids and digestibility, the major determinants of processing quality involve the composition and contents of grain storage proteins, starches and enzyme activities. Because cereals are cultivated to make various foods for different groups of consumers with different preferences, the scope of processing quality varies depending upon crop, food and consumer.

6.1. Nutritional Quality

Most cereal storage proteins, especially maize α-zein, are imbalanced in amino-acid composition and lack Lys, posing a major nutrition problem for cereals to be used as food and feed. In the last 50 years, many high-lysine mutants were identified in barley, maize, rice and sorghum, among which the maize mutants *opaque 2* (*o2*) and *floury 2* (*fl2*) are the best known. *o2* encodes a transcription factor of the basic domain/leucine zipper class (Schmidt *et al.,* 1987) and gives increased Lys content through suppressing the transcription of genes coding for a 22-kD α-zein (Unger *et al.,* 1993) and for the bifunctional enzyme lysine-ketoglutarate reductase/saccharopine dehydrogenase (LKR/SDH). The latter degrades lysine in maize endosperm (Kemper *et al.,* 1999), but enhances the expression of other lysine-containing proteins (Moro *et al.,* 1996) and the elongation factor-1 alpha (EF-1α) that is strongly correlated to the Lys content (Sun *et al.,* 1997). *o2* mutants also have a higher level of the free amino acids Lys, Thr, Met and Ile from the Asp pathway as well as Ala and Ser. Biochemically, Asp kinase (Ask), the first enzyme of the Asp pathway, is feedback regulated by its end product Lys. The mutants *Ask1* and *Ask2* of maize are less sensitive to Lys inhibition and cause overproduction of Lys, Thr, Met and Ile. *Ask1* and *Ask2* also seem to function under the control of *o2* gene. The double mutant of *Ask1* and *o2* is more insensitive than *Ask1* mutant to Lys inhibition (Brennecke *et al.,* 1996). Recently, a QTL influencing the free amino acid content was mapped to *Ask2* locus in the *o2* mutant background (Wang *et al.,* 2001). The expression in the endosperm of a large number of genes functioning in different pathways was affected by the *o2* mutation. Sixty genes were upregulated and 66 were downregulated by the wild-type allele of the *o2* gene (Hunter *et al.,* 2002).

The mutant, *fl2* corresponds to an Ala-to-Val substitution at the C-terminal position of the signal peptide of a 22-kD α-zein protein. This mutation led to a failure to correctly process the signal peptide (Coleman *et al.,* 1997), retention of the protein in the surface of endoplasmic reticulum (ER) and the creation of hydrophobic foci of unfolded α-zein protein bodies (Gillikin *et al.,* 1997). In response to the unfolded proteins in the endosperm cell, the syntheses of binding proteins and ER-resident chaperones were increased dramatically. Expression of 59 genes was regulated by the *fl2* gene; 22 were upregulated and 37 were downregulated in the *fl2* mutant (Hunter *et al.,* 2002).

Although the mutants *o2* and *fl2* greatly improved the nutritional quality, they were not used successfully early in maize breeding programs because the mutations caused substantial reduction in grain hardness and yield, and increased grain susceptibility to fungal diseases. Recently, genetic modifiers were identified for *o2* mutants that convert the soft and starchy endosperm of *o2* mutants into a hard and vitreous phenotype but maintain the higher Lys content by altering the expression of the 27-kDa γ-zein genes (Burnett and Larkins, 1999). By introducing the modifiers, CIMMYT has developed quality protein maize (QPM) varieties (Vasal, 2000).

In addition to the above, an effort was also made to increase the essential amino acids using other genetic and genomic approaches. Lai and Messing (2002) stabilized the expression of the high-Met protein Dzs10 by engineering the *cis*-acting site of its coding gene. Wu *et al.* (2003) raised the Lys level in rice significantly by changing the anticodon of $tRNA^{lys}$ species to match the stop codons. Segal *et al.* (2003) developed the novel "*opaque*" variants of maize by a single dominant RNA-interference-inducing transgene. In these *opaque* variants, the 22-kDa α-zein proteins were eliminated without affecting the accumulation of other zein proteins and grain hardness, and failed to give an *opaque* phenotype.

Digestibility is another factor affecting the nutritional quality. Protein digestibility is 81, 73, 66 and 46% for wheat, rice, maize and sorghum, respectively. A highly digestible sorghum mutant variety was identified, in which the protein bodies were irregular in shape and had many invaginations reaching into the central area. The relative locations of the α- and β-kafirins were similar in protein bodies of the mutant and normal varieties, but the γ-kafirins were concentrated in the dark-staining regions at the base in the mutant compared to the protein body's periphery in normal varieties (Oria *et al.,* 2000). The genetic identity causing the change in γ-kafirin deposition needs to be clarified.

6.2. Processing Quality

Many aspects of cereal processing quality have been investigated, and there are many terminologies regarding processing quality including milling, baking, boiling and steaming, malting, tasting, appearance and maintenance.

Here, we will discuss the progress made in the study of the genetic determinants of these processing quality traits at the genome level.

6.2.1. Grain Hardness

Grain hardness is one of the most important determinants of processing quality for the small grains, directly related with milling quality in wheat (i.e., flour yield, milling score and flour ash), and with malting quality in barley. Kernel texture is best documented genetically in wheat and barley. Wheat grain hardness is mainly under the control of a single genetic locus *Ha* on the short arm of chromosome 5D. Biochemical studies showed that the presence of the starch surface protein friabilin, which proved to be identical to the puroindolines a and b, was associated with grain hardness (Gautier *et al.*, 1994; Morris *et al.,* 1994). Puroindolines a and b were genetically mapped to the *Ha* locus (Sourdille *et al.,* 1996, Giroux and Morris, 1997). Molecular characterization of the *Ha* locus showed that when both puroindoline a and b are wild type (functional state), the grain texture is soft; when either one is absent or mutated, grain texture is hard (Giroux and Morris, 1997, 1998). The complete absence of puroindoline a or point mutations of Gly46-to-Ser46 (Giroux and Morris, 1997), Trp44-to-Arg44 and Leu60-to-Pro60 (Lillemo and Morris, 2000) substitutions in puroindoline b are associated with the hard allele of *Ha*. Expression of the wild-type puroindoline b gene in transgenic wheat complements a hard phenotype (Beecher *et al.*, 2002a). Surveying the genotypes of 71 hard spring wheat cultivars, Morris *et al.* (2001) found that 18 (25.4%) were null for puroindoline a, 47 (66.2%) carried mutation Gly46-to-Ser46, 4 (5.6%) had mutation Leu60-to-Pro60 and two (2.8%) were involved in a nonsense mutation changing Trp39 to a “stop” codon. Durum wheat has a uniform hard phenotype of grain hardness, due to the absence of both puroindoline a and b in this species.

Genes encoding puroindolines a and b are single copy and are tightly linked genetically and physically (Tranquilli *et al.*, 1999). Homologues of puroindolines a and b have been cloned and sequenced from genus *Aegilops* and other distant relatives of wheat, namely hordoindolines of barley, secaloindolines of rye and avenoindolines of oat (Gautier *et al.,* 2000). Hordoindolines were mapped to the distal part of the short arm of chromosome 5H, a location coinciding with a malt-exacting QTL (Beecher *et al.,* 2001) and associated with a major endosperm texture QTL (Beecher *et*

al., 2002b). In barley, two copies of hordoindoline b in tandem duplication were present in all cultivars. Hordoindoline a was present in all the soft and most hard cultivars (Darlington *et al.,* 2001).

6.2.2. Storage Proteins

Although most cereal storage proteins are poor in nutrition as discussed above, they are important for processing quality, particularly in wheat, and are adversely associated with malting quality in barley. There are two major classes of storage proteins in cereal endosperm: globulins and prolamins. Globulins are the major storage proteins in oats and rice, and prolamins are a major component in wheat, barley, rye and maize. As they are abundant in endosperm, cDNAs coding for most of the globulin and prolamin genes have been cloned, sequenced and mapped. Allelic variation also has remarkable effects on the processing quality. In wheat, allelic differences in the *Glu-A1*, *Glu-A3* and *Glu-B3* loci significantly affect the parameters for breadmaking quality: whole flour protein content (WFP), SDS sedimentation volume (SED) and mid-line peak volume (MPV); *Glu-A1* and *Glu-B3* loci also influence Pelshenke time (PEL). The combination of subunits 5+10 at the *Glu-D1* locus increased the PEL and SED. The *Glu-D3* locus affects all parameters except SED (Luo *et al.,* 2001).

In contrast to the coding genes, the genetic mechanisms underlying the grain protein content variations remain unclear in cereals except for maize in which the *o2* gene is responsible for accumulation of major storage protein 22-kDa α-zein. QTLs for grain protein content were identified in wheat, barley, rice and sorghum (Table 4). The QTL on the short arm of wheat chromosome 6B was transferred from wild wheat *Triticum turgidum* subsp. *dicoccoides* and is inherited monogenically explaining 70% PV (Chee *et al.,* 1998).

6.2.3. Starch Species

As a predominant component of the endosperm, starch plays an important role in determining processing quality. Cereal starch is composed of two types of polysaccharides: amylose and amylopectin, which account for 20-30% and 70-80% of the starch granule, respectively. Many genes coding for starch synthases have been isolated from cereals. Granule-bound starch

synthase (GBSS), encoded by the *Wx* gene, is responsible for amylose synthesis so that amylose content was dramatically reduced in cereal *wx* mutants. The *Wx-B1* locus on wheat chromosome 4A has major effect on starch peak viscosity, starch/flour swelling volume and noodle-making quality (Zhao *et al.,* 1998). Amylopectin is synthesized by the concerted action of ADP-glucose pyrophosphorylase (AGPase, encoded by gene *Sh2*), starch branching enzyme, soluble starch synthase (SSS) and starch debranching enzyme (Smith *et al.,* 1997). The subspecies, *japonica* and *indica* of rice differ significantly in structure and crystalline organization of starch granules as reflected by alkali disintegration and gelatinization. Alkali disintegration is controlled by a single gene *alk* on chromosome 6. Recent genomic analysis also mapped *alk*, genes for gelatinization and amylopectin chain length to a gene coding for SSIIa, an isoform of SSS (Umemoto *et al.,* 2002). The authors proposed that the allelic differences at the *SSIIa* locus are responsible for differences in amylopectin structure and are important for the chain elongation and further affect the gelinization (Umemoto *et al.,* 2002).

Reduction of enzyme activity for amylopectin synthesis has also been shown to increase the amylose content. The starch granule protein, SGP-1, an isoform of SSS and highly similar to SSIIa, was eliminated from wheat by crossing the three variants deficient in SGP-1A, SGP-1B and SGP-1D (Yamamori *et al.,* 2000). In this SGP-deficient wheat, the activity of other types of SSS was reduced considerably, A-type (large in size, 15-25 μm in diameter) granules were deformed and amylose content was increased to 37% (Yamamori *et al.,* 2000). With the availability of sequences for the genes coding for starch synthases, perfect markers were developed to monitor the starch quality in breeding programs.

6.2.4. Enzyme Activity

Amylase activity is very important for the malting quality of barley. Genes coding α- and β-amylases were isolated and many QTLs were identified for the their activities in barley and wheat (Table 4). Major QTLs for α-amylase activity were mapped to barley chromosomes 4H and 7H (Han *et al.*, 1997). These two QTLs also affected β-amylase activity, although effect was smaller (Clancy *et al.*, 2003). Compared to α-amylase, inheritance of β-amylase activity seems more complex, involving all the seven barley chromosomes, although the major QTLs were located on chromosomes 1H,

Table 4. Major QTLs for grain protein content and enzyme activity

Crop	Chromosome	Marker interval	% of P.V.[1]	Reference
I. Grain protein content				
Barley	6H	hvm	35.2	See *et al*. (2002)
Rice	6	C952 - Wx	13.0	Tan *et al*. (2001)
Sorghum	A	UMC23 - SSCIR88	18.4	Rami *et al*. (1998)
	C	BNL6.25 - UMC84	16.5 - 19.1	Rami *et al*. (1998)
	D	UMC138	11.0	Rami *et al*. (1998)
	H	BNL3.06 CSU94a	14.8	Rami *et al*. (1998)
Wheat	2D	wmc41	18.7	Prasad *et al*. (1999)
	6B	CDO365	72.0	Chee *et al*. (1998)
II. Alpha-amylase				
Barley	4H	WG622 - BCD402B	21.7	Han *et al*. (1997)
	7H	Brz - Amy2	29.5	Han *et al*. (1997)
Wheat	1B	gwm18 - glk483	7.9 - 14.7	Zanetti *et al*. (2000)
	3B	glk80 -psr1054	7.0 - 15.5	Zanetti *et al*. (2000)
	5A	prs644a - psr945a	13.0 - 13.9	Zanetti *et al*. (2000)
	5A	psr1149 - psr918b	20.0 -38.5	Zanetti *et al*. (2000)
	6A	psr008 - psr563a	13.5 - 17.7	Zanetti *et al*. (2000)
	7B	psr350 - pwir232b	7.7 -25.0	Zanetti *et al*. (2000)
III. Beta-amylase				
Barley	1H	MWG938 - MWG036A	20.7	Kaneko *et al*. (2001)
	1H	Hor	12.5 - 35.0	Clancy *et al*. (2003)
	2H	ABC156A - MWG858	20.1	Kaneko *et al*. (2001)
	4H	WG622 - BCD402B	10.5	Clancy *et al*. (2003)
	5H	CDO348B - ksuA1B	14.3	Kaneko *et al*. (2001)
	7H	Brz - Amy2	17.0	Clancy *et al*. (2003)
IV. Polyphenole oxidase				
Wheat	2D	FBA314	23.0	Demeke *et al*. (2001)
	2A	BCD307	12.0 - 16.0	Demeke *et al*. (2001)
	3B	BCD907	11.0 - 14.0	Demeke *et al*. (2001)
	6B	KSUG12a	12.0 - 14.0	Demeke *et al*. (2001)

% of P.V.[1] = percent of phenotypic variation

2H, 4H, 5H and 7H (Clancy *et al.*, 2003). All the major QTLs for activity of α- and β-amylases were strongly associated with the diastase power.

The α-amylase activity in the late-matured wheat grains is important because it is strongly associated with the preharvest sprouting, which reduces the wheat grain quality for food. Although five major QTLs were detected on five wheat chromosomes, chromosomes 5A and 7B had the largest effect (Zanetti *et al.*, 2000). No orthologous QTL was found between wheat and barley for α-amylase activity and, most probably, different regulatory mechanisms are involved in these two Triticeae crops as they experienced different selection processes.

Polyphenol oxidase (PPO) is an enzyme that seriously impacts noodle-making quality. High PPO activity causes the progressively browning or darkening of noodle and reduces the marketing value. Genes coding for PPO were located on wheat chromosomes of homoeologous group 6, 5B and 7D (Li *et al.*, 1999), although four major QTLs were detected on chromosomes 2A, 2D, 3B and 6B over the multiple years and at multiple locations.

7. SUMMARY AND OUTLOOK

The first four years of the 21st century saw great leap forward in cereal genomics. Draft sequences were reported in June of 2002 for the genomes of *indica* rice (Yu *et al.*, 2002) and *japonica* rice (Goff *et al.*, 2002); finished sequences of *japonica* rice chromosomes 1 (Sasaki *et al.*, 2002) and 4 (Feng *et al.*, 2002) were published at the end of 2002 and those for chromosome 10 were published in June of 2003 (The Rice Chromosome 10 Sequencing Consortium, 2003) and the maize genome sequencing project was initiated in 2002 (Chandler and Brendel, 2002). Furthermore, sequencing of the gene-rich regions of the wheat genome is being vigorously pursued (Gill and Appels, 2004). There is no doubt that genomes of all the cereals will be sequenced and most agronomically important genes, if not all, will be identified during the next decade. Once all the genes are sequenced, all available breeding lines can be genotyped and novel technologies developed for genetic engineering. As a result, we will be able to design varieties for specific areas or purposes, choose the best parents, target traits (genes) and perfect markers, and decide population size, generation and means of selection. All these can be done *in silico* before a cross is made. Genomics

will affect every aspect of cereal breeding and will modernize this old science, although genomic approaches will not succeed unless they are combined with traditional breeding programs.

8. APOLOGY AND ACKNOLEDGEMENTS

We apologize for any missing citation in the summarized data. This is contribution 04-212-B from Agriculture Experiment Station of Kansas State University.

9. REFERENCES

Ahmed T.A., Tsujimoto H., Sasakuma T. Identification of RFLP markers linked with heading date and its heterosis in hexaploid wheat. Euphytica 2000; 116: 111-119

Araki E., Miura H., Sawada S. Identification of genetic loci affecting amylose content and agronomic traits on chromosome 4A of wheat. Theor Appl Genet 1999; 98: 977-984

Autrique E., Nachit M.M., Monneveux P., Tanksley S.D., Sorrells M.E. Genetic diversity in durum wheat based on RFLPs, morphophysiological traits, and coefficient of parentage. Crop Sci 1996; 36: 735-742

Badr A., Müller K., Schafer-Pregl R., Rabey H.E., Effgen S., Ibrahim S.H.H., Pozzi C., Rohde W., Salamini F. On the origin and domestication history of barley. Mol Biol Evol 2000; 17: 499-510

Beecher B., Bettge A., Smidansky E., Giroux M. J. Expression of wild-type *Pinb* sequence in transgenic wheat complements a hard phenotype. Theor Appl Genet 2002a; 105:870-877

Beecher B., Bowman J., Martin J. M., Bettge A.D., Morris C.F., Blake T.K., Giroux M.J. Hordoindolines are associated with a major endosperm texture QTL in barley (*Hordeum vulgare*). Genome 2002b; 45: 584-591

Beecher B., Smidansky E.D., See D., Blake T.K. Giroux M.J. Mapping and sequence analysis of barley puroindolines. Theor Appl Genet 2001; 102: 833-840

Bernardo R. What if we knew all the genes for a quantitative trait in hybrid crops? Crop Sci 2001; 41:1-4

Berry D. A., Seltzer J.D., Xie C., Wright D.L., Jones E.S., Sebastian S., Smith J.S.C. Assessing probability of ancestry using simple sequence repeat profiles: applications to maize inbred lines and soybean varieties. Genetics 2003; 165: 331-342

Boppenmaier J., Melchinger A.E., Brunklaus-Jung E., Geiger H.H., Herrmann R.G. genetic distance for RFLPs in European maize inbreds. I. Relation to performance of flint x dent crosses for forage traits. Crop Sci 1992; 32:895-902

Brennecke K., Souz Neto A.J., Lugli J., Azevedo R.A. Asprate kinase in maize mutants *ask1-lt19* and *opaque-2*. Phytochemistry 1996; 41: 707-712

Brondani C., Rangel P., Brondani R., Ferreira M. QTL mapping and introgression of yield-related traits from *Oryza glumaepatula* to cultivated rice (*Oryza sativa*) using microsatellite markers. Theor Appl Genet 2002; 104: 1192-1203

Bullrich L., Appendino M.L., Tranquilli G., Lewis S. Dubcovsky J. Mapping of a thero-sensitive *earliness per se* gene on *Triticum monococcum* chromosome $1A^m$. Theor Appl Genet 2002; 105: 585-593

Burnett R.J., Larkins B.A. *Opaque 2* modifiers alter transcription of the 27-kDa gamma-zein genes in maize. Mol Gen Genet 1999; 261: 908-916

Cadalen T., Sourdille P., Charmet G., Tixier M.H., Gay G., Boeuf C., Bernard S., Leroy P., Bernard M. Molecular markers linked to genes affecting plant height in wheat using a doubled-haploid population. Theor Appl Genet 1998; 96: 933-940

Campbell K.G., Bergman C.J., Gualberto D.G., Anderson J.A., Girox M.J., Hareland G., Fulcher R.G., Sorrells M.E. Finney P.L. Quantitative trait loci associated with kernel traits in a soft X hard wheat cross. Crop Sci 1999; 39: 1184-1195

Chandler P.M., Marion-Poll A., Ellis M., Gubler F. Mutants at the *slender1* locus of barley cv Himalaya. molecular and physiological characterization. Plant Physiol 2002; 129: 181-190

Chandler V.L., Brendel V. The maize genome sequencing project. Plant Physiol 2002; 130: 1594-1597

Chee P.W., Elias E.M., Anderson J.A., Kianian S.F. Evaluation of a high grain protein QTL from *Triticum turgidum* L. var. *dicoccoides* in an adapted durum wheat background. Crop Sci 2001; 41: 295-301

Chen H.B., Martin J.M., Lavin M. Talbert L.E.. Genetic diversity in hard red spring wheat based on sequence-tagged-site PCR marker. Crop Sci 1994; 34: 1628-1632

Cheng C., Motohashi R., Tsuchimoto S., Fukuta Y., Ohtsubo H., Ohtsubo E. Polyphyletic origin of cultivated rice: based on the interspersion pattern of SINEs. Mol Biol Evol 2003; 20: 67-75

Cheng W.H., Taliercio E.W. Chourey P.S. The *miniature1 seed* locus of maize encodes a cell wall invertase required for normal development of endosperm and maternal cells in the pedicel. Plant Cell 1996; 8: 971-983

Childs K.L., Miller F.R., Cordonnier-Pratt M.M., Pratt L.H., Morgan P.W. Mullet J.E. The sorghum photoperiod sensitivity gene, *Ma3*, encodes a phytochrome B. Plant Physiol 1997; 113: 611-619

Clancy J.A., Han F., Ullrich S.E., the North American Barley Genome Project. Comparative mapping of β-amylase activity QTLs among three barley crosses. Crop Sci 2003; 43: 1043-1052

Coburn J.R., Temnykh S.V., Paul E.M., McCouch S.R. Design and application of microsatellite marker panels for semiautomated genotyping of rice (*Oryza sativa* L.). Crop Sci 2002; 42: 2092-2099

Coleman G.E., Clore A.M., JP Ranch, Higgins R., Lopes M.A., Larkins B.A. Expression of a mutant-zein creates the *floury2* phenotype in transgenic maize. Proc Natl Acad Sci USA 1997; 94: 7094-7097

Cubas P., Lauter N., Doebley J., Coen E. The TCP domain: a motif found in proteins regulating plant growth and development. Plant J 1999; 18: 215-222

Darlington H.F., Rouster J., Hoffmann L., Halford N.G., Shewry P.R., Simpson D.J. Identification and molecular characterisation of hordoindolines from barley grain. Plant Mol Biol 2001; 47: 785-794

Demeke T., Morris C.F., Campbell K.G., King G.E., Anderson J.A. Chang H-G. Wheat polyphenol oxidase: distribution and genetic mapping in three inbred line populations. Crop Sci 2001; 41: 1750-1757

Doebley J., Bacigalupo A., Stec A. Ineritance of kernel weight in two maize-teosinte hybrid populations: implications of crop evolution. J Heredity 1994; 85: 191-195

Doebley J., Stec A., Gustus C. *teosinte branched1* and the origin of maize: evidence for epistasis and the evolution of dominance. Genetics 1995; 141: 333-346

Doebley J., Stec A. Hubbard L. The evolution of apical dominance in maize. Nature 1997; 386: 485-488

Donini P., Law J.R., Koebner R.M.D., Reeves J.C., Cooke R.J. Temporal trends in the diversity of UK wheat. Theor Appl Genet 2000; 100: 912-917.

Dubreuil P., Charcosset A. Relationships among maize inbred lines and populations from European and North-American origins as estimated using RFLP markers. Theor Appl Genet 1999; 99: 473-480

Dudley J.W., Saghai Maroof M.A. and Rufener G.K. Molecular marker and grouping of parents in maize breeding programs. Crop Sci 1991; 31: 718-723

Dunford R.P., Yano M., Kurata N., Sasaki T., Huestis G., Rocheford T., Laurie D.A. Comparative mapping of the barley *Ppd-H1* photoperiod response gene region, which lies close to a junction between two rice linkage segments. Genetics 2002; 161: 825-834.

Dvorák J., Luo M-C., Yang Z-L., Zhang H-B. The structure of the *Aegilops tauschii* genepool and the evolution of hexaploid wheat. Theor Appl Genet 1998; 97: 657-670

Dvorák J., Luo M-C., Yang Z-L.. Restriction fragment length polymorphism and divergence in the genomic regions of high and low recombination in self-fertilizing and cross-fertilizing *Aegilops* species. Genetics 1998b; 148:423-434

Feng Q., Zhang Y., Wang S., Fu G., Huang Y., Li Y., Zhu J., Liu Y., Hu X., Jia P., *et al.* Sequence analysis of rice chromosome 4. Nature 2002; 420: 316-320

Friebe B., Jiang J., Raupp W.J., McIntosh R.A., Gill B.S. Characterization of wheat-alien translocations conferring resistance to disease and pests: current status. Euphytica 1996; 91: 59-87

Fu H., Dooner H.K. Intraspecific violation of genetic colinearity and its implications in maize. Proc Natl Acad Sci USA 2002; 99: 9573-9578

Gautier M.F., Aleman M.E., Guirao A., Marison D. Joudrier P. *Triticum aestivum* puroindolines, two basic cystine-rich seed proteins: cDNA sequence analysis and developmental gene expression. Plant Mol Biol 1994; 25: 43-57

Gautier M.F., Cosson P., Guirao A., Alary R. Joufrier P. Puroindoline genes are highly conserved in diploid ancestor wheats and related species but absent in tetraploid *Triticum* species. Plant Sci 2000; 153: 81-91

Gethi J.G., Labate J.A., Lamkey K.R., Smith M.E., Kresovich S. SSR Variation in important U.S. Maize Inbred Lines. Crop Sci 2002; 42: 951-957

Ghebru B., Schmidt R., Bennetzen J. Genetic diversity of Eritrean sorghum landraces assessed with simple sequence repeat (SSR) markers. Theor Appl Genet 2002; 105: 229-236

Gill B.S., Appels R. NSF/USDA/IGROW workshop on wheat genome sequencing. Genetics. 2004; 166: (in press).

Gillikin J.W., Zhang F., Coleman C.E., Bass H.W., Larkins B.A. Boston R.S. A defective signal peptide tethers the *floury-2* zein to the endoplasmic reticulum membrane. Plant Physiol 1997; 114: 345-352

Giroux M.J., Morris C.F. A glycine to serine change in puroindoline b is associated with wheat grain hardness and low levels of starch-surface friabilin. Theor Appl Genet 1997; 95: 857-864

Giroux M.J., Morris C.F. Wheat grain hardness results from highly conserved mutations in friabilin components puroindoline a and b. Proc Natl Acad Sci USA 1998; 95: 6262-6266

Giroux M.J., Shaw J., Barry G., Cobb B.G., Greene T., Okita T., Hannah L.C. A single gene mutation that increases maize seed weight. Proc Natl Acad Sci USA 1996; 93: 5824-5829

Goff S.A., Ricke D., Lan T.H., Presting G., Wang R., Dunn M., Clazebrook J., Sessions A., Oeller P., Varma H., *et al.* A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science 2002; 296: 92-100

Griffiths S., Dunford R.P., Coupland G., Laurie D.A. The evolution of CONSTANS-like gene families in barley, rice, and Arabidopsis. Plant Physiol 2003; 131: 1855-1867

Hayama R., Yokoi S., Tamaki S., Yano M., Shimamoto K. Adaptation of photoperiodic control pathways produces short-day flowering in rice. Nature 2003; 422: 719-722

Han F., Romagosa I., Ullrich S.E., Jones B.L., Hayes P.M., Wesenberg D.M. Molecular marker-assisted selection for malting quality traits in barley. Mol Breed 1997; 3: 427-437

Hart G.E., Schertz K.F., Peng Y., Syed N.H. Genetic mapping of *Sorghum bicolor* (L.) Moench QTLs that control variation in tillering and other morphological characters. Theor Appl Genet 2001; 103: 1232-1242

He P., Li J.Z., Zheng X.W., Shen L.S., Lu C.F., Chen Y., Zhu L. Comparison of molecular linkage maps and agronomic trait loci between DH and RIL population derived from the same rice cross. Crop Sci 2001; 41: 1240-1246

Heun M., Schäfer-Pregl R., Klawan D., Castagna R., Accerbi M., Borghi B., Salamini F. Site of einkorn wheat domestication identified by DNA fingerprinting. Science 1997; 278: 1312-1314

Hittalmani S., Huang N., Courtois B., Venuprasad R., Shashidhar H.E., Zhuang J-Y., Zheng K-L., Liu G-F., Wang G-C., Sidhu J.S., *et al.* Identification of QTL for growth- and grain yield-related traits in rice across nine locations of Asia. Theor Appl Genet 2003; 107: 679-690

Ho J., McCouch S., Smith M. Improvement of hybrid yield by advanced backcross QTL analysis in elite maize. Theor Appl Genet 2002; 105: 440-448

Hollick J.B., Chandler V.L. Epigenetic allelic states of a maize transcriptional regulatory locus exhibit overdominant gene action. Genetics 1998; 150: 891-897

Hua J.P., Xing Y.Z., Xu C.G., Sun X.L., Yu S.B., Zhang Q. Genetic dissection of an elite rice hybrid revealed that heterozygotes are not always advantageous for performance. Genetics 2002; 162: 1885-1895

Hua J., Xing Y., Wu W., Xu C., Sun X., Yu S., Zhang Q. Single-locus heterotic effects and dominance by dominance interactions can adequately explain the genetic basis of heterosis in an elite rice hybrid. Proc Natl Acad Sci USA 2003; 100: 2574-2579

Huang L., Brooks S.A., Li W., Fellers J.P., Trick H.N., Gill B.S. Map-based cloning of leaf rust resistance gene *Lr21* from the large and polyploid genome of bread wheat Genetics 2003; 164: 655-664

Huang X.Q., Börner A., Röder M. Ganal M. Assessing genetic diversity of wheat (*Triticum aestivum* L.) germplasm using microsatellite markers. Theor Appl Genet 2002; 105: 699-707

Huang X.Q., Cöster H., Ganal M.W., Röder M.S. Advanced backcross QTL analysis for the identification of quantitative trait loci alleles from wild relatives of wheat (*Triticum aestivum* L.). Theor Appl Genet 2003; 106: 1379-1389

Hubbard L., McSteen P., Doebley J., Hake S. Expression patterns and mutant phenotype of *teosinte branched1* correlate with growth suppression in maize and teosinte. Genetics 2002; 162: 1927-1935

Hunter B.G., Beatty M.K., Singletary G.W., Hamaker B.R., Dilkes B.P., Larkins B.A., Jung R. Maize *opaque* endosperm mutations create extensive changes in patterns of gene expression. Plant Cell 2002; 14: 2591-2612

Ikeda A., Ueguchi-Tanaka M., Sonoda Y., Kitano H., Koshioka M., Futsuhara Y., Matsuoka M., Yamaguchi J. *slender* rice, a constitutive gibberellin response mutant, is caused by a null mutation of the *slr1* gene, an ortholog of the height-regulating gene *GAI/RGA/RHT/D8*. Plant Cell 2001; 13: 999-1010

Ishimaru K., Yano M., Aoki N., Ono K., Hirose T., Lin S.Y., Monna L., Sasaki T., Ohsugi R. Toward the mapping of physiological and agronomic characters on a ricefunction map: QTL analysis and comparison between QTLs and expressed sequence tags. Theor Appl Genet 2001; 102: 793-800

Izawa T., Oikawa T., Tokutomi S., Okuno K., Shimamoto K. Phytochromes confer the photoperiodic control of flowering in rice (a short-day plant). Plant Jour 2000; 22: 391-399

Jordan D., Tao Y., Godwin I., Henzell R., Cooper M., McIntyre C. Prediction of hybrid performance in grain sorghum using RFLP markers. Theor Appl Genet 2003; 106: 559-567

Kaneko T., Zhang W., Takahashi H., Ito K., Takeda K. QTL mapping for enzyme activity and thermostability of β-amylase in barley (*Hordeum vulgare* L.). Breed Sci 2001; 51: 99-105

Kato K., Miura H., Sawada S. Detection of an earliness per se quantitative trait loci in the proximal region of wheat chromosome 5AL. Plant Breed 1999: 118: 391-394

Kato K., Miura H., Sawada S. Mapping QTLs controlling grain yield and its components on chromosome 5A of wheat. Theor Appl Genet 2000; 101: 1114-1121

Keller M., Karutz C., Schmid J.E., Stamp P., Winzeler M., Keller B., Messmer M.M. Quantitative trait loci for lodging resistance in a segregating wheat x spelt population. Theor Appl Genet 1999; 98: 1171-1182

Kemper E.L., Neto G.C., Papes F., Moraes K.C.M., Leite A., Arruda P. The role of *opaque2* in the control of lysine-degrading activities in developing maize endosperm. Plant Cell 1999; 11: 1981-1994

Kennard W., Phillips R., Porter R. Genetic dissection of seed shattering, agronomic, and color traits in American wildrice (*Zizania palustris* var. *interior* L.) with a comparative map. Theor Appl Genet 2002; 105: 1075-1086

Khush G.S. Green revolution: the way forward. Nat Rev Genet 2001; 2: 815–822

Kicherer S., Backes G., Walther U., Jahoor A. Localising QTLs for leaf rust resistance and agronomic traits in barley (*Hordeum vulgare* L.). Theor Appl Genet 2000; 100: 881-888

Klein R.R., Rodriguez-Herrera R., Schlueter J.A., Klein P.E., Yu Z.H., Rooney W.L. Identification of genomic regions that affect grain-mould incidence and other traits of agronomic importance in sorghum. Theor Appl Genet 2001; 102: 307-319

Kojima S., Takahashi Y., Kobayashi Y., Monna L., Sasaki T., Araki T., Yano M. *Hd3a*, a rice ortholog of the Arabidopsis *FT* gene, promotes transition to flowering downstream of *Hd1* under short-day conditions. Plant Cell Physiol 2002; 43: 1096-1105

Kosugi S., Ohashi Y. DNA binding and dimerization specificity and potential targets for the TCP protein family. Plant J 2002; 30: 337-348

Korzun V., Melz G., Börner A. RFLP mapping of the dwarfing (*Ddw1*) and a hairy peduncle (*Hp*) genes on chromosome 5 of rye (*Secale cereale* L.). Theor Appl Genet 1996; 92: 1073-1077

Korzun V., Röder M.S., Ganal M.W., Worland A.J., Law C.N. Genetic analysis of the dwarfing gene (Rht8) in wheat. Part I. Molecular mapping of Rht8 on the short arm of chromosome 2D of bread wheat (*Triticum aestivum* L.). Theor Appl Genet 1998; 96: 1104-1109

Labate J.A., Lamkey K.R., Mitchell S.E., Kresovich S., Sullivan H., Smith J.S.C. Molecular and historical aspects of corn belt dent diversity. Crop Sci 2003; 43: 80-91

Lai J., Messing J. Increasing maize seed methionine by mRNA stability. Plant Jour 2002; 30: 395-402

Laurie D.A., Pratchett N., Bezant J.H., Snape J.W. RFLP mapping of five major genes and eight quantitative trait loci controlling flowering time in a winter x spring barley (*Hordeum vulgare* L.) cross. Genome 1995; 38: 575-585

Li W., Faris J.D., Chittoor J., Leach J.E., Liu D.J., Chen P.D., Gill B.S. Genomic mapping of defense response genes in wheat. Theor Appl Genet 1999; 98:226-233.

Li W., Nelson J.C., Chu C.Y., Shi L.H., Huang S.H., Liu D.J. Chromosomal locations and genetic relationship of tiller and spike characters in wheat. Euphytica 2002; 125: 357-366

Li Z.K., Luo L.J., Mei H.W., Wang D.L., Shu Q.Y., Tabien R., Zhong D.B., Ying C.S., Stansel J.W., Khush G.S., *et al.* Overdominant epistatic loci are the primary genetic basis of inbreeding depression and heterosis in rice. I. Biomass and grain yield. Genetics 2001; 158: 1737-1753

Li Z., Pinson S.R.M., Park W.D., Paterson A.H., Stansel J.W. Epistasis for three grain yield components in rice (*Oryza sativa* L.). Genetics 1997; 145: 453-465

Li Z., Pinson S.R.M., Stansel J.W., Park W.D. Identification of quantitative trait loci (QTLs) for heading date and plant height in cultivated rice (*Oryza sativa* L.). Theor Appl Genet 1995; 91: 374-381

Li X., Qian Q., Fu Z., Wang Y., Xiong G., Zeng D., Wang X., Liu X., Teng S., Hiroshi F., *et al.* Control of tillering in rice. Nature 2003; 422: 618-621

Lillemo M., Morris C.F. A leucine to proline mutaion in proindoline b is frequently present in hard wheat from northern Europe. Theor Appl Genet 2000; 100: 1100-1107

Lin Y-R., Schertz K.F. Paterson A.H. Comparative analysis of QTLs affecting plant height and maturity across the *Poaceae*, in reference to an interspecific sorghum population. Genetics 1995; 141: 391-411

Lin H.X., Yamamoto T., Sasaki T., Yano M. Characterization and detection of epistatic interactions of 3 QTLs, *Hd1*, *Hd2*, and *Hd3*, controlling heading date in rice using nearly isogenic lines. Theor Appl Genet 2000; 101: 1021-1028

Liu S., Anderson J.A. Marker assisted evaluation of Fusarium head blight resistant wheat germplasm. Crop Sci 2003; 43: 760-766

Lou C., Griffin W.B., Branlard G., McNeil D.L. Comparison of low- and high molecular-weight wheat glutenin allele effects on flour quality. Theor Appl Genet 2001; 102: 1088-1098

Lu H., Bernardo R. Molecular marker diversity among current and historical maize inbreds. Theor Appl Genet 2001; 103: 613-617

Lu C., Shen L., Tan Z., Xu Y., He P., Chen Y., Zhu L. Comparative mapping of QTLs for agronomic traits of rice across environments using a doubled haploid population. Theor Appl Genet 1996; 93: 1211-1217

Lu H., Romero-Severson J., Bernardo R. Genetic basis of heterosis explored by simple sequence repeat markers in a random-mated maize population. Theor Appl Genet 2003; 107: 494-502

Lubbers E.L., Gill K.S., Cox T.S. Gill B.S. Variation of molecular markers among geographically diverse accessions of *Triticum tauschii*. Genome 1991; 34: 354-361

Lübberstedt T., Melchinger A.E., Dußlea C., Vuylstekeb M., Kuiperb M. Relationships among early European maize inbreds. IV. Genetic diversity revealed with aflp markers and comparison with rflp, rapd, and pedigree data. Crop Sci 2000; 40: 783-791

Luo L.J., Li Z.K., Mei H.W., Shu Q.Y., TabienM R., Zhong D.B., Ying C.S., Stansel J.W., Khush G.S., Paterson A.H. Overdominant epistatic loci are the primary genetic basis of inbreeding depression and heterosis in rice. II. Grain yield components. Genetics 2001; 158: 1755-1771

Malyshev S., Korzun V., Voylokov A., Smirnov V., Börner A. Linkage mapping of mutant loci in rye (*Secale cereale* L.). Theor Appl Genet 2001; 103:70-74

Manifesto M.M., Schlatter A.R., Hopp H.E., Suárez E.Y., Dubcovsky J. Quantitative evaluation of genetic diversity in wheat germplasm using molecular markers. Crop Sci 2001; 41: 682-690

Marquez-Cedillo L.A., Hayes P.M., Kleinhofs A., Legge W.G., Rossnage B.G., Sato K., Ullrich S.E., Wesenberg D.M. QTL analysis of agronomic traits in barley based on the doubled haploid progeny of two elite North American varieties representing different germplasm groups. Theor Appl Genet 2001; 103: 625-637

Matsuoka Y., Vigouroux Y., Goodman M.M., Sanchez J., Buckler E., Doebley J.F. A single domestication for maize shown by multilocus microsatellite genotyping. Proc Natl Acad Sci USA 2002; 99: 6080-6084

Melchinger A.E., Lee M., Lamkey K.R., Woodman W.W. Genetic diversity for restriction fragment length polymorphisms: relation to estimated genetic effects in maize inbreds. Crop Sci 1990; 30: 1033-1040

Milach S.C.K., Rines H.W. Phillips R.L. Molecular genetic mapping of dwarfing genes in oat. Theor Appl Genet 1997; 95: 783-790

Moeller D.A., Schaal B.A. Genetic relationships among native American maize accessions of the Great Plains assessed by RAPDs. Theor Appl Genet 1999; 99: 1061-1067

Moncada P., Martínez C.P., Borrero J., Chatel M., Gauch Jr H., Guimaraes E., Tohme J., McCouch S.R. Quantitative trait loci for yield and yield components in an *Oryza sativa* x *Oryza rufipogon* BC_2F_2 population evaluated in an upland environment. Theor Appl Genet 2001; 102: 41-52

Monna L., Kitazawa N., Yoshino R., Suzuki J., Masuda H., Maehara Y., Tanji M., Sato M., Nasu S., Minobe Y. Positional cloning of rice semidwarfing gene, *sd-1*: rice "green revolution gene" encodes a mutant enzyme involved in gibberellin synthesis. DNA Res. 2002a; 9: 11-17

Monna L., Lin H., Kojima S., Sasaki T., Yano M. Genetic dissection of a genomic region for a quantitative trait locus, *Hd3*, into two loci, *Hd3a* and *Hd3b*, controlling heading date in rice. Plant Cell 2002b; 12: 2473-2484

Moro G.L., Habben J.E., Hamaker B.R., Larkins B.A. Characterization of the variability for lysine content in normal and *opaque-2* maize endosperm. Crop Sci 1996; 36: 1651-1659

Morris C.F., Greenblatt G.A., Bettge A.D., Malkawi I. Isolation and characterization of multiple forms of friabilin. J Cereal Sci 1994; 20: 167-174

Morris C.F., Lillemo M., Simeone M.C., Giroux M.J., Babb S.L., Kidwell K.K. Prevalence of puroindoline grain hardness genotypes among historically significant North American spring and winter wheats. Crop Sci 2001; 41: 218-228

Mouradov A., Cremer F., Coupland G. Control of flowering time: interacting pathways as a basis for diversity. Plant Cell 2002; 14: S111-S130

Multani D.S., Briggs S.P., Chamberlin M.A., Blakeslee J.J., Murphy A.S. Johal G.S. Loss of an MDR transporter in compact stalks of maize *br2* and sorghum *dw3* mutants. Science 2003; 302: 81-84

Nagaraju J., Kathirvel M., Kumar R.R., Siddiq E.A., Hasnain S.E. Genetic analysis of traditional and evolved Basmati and non-Basmati rice varieties by using fluorescence-based ISSR-PCR and SSR markers. Proc Natl Acad Sci USA 2002; 99: 5836-5841

Nemoto Y., Kisaka M., Fuse T., Yano M., Ogihara Y. Characterization and functional analysis of three wheat genes with homology to the *CONSTANS* flowering time gene in transgenic rice. Plant J 2003; 36: 82-93

Ni J., Colowit P.M. Mackill D.J. Evaluation of genetic diversity in rice subspecies using microsatellite markers. Crop Sci 2002; 42: 601-607

Oria M.P., Hamaker B.R., Axtell J.D., Huang C-P. A highly digestible sorghum mutant cultivar exhibits a unique folded structure of endosperm protein bodies. Proc Natl Acad Sci USA 2000; 97: 5065-5070

Özkan. H., Brandolini A., Schäfer-Pregl R., Salamini F. AFLP Analysis of a collection of tetraploid wheats indicates the origin of emmer and hard wheat domestication in southeast Turkey. Mol Biol Evol 2002; 19: 1797-1801

Peng J., Carol P., Richards D.E., King K.E., Cowling R.J., Murphy G.P. Harberd N.P. The Arabidopsis GAI gene defines a signaling pathway that negatively regulates gibberellin response. Genes & Devel 1997; 11: 3194-3205

Peng, J., Richards D.E., Hartley N.M., Murphy G.P., Devos K.M., Flintham J.E., Beales J., Fish L.J., Worland A.J., Pelica F., *et al.* 'Green revolution' genes encode mutant gibberellin response modulators. Nature 1999; 400: 256-261

Pestsova E., Korzun V., Goncharov N.P., Hammer K., Ganal M.W., Röder M.S. Microsatellite analysis of *Aegilops tauschii* germplasm. Theor Appl Genet 2000; 101: 100-106

Pillen K., Zacharias A., Léon J. Advanced backcross QTL analysis in barley (*Hordeum vulgare* L.). Theor Appl Genet 2003; 107: 340-352

Poncet V., Martel E., Allouis S., Devos K., Lamy F., Sarr A., Robert T. Comparative analysis of QTLs affecting domestication traits between two domesticated 2 wild pearl millet (*Pennisetum glaucum* L., Poaceae) crosses. Theor Appl Genet 2002; 104: 965-975

Prasad M., Varshney R.K., Kumar A., Baylan H.S., Sharma P.C., Edwards K.J., Singh H., Dhaliwal H.S., Roy J.K., Gupta P.K. A microsatellite marker associated with a QTL from grain protein content on chromosome arm 2DL of bread wheat. Theor Appl Genet 1999; 99: 341-345

Prasad M., Varshney R.K., Roy J.K., Balyan H.S., Gupta P.K. The use of microsatellites for detecting DNA polymorphism, genotype identification and genetic diversity in wheat. Theor Appl Genet 2000; 100: 584-592

Rami J.-F., Dufour P., Trouche G., Fliedel G., Mestres C., Davrieux F., Blanchard P., Hamon P. Quantitative trait loci for grain quality, productivity, morphological and agronomical traits in sorghum (*Sorghum bicolor* L. Moench). Theor Appl Genet 1998; 97: 605-616

Röder M.S., Huang X-Q., Borner A., Ganal M. Wheat microsatellite diversity of a genebank collection in comparison to registered varieties. In: N.E. Pogna, M. Romano, E.A. Pogna, G. Galterio (eds.) *Proc 10th Intern Wheat Genet Symp*, Instituto Spermentale per la Cerealicoltura, Roma, Italy, 2003, pp. 625-627

Saito K., Miura K., Nagano K., Hayano-Saito Y., Araki H., Kato A. Identification of two closely linked quantitative trait loci for cold tolerance on chromosome 4 of rice and their association with anther length. Theor Appl Genet 2001; 103: 862-868

Sari-Gorla M., Krajewski P., Di Fonzo N., Villa M., Frova C. Genetic analysis of drought tolerance in maize by molecular markers. II. Plant height and flowering. Theor Appl Genet 1999; 99: 289-295

Sasaki A., Ashikari M., Ueguchi-Tanaka M., Itoh H., Nishimura A., Swapan D., Ishiyama K., Saito T., Kobayashi M., Khush G.S., *et al.* Green revolution: A mutant gibberellin-synthesis gene in rice. Nature 2002a; 416: 701-702

Sasaki A., Itoh H., Gomi K., Ueguchi-Tanaka M., Ishiyama K., Kobayashi M., Jeong D.-H., An G., Kitano H., Ashikari M. *et al.* Accumulation of phosphorylated repressor for gibberellin signaling in an F-box mutant. Science 2003; 299: 1896-1898

Sasaki T., Matsumoto T., Yamamoto K., Sakata K., Baba T., Katayose Y., Wu J., Nilimura Y., Cheng Z., Nagamura Y., *et al.* The genome sequence and structure of rice chromosome 1. Nature 2002b; 420: 312-316

Sayed-Tabatabaei B.E., Shahnejat-Bushehri A.A. Assessment of genetic similariy among wheat cultivars using RAPD and AFLP techniques. In: N.E. Pogna, M. Romano, E.A. Pogna, G. Galterio (eds.) *Proc 10th Intern Wheat Genet Symp*, Instituto Spermentale per la Cerealicoltura, Roma, Italy, 2003, pp. 631-633

Schmidt R.J., Burr F.A., Burr B. Transposon tagging and molecular analysis of maize regulatory locus *opaque-2*. Science 1987; 238: 960-963

Schon C.C., Melchinger A.E., Boppenmaier J., Brunklaus-Jung E., Herrmann R.D., Seitzer J.F. RFLP mapping in maize: quantitative trait loci affecting testcross performance of elite European flint lines. Crop Sci 1994; 34: 378-389

See D., Kanazin V., Kephart K., Blake T. Mapping genes controlling variation in barley grain protein concentration. Crop Sci 2002; 42: 680-685

Segal G., Song R., Messing J. A new opaque variant of maize by a single dominant RNA-interference-inducing transgene. Genetics 2003; 165: 387-397

Senior M.L., Murphy J.P., Goodman M.M. Stuber C.W. Utility of SSRs for determining genetic similarities and relationships in maize using an agarose gel system. Crop Sci 1998; 38: 1088-1098

Shah M.M., Gill K.S., Baeniziger P.S., Yen Y., Kaeppler S.M., Ariyarathne H.M. Molecular mapping of loci for agronomic traits on chromosome 3A of bread wheat. Crop Sci 1999; 39: 1728-1732

Smidansky E.D., Clancy M., Meyer F.D., Lanning S.P., Blake N.K., Talbert L.E. Giroux M.J. Enhanced ADP-glucose pyrophosphorylase activity in wheat endosperm increases seed yield. Proc Natl Acad Sci USA 2002; 99: 1724-1729

Smith A.M., Denyer K., Martin C. The synthesis of the starch granule. Ann Rev Plant Physiol Plant Mol Biol 1997; 48: 67-87

Song R., Messing J. Gene expression of a gene family in maize based on noncollinear haplotypes. Proc Natl Acad Sci USA 2003; 100: 9055-9060

Song W.-Y., Wang G.-L., Chen L.-L., Kim H.-S., Pi L.-Y., Holsten T., Gardner J., Wang B., Zhai W.-X., Zhu L.-H., *et al.* A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa21*. Science 1995; 270: 1804-1806

Sourdille P., Perretant M.R., Charmet G., Leroy P., Gautier M.F., Joudrier P., Nelson J.C., Sorrells M.E., Bernard M. Linkage between RFLP markers and genes affecting kernel hardness in wheat. Theor Appl Genet 1996; 93: 580-586

Souza E., Fox P.N., Byerlee D., Skovmand B. Spring wheat diversity in two developing countries. Crop Sci 1994; 34: 774-783

Spielmeyer W., Ellis M.H., Chandler P.M. Semidwarf (*sd-1*), "green revolution" rice, contains a defective gibberellin 20-oxidase gene. Proc Natl Acad Sci USA 2002; 99: 9043-9048

Stuber C.W., Lincoln S.E., Wolff D.W., Helentjaris T., Lander E.S. Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. Genetics 1992; 132: 823-839

Sun Y., Carneiro N., Clore A.M., Moro G.L., Habben J.E., Larkins B.A. Characterization of maize elongation factor 1a and its relationship to protein quality in the endosperm. Plant Physiol 1997; 115: 1101-1107

Takahashi Y., Shomura A., Sasaki T., Yano M. *Hd6*, a rice quantitative trait locus involved in photoperiod sensitivity, encodes the subunit of protein kinase CK2. Proc Natl Acad Sci USA 2001; 98: 7922-7927

Takeda T., Suwa Y., Suzuki M., Kitano H., Ueguchi-Tanaka M., Ashikari M., Matsuoka M., Ueguchi C. The *OsTB1* gene negatively regulates lateral branching in rice. Plant Jour 2003; 33: 513-520

Takeuchi Y., Lin S.Y., Sasaki T., Yano M. Fine linkage mapping enables dissection of closely linked quantitative trait loci for seed dormancy and heading in rice. Theor Appl Genet 2003; 107: 1174-1180

Talbert L.E., Smith L.Y. Blake N.K. More than one origin of hexaploid wheat is indicated by sequence comparison of low-copy DNA. Genome 1998; 41: 402-407

Tan Y.F., Sun M., Xing Y.Z., Hua J.P., Sun X.L., Zhang Q.F. Corke H. Mapping quantitative trait loci for milling quality, protein content and color characteristics of rice using a recombinant inbred line population derived from an elite rice hybrid. Theor Appl Genet 2001; 103: 1037-1045

Tanksley S.D., McChouch S.R. Seed banks and molecular maps: unlocking genetic potential from the wild. Science 1997; 277: 1063-1066.

Tanksley S.D., Nelson J.C. Advanced backcross QTL analysis: a method for the simultaneous discoveryand transfer of valuable QTLs from unadapted germplasm into elite breeding lines. Theor Appl Genet 1996; 92: 191-203

Tenaillon M.I., Sawkins M.C., Long A.D., Gaut R.L., Doebley J.F., Gaut B.S. Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). Proc Natl Acad Sci USA 2001; 98: 9161-9166

Teulat B., Merah O., Souyris I., This D.. QTLs for agronomic traits from a Mediterranean barley progeny grown in several environments. Theor Appl Genet 2001; 103: 774-787

The Rice Chromosome 10 Sequencing Consortium. In-depth view of structure, activity, and evolution of rice chromosome 10. Science 2003; 300: 1566-1569

Thomson M.J., Tai T.H., McClung A.M., Lai X.-H., Hinga M.E., Lobos K.B., Xu Y., Martinez C.P., McCouch S.R. Mapping quantitative trait loci for yield, yield components and morphological traits in an advanced backcross population between

Oryza rufipogon and the *Oryza sativa* cultivar Jefferson. Theor Appl Genet 2003; 107: 479-493

Thornsberry J.M., Goodman M.M., Doebley J., Kresovich S., Nielsen D. Buckler E.S. *Dwarf8* polymorphisms associate with variation in flowering time. Nature Genet 2001; 28: 286-289

Tranquilli G., Dubcovsky J. Epistatic interaction between vernalization genes *Vrn-A^{m}1* and *Vrn-A^{m}2* in diploid wheat. Jour Hered 2000; 91: 304-306

Tranquilli G., Lijavetzky D., Muzzi G., Dubcovsky J. Genetic and physical characterization of grain texture-related loci in diploid wheat. Mol Gen Genet 1999; 262: 846-850

Trevaskis B., Bagnall D.J., Ellis M.H., Peacock W.J. Dennis E.S. MADS box genes control vernalization-induced flowering in cereals. Proc Natl Acad Sci USA 2003; 100:13099-13104

Troyer A.F., Rocheford T.R. Germplasm ownership: related corn inbreds. Crop Sci 2002; 42: 3-11

Umemoto T., Yano M., Satoh H., Shomura A., Nakamura Y. Mapping of a gene responsible for the difference in amylopectin structure between *japonica*-type and *indica*-type rice varieties. Theor Appl Genet 2002; 104: 1-8

Unger E., Parsons R.L., Schmidt R.J., Bowen B., Roth B.A. Dominant negative mutants of *opaque2* suppress transactivation of a 22-kd zein promoter by *opaque2* in maize endosperm cells. Plant Cell 1993; 5: 831-841

Uptmoor R., Wenzel W., Friedt W., Donaldson G., Ayisi K., Ordon F. Comparative analysis on the genetic relatedness of *Sorghum bicolor* accessions from Southern Africa by RAPDs, AFLPs and SSRs. Theor Appl Genet 2003; 106: 1316-1325

Vasal S.K. The quality protein maize story. Food Nutr Bull 2000; 21: 445-450

Veldboom L.R., Lee M. Molecular-marker-facilitated studies of morphological traits in maize. II. Determination of QTLs for grain yield and yield components. Theor Appl Genet 1994; 89: 51-458

Venuprasad R., Shashidhar H.E., Hittalmani S., Hemamalini G.S. Tagging quantitative trait loci associated with grain yield and root morphological traits in rice (*Oryza sativa* L.) under contrasting moisture regimes. Euphytica 2002; 128: 293-300

Vigouroux Y., McMullen M., Hittinger C.T., Houchins K., Schulz L., Kresovich S., Matsuoka Y., Doebley J.F. Identifying genes of agronomic importance in maize by screening microsatellites for evidence of selection during domestication. Proc Natl Acad Sci USA 2002; 99: 9650-9655

Wang R-L., Stec A., Hey J., Lukens L., Doebley J. The limits of selection during maize domestication. Nature 1999; 398: 236-239

Wang X., Stumpf D.K. Larkins B.A. Aspartate kinase 2. A candidate gene of a quantitative trait locus influencing free amino acid content in maize endosperm. Plant Physiol 2001; 125: 1778-1787

Warburton M.L., Xia X., Crossa J., Franco J., Melchinger A.E., Frisch M., Bohn M., Hoisington D. Genetic Characterization of CIMMYT inbred maize lines and open pollinated populations using large scale fingerprinting methods. Crop Sci 2002; 42: 1832-1840

Wu X.R., Chen Z.H., Folk W.R. Enrichment of cereal protein line content by altered tRNA^{lys} coding during protein synthesis. Plant Biotech Jour 2003; 1: 187-194

Xiao J., Li J., Yuan L., Tanksley S.D. Dominance is the major genetic basis of heterosis in rice as revealed by QTL analysis using molecular markers. Genetics 1995; 140: 745-754

Xiao J., Li J., Grandillo S., Ahn S.N., Yuan L., Tanksley S.D., McCouch S.R. Identification of trait-improving quantitative trait loci alleles from a wild rice relative, *Oryza rufipogon*. Genetics 1998; 150: 899-909

Xiao J., Li J., Yuan L., McChouch S.R., Tanksley S.D. Genetic diversity and its relationship to hybrid performance and heterosis in rice as revealed by PCR-based markers. Theor Appl Genet 1996a; 92: 637-643

Xiao J., Li J., Yuan L., Tanksley S.D. Identification of QTLs affecting traits of agronomic importance in recombinant inbred population derived from a subspecific rice cross. Theor Appl Genet 1996b; 92: 230-244

Xing Y., Tan Y., Hua J., Sun X., Xu C., Zhang Q. Characterization of the main effects, epistatic effects and their environmental interactions of QTLs on the genetic basis of yield traits in rice. Theor Appl Genet 2002; 105: 248-257

Yamamori M., Fujita S., Hayakawa K., Matsuki J., Yasui T. Genetic elimination of a starch granule protein, SGP-1, of wheat generates an altered starch with apparent high amylose. Theor Appl Genet 2000; 101: 21-29

Yan L., Loukoianov A., Tranquilli G., Helguera M., Fahima T., Dubcovsky J. Positional cloning of the wheat vernalization gene VRN1. Proc Natl Acad Sci USA 2003; 100: 6263-6268

Yang G.P., Saghai Maroof M.A., Xu C.G., Zhang Q., Biyashev R.M. Comparative analysis of microsatellite DNA polymorphism in landraces and cultivars of rice. Mol Gen Genet 1994; 245: 187-194

Yang W., de Oliveira A.C., Godwin I., Schertz K., Bennetzen J.L. Comparison of DNA marker technologies in characterizing plant genome diversity: variability in Chinese Sorghums. Crop Sci 1996; 36: 1669-1676

Yano M., Katayose Y., Ashikari M., Yamanouchi U., Monna L., Fuse T., Baba T., Yamamoto K., Umehara Y., Nagamura Y., *et al. Hd1*, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the Arabidopsis flowering time gene CONSTANS. Plant Cell 2000; 12: 2473-2484

Yano M., Kojima S., Takahashi Y., Lin H., Sasaki T. Genetic control of flowering time in rice, a short-day plant. Plant Physiol 2001; 127: 1425-1429

Yin X., Stam P., Dourleijn C.J., Kropff M.J. AFLP mapping of quantitative trait loci for yield-determining physiological characters in spring barley. Theor Appl Genet 1999; 99: 244-253

Yu J., Hu S., Wang J., Wong G.K.S., Li S., Liu B., Deng Y., Dai L., Zhou Y., Zhang X., *et al.* A draft sequence of the rice genome (*Oryza sativa* L. ssp *indica*). Science 2002; 296: 79-92

Yu S.B., Li J.X., Xu C.G., Tan Y.F., Gao Y.J., Li X.H., Zhang Q., Saghai-Maroof M.A. Importance of epistasis as the genetic basis of heterosis in an elite rice hybrid. Proc Natl Acad Sci USA 1997; 94: 9226-9231

Zanetti S., Winzeler M., Keller M., Keller B., Messmer M. Genetic analysis of pre-harvest sprouting resistance in a wheat × spelt cross. Crop Sci 2000; 40: 1406-1417

Zhang X., Li C., Wang L., Wang H., You G., Dong Y. An estimation of the minimum number of SSR alleles needed to reveal genetic relationships in wheat varieties. I. Information from large-scale planted varieties and cornerstone breeding parents in Chinese wheat improvement and production. Theor Appl Genet 2002; 106: 112-117

Zhang X., You G.X., Wang L.F. An estimation of the minimum number of SSR alleles needed to reveal genetic relationships in wheat varieties: information from 96 random accessions with maximized genetic diversity. In: N.E. Pogna, M. Romano, E.A. Pogna,

G. Galterio (eds.) *Proc 10th Intern Wheat Genet Symp*, Instituto Spermentale per la Cerealicoltura, Roma, Italy, 2003, pp. 545-547

Zhang Q., Zhou Z.Q., Yang G.P., Xu C.G., Liu K.D., Saghai Maroof M.A. Molecular marker heterogosity and hybrid performance in *indica* and *japonica* rice. Theor Appl Genet 1996; 93: 1218-1224

Zhao X.C., Batey I.L., Sharp P.J., Crosbie G.B., Barclay I., Wilson R., Moell M.K., Appels R. A single genetic locus associated with starch granule properties and noodle quality in wheat. J Cereal Sci 1998; 27: 7-13

Zhuang J-Y., Lin H-X., Lu J., Qian H-R., Hittalmani S., Huang N., Zheng K-L. Analysis of QTL2environment interaction for yield components and plant height in rice. Theor Appl Genet 1997; 95: 799-808

Zhuang J-Y., Fan Y-Y., Rao Z-M., Wu J-L., Xia Y-W., Zheng K-L. Analysis on additive effects and additive-by-additive epistatic effects of QTLs for yield traits in a recombinant inbred line population of rice. Theor Appl Genet 2002; 105: 1137-1145

LIST OF CONTRIBUTORS

OLIN D. ANDERSON
USDA-ARS Western Regional Research Center, Albany, CA 94710, USA

BALTAZAR A. ANTONIO
National Institute of Agrobiological Sciences, 1-2, Kannondai 2-chome, Tsukuba, Ibaraki 305-8602, Japan

GUNTER BACKES
Risø National Laboratory, Plant Research Department, Frederiksborgvej 399, DK–4000 Roskilde, Denmark

ANDREAS BÖRNER
Institute of Plant Genetics and Crop Plant Research (IPK), Corrensstrasse 3, D06466 Gatersleben, Germany

VICTORIA CAROLLO
USDA-ARS, Cornell University, Ithaca, NY 14853, USA

LARS ERIKSEN
Sejet Plantbreeding, Nørremarksvej 67, DK–8700 Horsens, Denmark

BIKRAM S. GILL
WGRC, Department of Plant Pathology, Throckmorton Plant Science Center, Kansas State University, Manhattan, KS 66506-5502, USA

KULVINDER S. GILL
Department of Crop and Soil Sciences, Washington State University, Pullman, WA 99164-6420, USA

ANDREAS GRANER
Institute of Plant Genetics and Crop Plant Research (IPK), Corrensstrasse 3, D06466 Gatersleben, Germany

PUSHPENDRA K. GUPTA
Department of Genetics and Plant Breeding, Ch. Charan Singh University, Meerut- 250004, India

AHMED JAHOOR
Risø National Laboratory, Plant Research Department, Frederiksborgvej 399, DK–4000 Roskilde, Denmark

POLAVARPU B. KAVIKISHORE
Department of Genetics, Osmania University, Hyderabad 500 007, India

ROBERT M.D. KOEBNER
Crop Genetics Department, John Innes Centre, Colney Lane, Norwich NR4 7UH, UK

VIKTOR KORZUN,
Lochow Petkus GmbH, PF 1197, D-29296 Bergen, Germany

PETER LANGRIDGE
Australian Centre for Plant Functional Genomics, The University of Adelaide, PMB1, Glen Osmond, SA 5064, Australia

GERRAD LAZO
USDA-ARS Western Regional Research Center, Albany, CA 94710, USA

WANLONG LI
WGRC, Department of Plant Pathology, Throckmorton Plant Science Center, Kansas State University, Manhattan, KS 66506-5502, USA

SERGIY LOPATO
Australian Centre for Plant Functional Genomics, The University of Adelaide, PMB1, Glen Osmond, SA 5064, Australia

KLAUS P. MAYER
MIPS, Institute for Bioinformatics, National Research Centre for Environment and Health (GSF), Ingolstädter Landstr. 1, 85764 Neuherberg, Germany

DAVID E. MATTHEWS
USDA-ARS, Cornell University, Ithaca, NY 14853, USA

ANDREW S. MILLIGAN
Australian Centre for Plant Functional Genomics, The University of Adelaide, PMB1, Glen Osmond, SA 5064, Australia

EVIATAR NEVO,
Institute of Evolution, University of Haifa 31905, Israel

ANDREW H. PATERSON,
University of Georgia, 111 Riverbend Road, Rm 228, GA 30602 Athens, USA

CARLO POZZI,
Fondazione Parco Tecnologico Padano, via Haussmann 7, 26900 Lodi, Italy

LAURA ROSSINI
Università degli Studi di Milano, Dipartimento di Produzione Vegetale, via Celoria 2, 20133 Milan, Italy

STEPHEN RUDD
MIPS, Institute for Bioinformatics, National Research Centre for Environment and Health (GSF), Ingolstädter Landstr. 1, 85764 Neuherberg, Germany

FRANCESCO SALAMINI
Max-Planck Institute for Plant Breeding, Carl von Linnè weg 10, 50829 Cologne, Germany

SILVIO SALVI
Department of Agroenvironmental Sciences and Technology, Viale Fanin 44, University of Bologna, Italy

TAKUJI SASAKI
National Institute of Agrobiological Sciences, 1-2, Kannondai 2-chome, Tsukuba, Ibaraki 305-8602, Japan

ALAN H. SCHULMAN,
MTT Agrifood Research, Jokioinen, and University of Helsinki, Helsinki, Finland

DARYL J. SOMERS,
Agriculture and Agri-Food Canada – Cereal Research Centre, 195 Dafoe Rd, Winnipeg, Manitoba, R3T-2M9 Canada

HEIKO SCHOOF
MIPS, Institute for Bioinformatics, National Research Centre for Environment and Health (GSF), Ingolstädter Landstr. 1, 85764 Neuherberg, Germany

MARK E. SORRELLS
Department of Plant Breeding, Cornell University, Ithaca, New York 14853, USA

NESE SREENIVASULU
Institute of Plant Genetics and Crop Plant Research (IPK), Corrensstrasse 3, D06466 Gatersleben, Germany

NILS STEIN
Institute of Plant Genetics and Crop Plant Research (IPK), Corrensstrasse 3, D06466 Gatersleben, Germany

ROBERTO TUBEROSA
Department of Agroenvironmental Sciences and Technology, Viale Fanin 44, University of Bologna, Italy

RAJEEV K. VARSHNEY,
Institute of Plant Genetics and Crop Plant Research (IPK), Corrensstrasse 3, D06466 Gatersleben, Germany

ALBERTO VECCHIETTI,
Fondazione Parco Tecnologico Padano, via Haussmann 7, 26900 Lodi, Italy

YEISOO YU
Arizona Genomics Institute, Department of Plant Sciences, University of Arizona, 303 Forbes Building, Tucson, AZ 85721, USA

WINFRIEDE WESCHKE
Institute of Plant Genetics and Crop Plant Research (IPK), Corrensstrasse 3, D06466 Gatersleben, Germany

ROD A. WING
Arizona Genomics Institute, Department of Plant Sciences, University of Arizona, 303 Forbes Building, Tucson, AZ 85721, USA

LIST OF REVIEWERS

Olin D. Anderson, USDA-ARS, Albany, USA
Rudi Appels, Department of Agriculture Western Australia, Australia
Michael Baum, ICARDA, Aleppo, Syria
Michel Beckert, UMR INRA-UBP,Clermont-Ferrand Cedex 2, France
Michel Bernard, UMR INRA-UBP, Clermont-Ferrand Cedex 2, France
Tom Blake, Montana State University, Bozeman, USA
Benzamin Burr, Brookhaven National Laboratory, Upton, USA
Michel Delseny, University of Perpignan, Perpignan, France
Catherine Feuillet, University of Zurich, Zurich, Switzerland
Mike Gale, John Innes Centre, Norwich, UK
Anil Grover, University of Delhi South Campus, New Delhi, India
Michael Holdsworth, University of Nottingham, Nottingham, UK
Pankaj Jaiswal, Cornell University, Ithaca, USA
Beat Keller, University of Zurich, Zurich, Switzerland
Peter Langridge, University of Adelaide, Glen Osmond, Australia
Diane E. Mather, McGill University, Quebec, Canada
Andrew H. Paterson, University of Georgia, Athens, USA
Wayne Powell, Scottish Crop Research Institute, Dundee, UK
Wusrika Ramakrishna, Michigan Tech University, Houghton,USA
J.-Antoni Rafalski, DuPont Crop Genetics, Wilmington, USA
Stephen Rudd, MIPS, GSF, Munich,Germany
Renate Schmidt, MPI of Molecular Plant Physiology, Golm, Germany
Rama Shankar Singh, McMaster University, Hamilton, Canada
Sudhir K. Sopory, ICGEB, New Delhi, India
Satoshi Tabata, Kazusa DNA Research Institute, Kisarazu, Chiba, Japan
WT Bill Thomas, Scottish Crop Research Institute, Dundee, UK
Roberto Tuberosa, University of Bologna, Bologna, Italy
Akhilesh K. Tyagi, University of Delhi South Campus, New Delhi, India
Mark Vaudin, Monsanto, St. Louis, USA
Robbie Waugh, Scottish Crop Research Institute, Dundee, UK
Roger Wise, Iowa State University, Ames, USA
Ulrich Wobus, IPK, Gatersleben, Germany