Genomics-Assisted Crop Improvement

Volume 2
Genomics Applications in Crops

Rajeev K. Varshney
Roberto Tuberosa
Editors

Springer

Genomics-Assisted Crop Improvement

Genomics-Assisted Crop Improvement

Vol 2: Genomics Applications in Crops

Edited by

Rajeev K. Varshney
ICRISAT, Patancheru, India

and

Roberto Tuberosa
University of Bologna, Italy

Springer

A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN 978-1-4020-6296-4 (HB)
ISBN 978-1-4020-6297-1 (e-book)

Published by Springer,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

www.springer.com

Printed on acid-free paper

CONTENTS

FOREWORD TO THE SERIES: GENOMICS-ASSISTED CROP IMPROVEMENT

Genetic markers and their application in plant breeding played a large part in my research career, so I am delighted to have the opportunity to write these notes to precede the two volumes on 'Genomics-assisted crop improvement'. Although I am not so old, I go right back to the beginning in 1923 when Karl Sax described how 'factors for qualitative traits' (today's genetic markers) could be used to select for 'size factors' (today's QTLs and genes for adaptation). But it was clear to me 40 years ago that even then plant breeders clearly understood how genetic markers could help them - if only they actually had the markers and understood the genetics underlying their key traits. It was not clear to me that it was going to take until the next century before marker-aided selection would become routine for crop improvement.

In the 1960s only 'morphological' markers were available to breeders. As a research student at Aberystwyth, I worked with Des Hayes at the Welsh Plant Breeding Station when he was trying to develop an F_1 hybrid barley crop based on a male sterility gene linked to a DDT resistance gene. The idea was to link the male fertile allele with susceptibility and then kill the fertile plants off in segregation populations by dousing the field with DDT. Rachel Carson's 'Silent Spring' ensured that idea never flew.

Then I moved to the Plant Breeding Institute in Cambridge where anyone working alongside the breeders in those early days could not help but be motivated by breeding. Protein electrophoresis raised the first possibility of multiple neutral markers and we were quick to become involved in the search for new isozyme markers in the late 1970s and early 1980s. Probably only the linkage between wheat endopeptidase and eyespot resistance was ever used by practical breeders, but we had an immense amount of fun uncovering the genetics of a series of expensive markers with hardly any polymorphism, all of which needed a different visualisation technology!

During this same period, of course, selection for wheat bread-making quality using glutenin subunits was being pioneered at the PBI, and is still in use around the world. These were the protein equivalent of today's 'perfect' or 'functional' markers for specific beneficial alleles. Such markers - although of course DNA-based, easy and economical to use, amenable to massively high throughput and available for all key genes in all crops - are exactly where we want to end up.

Proteins were superseded by RFLPs and in 1986 we set out to make a wheat map, only with the idea of providing breeders with the effectively infinite number of mapped neutral markers that they had always needed. We revelled in this massively expensive job, funded by a long-suffering European wheat breeding industry, of creating the first map with a marker technology so unwieldy that students today would not touch it with a bargepole, let alone plant breeders. This was, of course, before the advent of PCR, which changed everything.

The science has moved quickly and the past 20 years have seen staggering advances as genetics segued into genomics. We have seen a proliferation of maps, first in the major staples and later in other crops, including 'orphan' species grown only in developing countries. The early maps, populated with isozyme markers and RFLPs, were soon enhanced with more amenable PCR-based microsatellites, which are now beginning to give way to single nucleotide polymorphisms. These maps and markers have been used, in turn, to massively extend our knowledge of the genetic control underlying yield and quality traits. The relatively dense maps have allowed whole genome scans which have uncovered all regions of the genome involved in the control of key adaptive traits in almost all agricultural crops of any significance.

More amazing is the fact that we now have the whole genome DNA sequences of not one but four different plant genomes - *Arabidopsis*, rice and poplar and sorghum. Moreover cassava, cotton, and even maize could be added to the list before these volumes are published. Other model genomes where sequencing has been started include Aquilegia (evolutionary equidistant between rice and arabidopsis), Mimulus (for its range of variation) and Brachypodium (a small-genome relative of wheat and barley).

Two other components deserve mention. The first is synteny, the tendency for gene content and gene order to be conserved over quite distantly related genomes. Ironically, synteny emerged from comparisons between early RFLP maps and probably would not have been observed until we had long genomic sequences to compare had we started with PCR-based markers that require perfect DNA primer sequence match. The possibility of being able to predict using genetic information and DNA sequence gained in quite distantly related species has had a remarkable unifying effect on the research community. Ten years ago you could work away at your own favourite crop without ever talking to researchers and breeders elsewhere. Not so today. Synteny dictates that genome researchers are part of one single global community.

The second component is the crop species and comparative databases that we all use on a daily basis. The selfless curators, that we have all taken for granted, deserve mention and ovation here because, while the rest of us have been having fun in the lab, they have been quietly collecting and collating all relevant information for us to access at the press of a button. This is a welcome opportunity to acknowledge these unsung heroes, and of course, their sponsors.

The practical application of markers and genomics to crop improvement has been much slower to emerge. While endopeptidase and the glutenin gels continue to see

use in wheat breeding, marker-aided selection (MAS) using DNA markers has, in both public breeding and the multinationals, emerged only in the last few years and examples of new varieties that have been bred using MAS are still few and far between. This will change, however, as the cost of marker data points continues to plummet and the application of high-throughput methods moves the technology from breeding laboratories to more competitive outsourced service providers.

The post-RFLP period and the new opportunities for deployment of economical high-throughput markers are the subjects of these volumes. The first volume deals with platforms and approaches while the second covers selected applications in a range of crop plants. The editors, Rajeev Varshney and Roberto Tuberosa, are to be congratulated on bringing together an authorship of today's international leaders in crop plant genomics.

The end game, where plant breeders can assemble whole genomes by manipulating recombination and selecting for specific alleles at all key genes for adaptation is still a very long way off. But these two volumes are a unique opportunity to take stock of exactly where we are in this exciting arena, which is poised to revolutionise plant breeding.

Mike Gale, FRS
John Innes Foundation Emeritus
Fellow
John Innes Centre
Norwich
United Kingdom

FOREWORD

According to the World Bank, approximately 1.2 billion people in absolute poverty live on less than US$ 1 per day, while nearly twice this number live on less than US$ 2 per day. About 90% of the world's poor comprise rural, resource-poor farmers and their families, and landless poor who depend on agriculture for their livelihoods. However, poor farmers live and work in regions that continue to be important sources of genetic diversity. While poor farmers, in general, cannot use "modern" (high-input) crop varieties, selected for optimal performance within a narrow range of highly managed environmental conditions, the more wealthy ones have replaced a wide range of traditional crops and varieties with a limited number of "modern" varieties of major crops.

However, it must be said to the credit of resource-poor farmers that they plant their own seeds (landraces) and manage their farms in a manner that allows the varieties to evolve. They select plant types rather than varieties based on their own observations and specific needs. They are in a way, responsible for maintaining the genetic diversity that is essential to the continued evolution and adaptation of plant genotypes.

They are therefore well placed to supply formal plant breeding systems with new genetic material that is urgently needed. Their empirical knowledge of the characteristics of specific plant types could help breeders identify the source of valuable traits for introgression into elite crop varieties. In this context, plant breeders need to screen germplasm from regions of low-resource agriculture but rich in genetic diversity and from seed banks for traits they consider useful, and then find ways to introgress the desirable traits into the varieties they are developing.

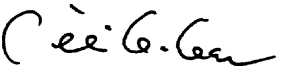
In recent years, genomics, the modern science of genetics, has been providing breeders with new tools and novel approaches to perform their tasks with high precision and efficiency. For example, applications of molecular markers in breeding through marker-assisted selection (MAS) have already been demonstrated in several crop species to develop improved varieties with better agronomic traits and enhanced resistance or tolerance to biotic or abiotic stresses. Indeed, MAS is being used widely both in developed and developing countries, and is enabling breeders to make use of unconventional plant materials.

Tremendous progress has been made in genome science in recent years. For example, the complete genomes for several plant species e.g., rice and sorghum, have already been sequenced, and similar efforts are underway for many other crop

species. Comparative and functional genomics approaches are helping scientists to better understand gene functions and to more effectively tailor the desirable genotype so that improved varieties can be released in a more timely fashion to farmers. If these new varieties prove to be appropriate for resource-poor farmers, then farmers could use them to enhance their livelihood security. This is probably the most obvious way to demonstrate how genomics empowers poor farmers.

This volume deals with applications of genomics towards crop improvement. I am glad to note the variety of opinions and experiences that the editors of this volume have gathered - eminent scientists representing international agriculture centers, advanced research institutes and national agricultural research systems from several countries - providing a treasure trove of information.

I am sure the book will be useful for the community involved in applying genomics research for crop improvement as well as for teachers and students to enhance their knowledge in the latest tools and approaches to genomics.



William D. Dar
Director General
International Crops Research Institute for the Semi-Arid Tropics (ICRISAT)
Patancheru, Andhra Pradesh
India

PREFACE

Genomics, dealing with the collection and characterization of genes and analysis of the relationships between gene activity and cell function, is a rapidly evolving, interdisciplinary field of study aimed at understanding and exploiting the biological information encoded in DNA. The genomics toolbox includes automated genetic and physical mapping, DNA sequencing, bioinformatics software and databases, transcriptomics, proteomics, metabolomics, and high-throughput profiling approaches. Indeed, the past two decades have witnessed spectacular advances in genomics. For example, at the dawn of the genomics era, Arabidopsis was chosen as the first model genome for sequencing, which was then quickly followed by the sequencing of other model genomes (rice for monocots, Medicago and Lotus for legume crops and poplar for tree species) and crop species (soybean, cassava, sorghum, etc.). While new crops (e.g. maize, wheat, finger millet, etc.) are being added to the list for sequencing the genome or gene space, the generated sequence data are being analyzed in parallel for characterizing the genes and validating their functions through comparative and functional genomics approaches including bioinformatics, transcriptomics, and genetical genomics. Candidate genes are becoming increasingly useful for the development of markers for assaying and understanding functional diversity, association studies, allele mining, and most importantly, marker-assisted selection. Therefore, genomics research has great potential to revolutionize the discipline of plant breeding in order to face the challenges posed by feeding an ever-growing human population expected to top 10 billion by 2050, while decreasing the environmental footprint of agriculture and preserving the remaining biodiversity.

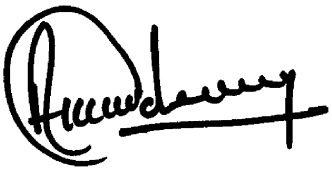
Several high-throughput approaches, genomics platforms, and strategies are currently available for applying genomics to crop breeding. However, the high costs invested in, and associated with, genomics research currently limit the implementation of genomics-assisted crop improvement, especially for autogamous and/or minor and orphan crops. This book presents a number of articles illustrating different contributions which genomics can offer to unravel the path from genes to phenotypes and vice versa, and how this knowledge can help to improve crops' performance and reduce the impact of agriculture on the environment. Each article shows how structural and/or functional genomics can improve our capacity to unveil and deploy natural and artificial allelic variation for the benefit of plant breeders. Volume 1, entitled "Genomics Approaches and Platforms", presents state-of-the-art genomic

resources and platforms and also describes the strategies and approaches for effectively exploiting genomics research for crop improvement. Volume 2, entitled "Genomics Applications in Crops", presents a number of case studies of important crop and plant species that summarize both the achievements and limitations of genomics research for crop improvement.

More than 90 authors, representing 16 countries from five continents have contributed 16 chapters for Volume I and 18 chapters for Volume II (see Appendix I). The editors are grateful to all the authors, who not only provided a timely review of the published research work in their area of expertise but also shared their unpublished results to offer an updated view. We also appreciate their cooperation in meeting the deadlines, revising the manuscripts, and in checking the galley proofs. While editing this book, we received strong support from many reviewers (see Appendix II) who provided useful suggestions for improving the manuscripts. We would like to thank our colleagues and research scholars, especially Yogendra, Rachit, Mahender, Priti, and Spurthi working at ICRISAT for their help in various ways. Nevertheless, we take responsibility for any errors that might have crept in inadvertently during the editorial work.

The cooperation and encouragement received from Jacco Flipsen and Noeline Gibson of Springer during various stages of the development and completion of this project, together with the help of Rajeshwari Pal of Integra Software Services for typesetting and correcting the galley proofs, have been instrumental for the completion of this book and are gratefully acknowledged. We also recognize that our editorial work took away precious time that we should have spent with our respective families. RKV acknowledges the help and support of his wife, Monika and son, Prakhar (Kutkut) who allowed their time to be taken away to fulfill RKV's editorial responsibilities in addition to research and other administrative duties at ICRISAT. Similarly, RT is grateful to his wife Kay for her precious help in editing and proof-reading a number of manuscripts.

We hope that our efforts will help those working in crop genomics as well as conventional plant breeding to better focus their research plans for crop improvement programs. The book will also help graduate students and teachers to develop a better understanding of this fundamental aspect of modern plant science research. Finally, we would appreciate receiving readers' feedback on the errors and omissions, if any, as well as their suggestions, so that a future revised and updated edition, if planned, may prove more useful.



Rajeev K. Varshney

Roberto Tuberosa

COLOR PLATES

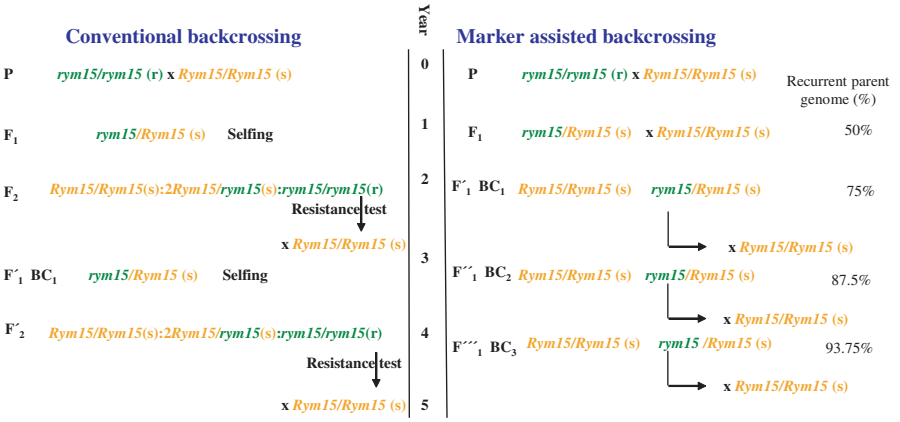


Plate 1. Comparison of conventional and marker-assisted backcrossing programmes for the incorporation of *rym15* (from Ordon et al. 1999, mod.) *(See Fig. 1, on page 88)*

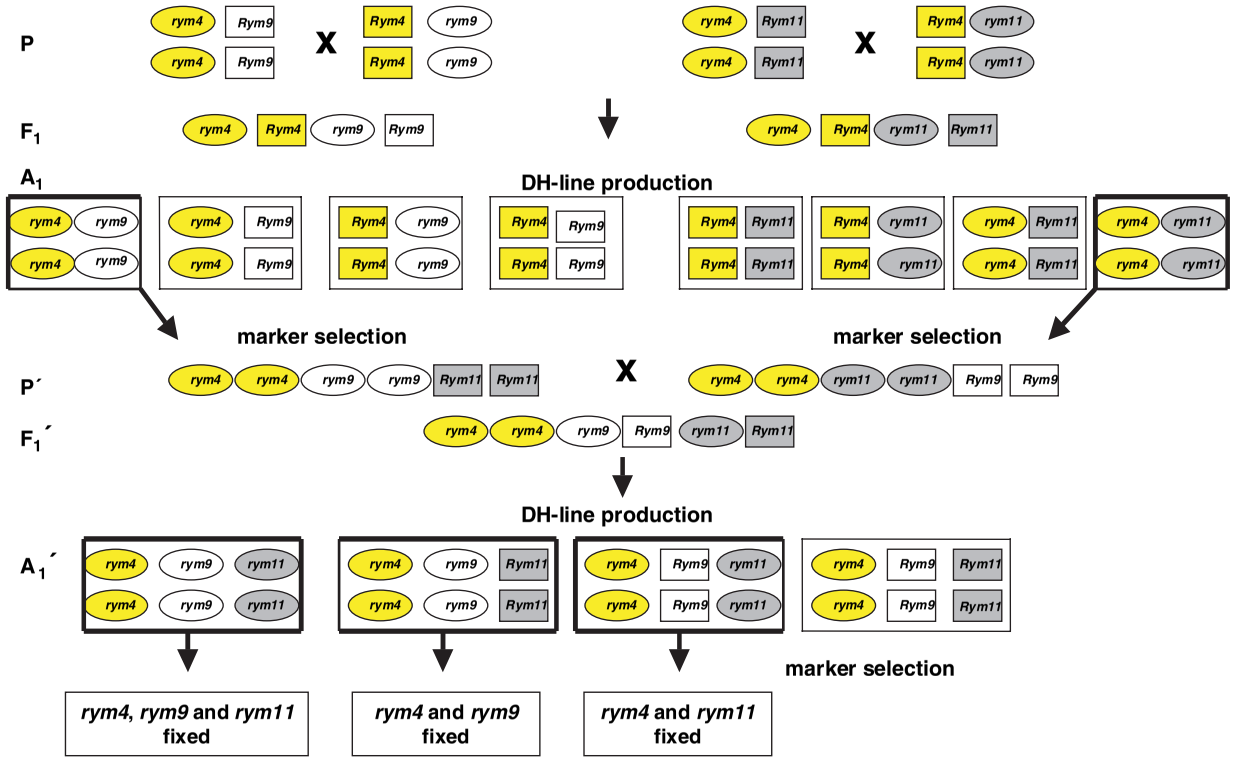


Plate 2. Scheme of pyramiding resistance genes *rym4*, *rym9* and *rym11* by two haploidy steps (Werner et al. 2005) *(See Fig. 2, on page 91)*

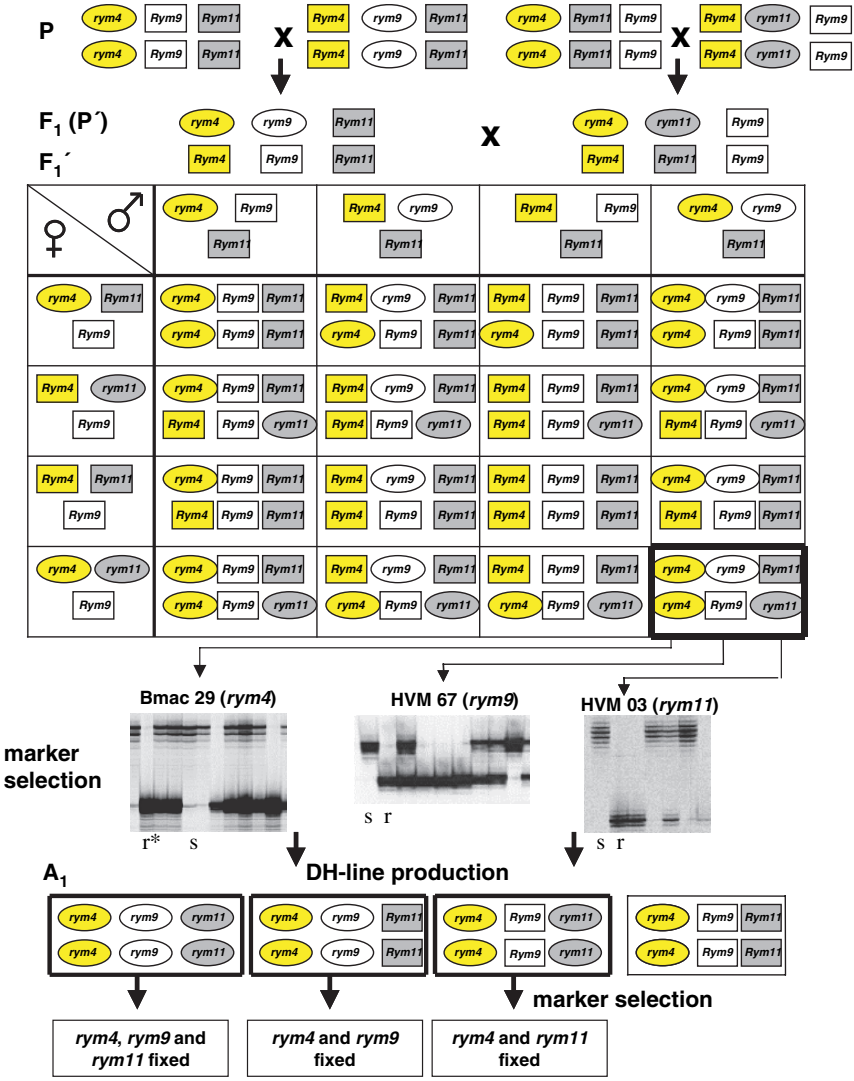


Plate 3. Scheme of pyramiding resistance genes *rym4*, *rym9* and *rym11* by one haploidy step (Werner et al. 2005) *(See Fig. 3, on page 92)*

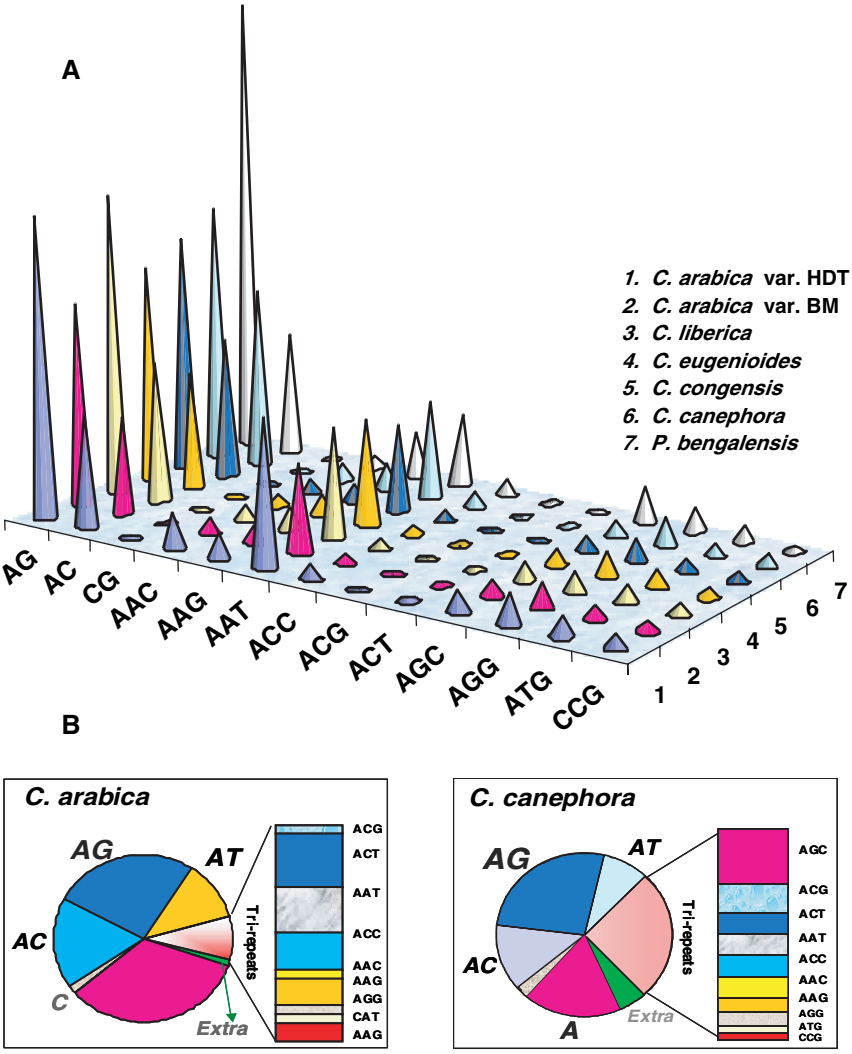


Plate 4. Relative abundance of different SSR motifs in genomes of different coffee species as revealed by: A) semi-quantitative Southern hybridization based slot-blot analysis, and B) *in-silico* sequence analysis of >1000 SSR positive clones from *C. arabica* and *C. canephora* (robusta) small-insert genomic libraries. Note almost similar comparable pattern/relative frequencies of different SSRs across coffee species and also between the two approaches of evaluation (our unpublished data) *(See Fig. 1, on page 405)*

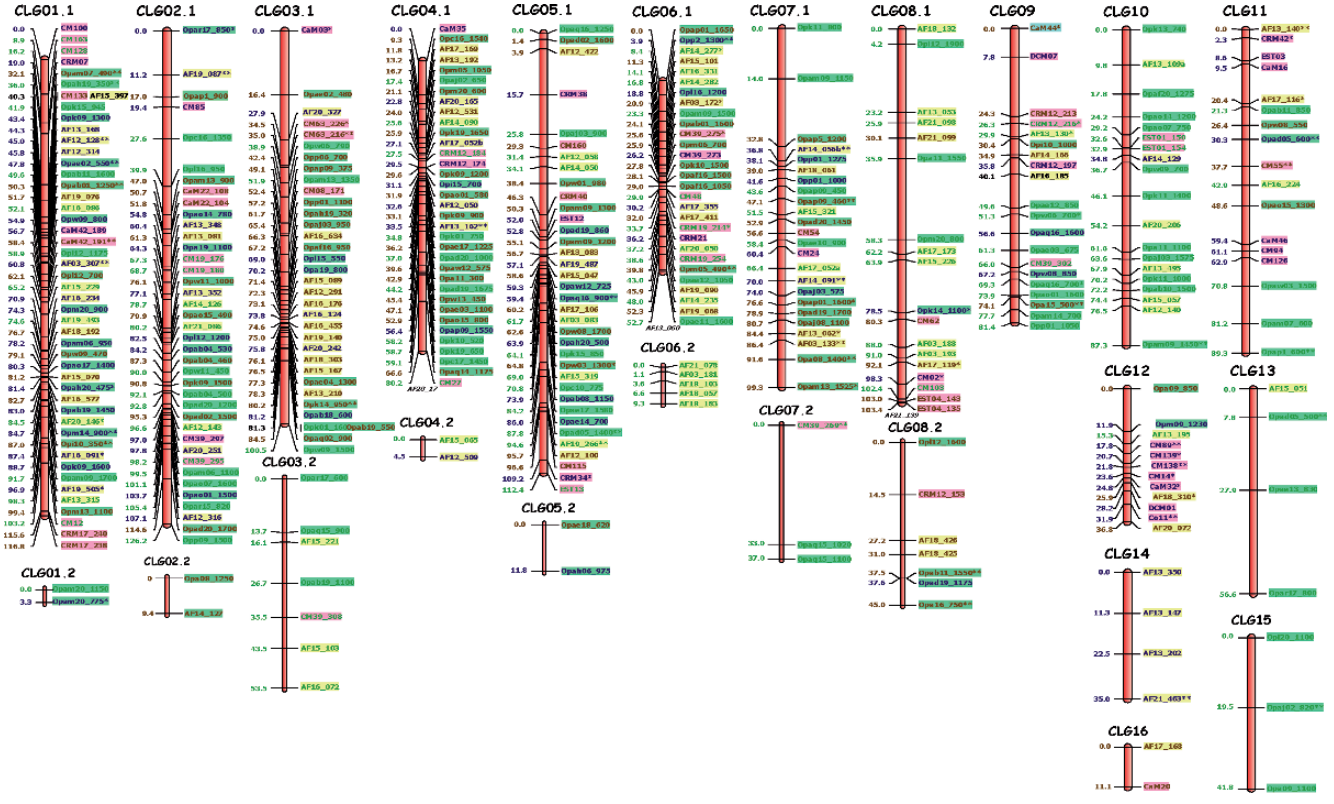


Plate 5. An integrated framework linkage map of robusta coffee (*C. canephora*) developed using a *pseudo-testcross* strategy and DNA markers (AFLP, RAPD, SSRs). The map comprises 11 major and 5 minor linkage groups, and has a total of 374 mapped markers. Prefixes Op-, AF- in the marker names indicates RAPD, AFLP respectively, while CM-/CR-/CaM- represents the mapped SSRs. The small groups shown below some of the major CLGs, represent the floating markers of the respective CLG placed in an 'Accessory group/map" (unpublished, Hendre 2006) *(See Fig. 2, on page 413)*

CHAPTER 1

MICROSATELLITE AND SNP MARKERS IN WHEAT BREEDING

MARTIN W. GANAL[1,*] AND MARION S. RÖDER[2]

[1]*TraitGenetics GmbH, Am Schwabeplan 1b, D-06466 Gatersleben, Germany*
[2]*Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Corrensstr. 3, D-06466 Gatersleben, Germany*

Abstract: Bread wheat (*Triticum aestivum* L.) is one of the most important crop plants. Due to its hexaploid nature consisting of three different genomes (A, B and D) and its large genome size of approximately 15 billion base pairs, it is also one of the most complex crop genomes. This has rendered the use of molecular markers in wheat genome analysis and breeding slow and difficult. Mainly, through the use of microsatellite or SSR (simple sequence repeat) markers, wheat molecular marker analysis has gained momentum during the last ten years. The advantage of microsatellite markers is that they detect an unsurpassed level of polymorphism in this recently polyploidised organism with a generally low level of sequence variation. Furthermore, a large proportion of the microsatellite markers is genome-specific, thus amplifying a defined single product from one of the three wheat genomes. Currently, 2.000 to 2.500 mapped microsatellite markers are available for the wheat genome. With microsatellite markers, the chromosomal position of many relevant breeding traits such as disease resistance genes and quality traits has been identified and they are increasingly used in marker-assisted selection during wheat breeding. For the future, high expectations are being put into another marker type that is called single nucleotide polymorphisms (SNPs) since their number in the wheat genome should be much higher and cost-efficient, highly multiplexed technologies are available for the analysis of SNP markers in plants. SNP marker development and use are, however, still in their infancy. Based on recent results, we discuss here the advantages and disadvantages of SNPs compared to microsatellite markers for future wheat breeding.

1. INTRODUCTION

Hexaploid bread wheat ($n = 21$) consists of three genomes with a basic chromosome number of $n = 7$ for each genome. The three genomes are derived from *Triticum urartu* (A-genome), an unknown species from the *Sitopsis* section (B-genome)

*Author for correspondence: ganal@traitgenetics.de

R.K. Varshney and R. Tuberosa (eds.), Genomics Assisted Crop Improvement: Vol. 2: Genomics Applications in Crops, 1–24.

and *Triticum tauschii* (D-genome) respectively (Feldman and Levy, 2005). The hexaploid bread wheat genome is one of the largest crop genomes with, all three genomes combined, approximately 15 billion base pairs (Arumuganathan and Earle, 1991) and each gene usually occurring in at least three copies. It is estimated that each of the three wheat genomes contains approximately 40.000 to 50.000 genes. More than 80% of the DNA of bread wheat consists of repeated DNA sequences with transposons and retrotransposons representing the highest proportion (Devos et al., 2005). Intensive genetic analyses showed that synteny perturbations between the three genomes are correlated with recombination rates along chromosomes (Akhunov et al., 2003a,b). Physical mapping through the use of deletion lines led to the conclusion that the genes are not evenly distributed along the chromosomes but are predominantly located in telomeric regions (Qi et al., 2004). However, first data on the DNA level obtained through the comparative sequencing of large DNA regions have shown that the three genomes differ considerably at their DNA sequence level predominantly caused by the presence/absence of retrotransposons and other repetitive sequences but also through the shuffling of genes within the individual genomes (Wicker et al., 2001; Appels et al., 2003; Keller et al., 2005).

Over more than fifteen years, molecular markers have been used in wheat genome analysis. Initially, with the RFLP technology, first genetic maps of wheat have been constructed for all 21 chromosomes. The advantage of the RFLP technology in wheat genome analysis includes usually the detection of one copy of the respective locus in each of the three wheat genomes simultaneously so that very large numbers of loci can be assigned to the wheat chromosomes or specific chromosomal regions through the use of deletion lines (Qi et al., 2004). Another advantage of the RFLP analysis is that through the hybridization of probes from closely related other Gramineae, it was possible to deduce that the wheat genomes and the genomes of other Gramineae, such as rice, maize, rye and barley share a very similar order of genes called synteny on the chromosomal level (Van Denze et al., 1995; Gale and Devos, 1998). RFLP analysis has however, only in a few cases been used for segregation analysis in wheat breeding except for the mapping of genes in wide crosses. This is mainly due to the fact, that the level of RFLP polymorphism is very low in wheat breeding material. Other PCR-based marker analyses such as multilocus systems, for example, AFLP have also not been widely used in wheat since this system also identifies only a limited level of genetic polymorphism in breeding material and routine analysis is being further complicated through the large genome size and the amount of repetitive sequences (Manifesto et al., 2001; Zhou et al., 2002).

2. MICROSATELLITE MARKERS

Only through the advent of microsatellite or SSR markers, the use of molecular markers has gained momentum in the last ten years since the first SSR markers have been described for hexaploid wheat (Devos et al., 1995; Röder et al., 1995; Bryan et al., 1997). Microsatellite markers are based on tandemly repeated DNA

sequences of short repetitive motives (e.g. poly CA, poly CT, poly AT and other repeated sequences of 3–5 bases). The variability of microsatellite sequences in a genome is not based on point mutations but on the variation in the number of the simple sequence repeats. Such variation occurs approximately ten times more frequently through processes such as slippage during replication or unequal crossing over (Hancock, 1999). This makes SSRs the most suitable and polymorphic marker system in species with a low level of polymorphism such as wheat.

2.1. Wheat Microsatellite Markers

The first large set of microsatellite markers for the wheat genome has been published in 1998 (Röder et al., 1998b). With these markers it could be shown that microsatellite markers have a number of characteristics that make them the currently best-suited marker system for the analysis of hexaploid wheat. Wheat microsatellite markers detect a much higher level of variability than RFLPs and other marker types (e.g. AFLPs), especially in closely related wheat germplasm and varieties such as they are used for breeding. Approximately 50% of the wheat SSR markers detect only a specific locus on one of the three genomes and thus are genome-specific. If they amplify from more than one of the three wheat genomes, the amplified fragments are often clearly distinguishable on high resolution gels. SSR markers are multiallelic and detect up to more than 30 different alleles in the wheat germplasm for a given locus which makes them in their information content significantly superior to biallelic marker systems (Plaschke et al., 1995; Röder et al., 2002). The distribution of microsatellite markers along the wheat chromosomes does not significantly differ from that genes, which show partial clustering at the physical end of the chromosomes, although most of the currently used SSR markers are not in genes (Röder et al., 1998a). Finally, wheat microsatellite markers are amenable to high-throughput so that large numbers of markers can be analyzed with large numbers of plants. This make them the preferred marker system for marker assisted selection and mapping in wheat and for breeding (Koebner et al., 2001; Koebner and Summers, 2003).

Over the last eight years, large sets of wheat microsatellite markers have been developed from various sources. At Gatersleben, most of the approximately 1.000 identified SSR markers are derived from genomic clones generated from single-copy sequences out of hypomethylated regions (Röder et al., 1998b; Pestsova et al., 2000; unpublished results). Other microsatellite markers, such as the wmc markers, are derived from libraries of wheat sequences that were generated through a variety of enrichment procedures (Song et al., 2002; Somers et al., 2004). Recently, a considerable number of wheat microsatellite markers have been identified from microsatellite motives identified within EST sequences through bioinformatic data mining (Eujayl et al., 2002; Holton et al., 2002; Gao et al., 2004; Yu et al., 2004; Peng and Lapitan, 2005). In general, the developed wheat SSR markers are of different amplification quality and have a varying number of detected loci. Nevertheless, it becomes now clear that microsatellite markers generated

from microsatellite motives within ESTs are usually more highly conserved and can be used through a wider range of germplasm (Zhang et al., 2006) but on the other side are also significantly less polymorphic than markers derived from genomic sequences since EST-SSR contain on average less repeating units of the microsatellite motif (Varshney et al., 2005).

Currently, the mapping of approximately 2.000 to 2.500 SSR markers on the 21 wheat chromosomes has been published in a variety of publications (Table 1). The precise number of mapped SSR markers in wheat is at present not easily determinable since the published maps show considerable overlap in the used SSR markers and frequently in different mapping populations different loci on the three homeologous chromosomes have been identified. Furthermore, in a number of cases only the number of detected loci is specifically described. Finally, no comprehensive sequence comparison between the sequences used for marker generation and no mapping data integration has been performed for all markers so that it is possible that a considerable number of microsatellites markers that detect the same sequence has been developed independently in different laboratories, especially in the case of EST-SSRs mined from public databases. Figure 1 shows a map generated with more than 1.000 unique SSR markers generating 1169 mapped microsatellite loci which was established in collaboration between the IPK and TraitGenetics representing currently the genetic map with the largest number of unique wheat SSR markers.

In wheat breeding, SSRs are increasingly being used as the marker backbone for a variety of purposes. These include the localization of individual genes onto the 21 wheat chromosomes such as, for example, disease resistance genes or genes affecting other agriculturally important traits. A large variety of papers has been summarized in a review (Röder et al., 2004) and a website (http://maswheat.ucdavis.edu). Furthermore, wheat microsatellite markers have been used for the localization of a large set of QTLs (quantitative trait loci) affecting morphological and agronomically important traits (e.g. Perretant et al., 2000; Börner et al., 2002; Huang et al., 2003; 2004). Other examples for important mapped QTLs are loci affecting resistance against the scab disease caused by *Fusarium* for which several populations have been used (Anderson et al., 2001; Buerstmayr et al., 2002; del Blanco et al., 2003; Bai and Shaner, 2004). For these QTLs, currently large SSR marker assisted selection projects are in progress. Further applications for which wheat microsatellite markers have been used include the characterization of large numbers of wheat varieties grown in Europe and North America, as well as the characterization of wheat lines from germplasm collections for the determination of genetic diversity over time (Donini et al., 2000; Christiansen et al., 2002; Huang et al., 2002; Pestsova and Röder, 2002; Röder et al., 2002). Many of these aspects have recently been published in a number of reviews (Koebner et al., 2001, Koebner and Summers, 2003; Röder et al., 2004). Finally, it is very likely, that wheat microsatellite markers will in the future also be used as backbone for the anchoring of a physical map of the wheat genome onto the genetic map and within the map-based isolation procedure for genes from the wheat genome, as it has been described already in a few cases (Stein et al., 2000; Keller et al., 2005).

Table 1. Published SSR maps in wheat with the number of mapped loci

Lab-designator of mapped SSRs (Number of mapped SSR loci)	Mapping population	Remarks	Reference
gwm (279)	Opata 85 × W7984 (ITMI-population)	SSRs isolated from genomic libraries of A, B and D genomes	Röder et al. (1998)
psp (53)	CS × synthetic population	SSRs isolated from genomic libraries of A, B and D genomes	Stephenson et al. (1998)
gdm (46)	ITMI-population	SSRs isolated from genome D (*Ae. tauschii*)	Pestsova et al. (2000)
wmc (66)	ITMI-population	SSRs isolated from enriched genomic libraries of A, B and D genomes	Gupta et al. (2002)
DupW (22)	ITMI-population	SSRs isolated from ESTs	Eujayl et al. (2002)
cfd (84)	Courtot × CS population	SSRs isolated from genome D (*Ae. tauschii*)	Guyomarc'h et al. (2002)
cnl, ksum (149)	ITMI-population, physical mapping	SSRs isolated from ESTs	Yu et al. (2004)

(Continued)

Table 1. (Continued)

Lab-designator of mapped SSRs (Number of mapped SSR loci)	Mapping population	Remarks	Reference
wmc, gwm, gdm, cfa, cfd, barc (1235)	Consensus map	SSRs from various sources	Somers et al. (2004)
cfa, cfd, gpw, ksu, gwm, barc (725)	Physical mapping on nulli-tetrasomic and deletion lines	SSRs from various sources	Sourdille et al. (2004)
xx-SSR (101)	ITMI-population and WSpop and LHpop	SSRs isolated from ESTs	Gao et al. (2004)
cwem (48 as PCR-based markers of total of 266 eSSR loci)	Physical mapping on nulli-tetrasomic and deletion lines	SSRs isolated from ESTs	Peng and Lapitan (2005)
barc (347)	ITMI-population, physical mapping	SSRs isolated from sheared genomic libraries of A, B and D genomes	Song et al. (2005)
hbg, hbe, hbd (250)	Kitamoe × Münstertaler, intraspecific DH-population	SSRs isolated from genomic libraries and ESTs	Torada et al. (2006)

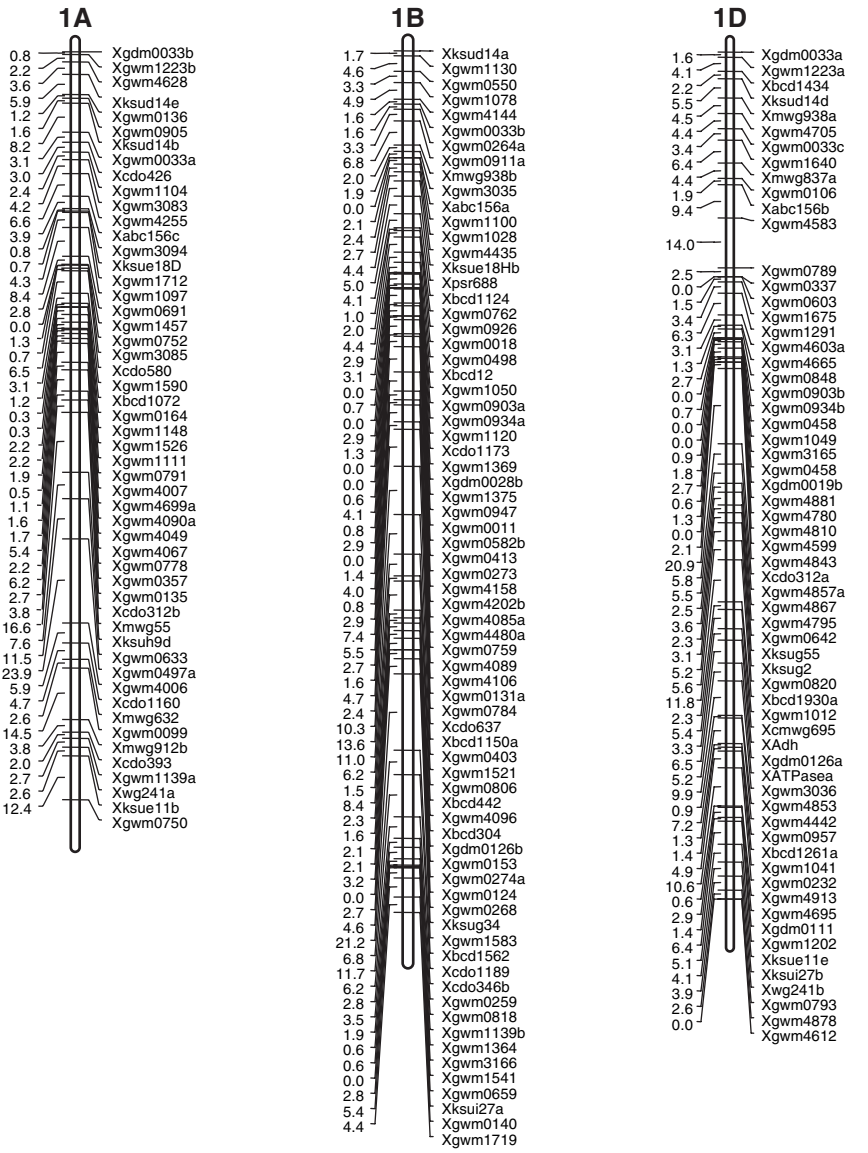


Figure 1. Molecular linkage map of wheat based on 70 recombinant inbred lines derived of the cross Opata × W-7984, the so-called ITMI-population (International Triticeae Mapping Initiative). The microsatellite loci carrying the lab designators "gwm" (Gatersleben wheat microsatellite) or "gdm" (Gatersleben D-genome microsatellite) are placed in a framework of previously mapped RFLP markers

In summary, wheat SSR will still be the tool of choice for wheat genome analysis and wheat breeding over the next several years, especially if the number of useful markers can be increased. Plant breeding companies are increasingly using the

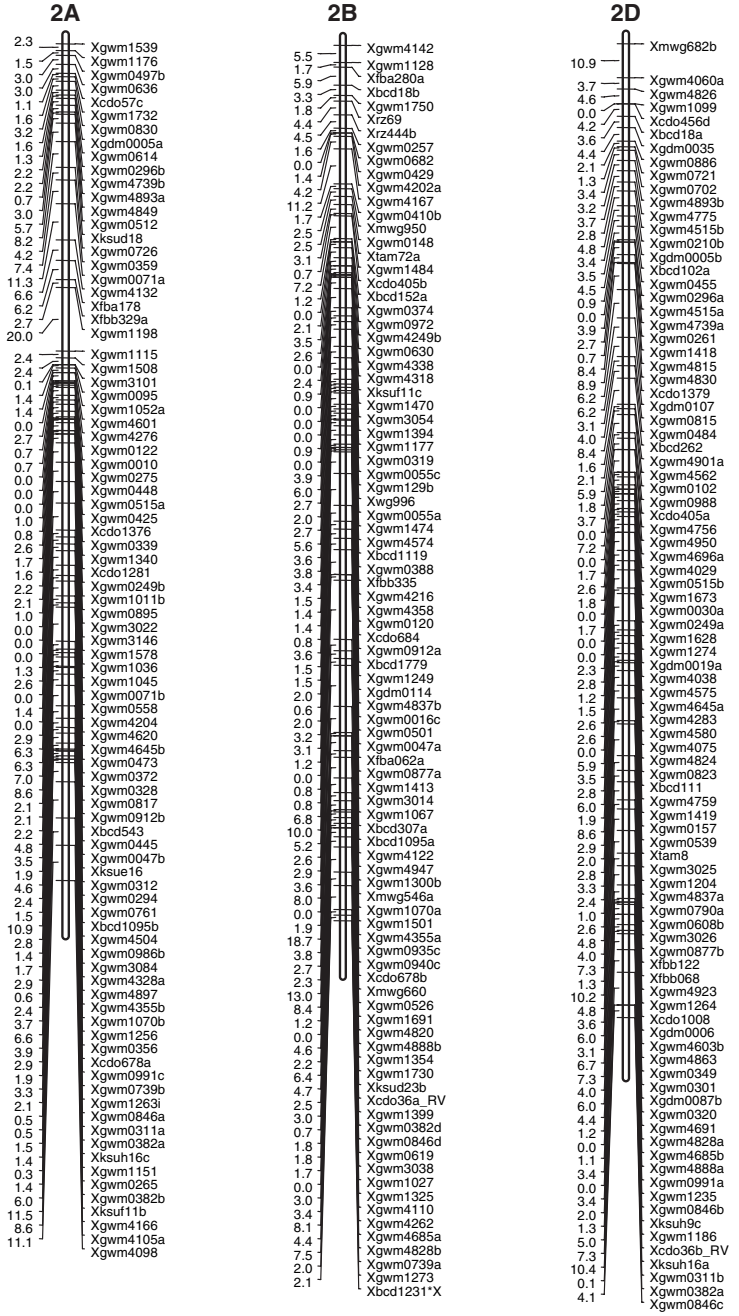


Figure 1. (Continued)

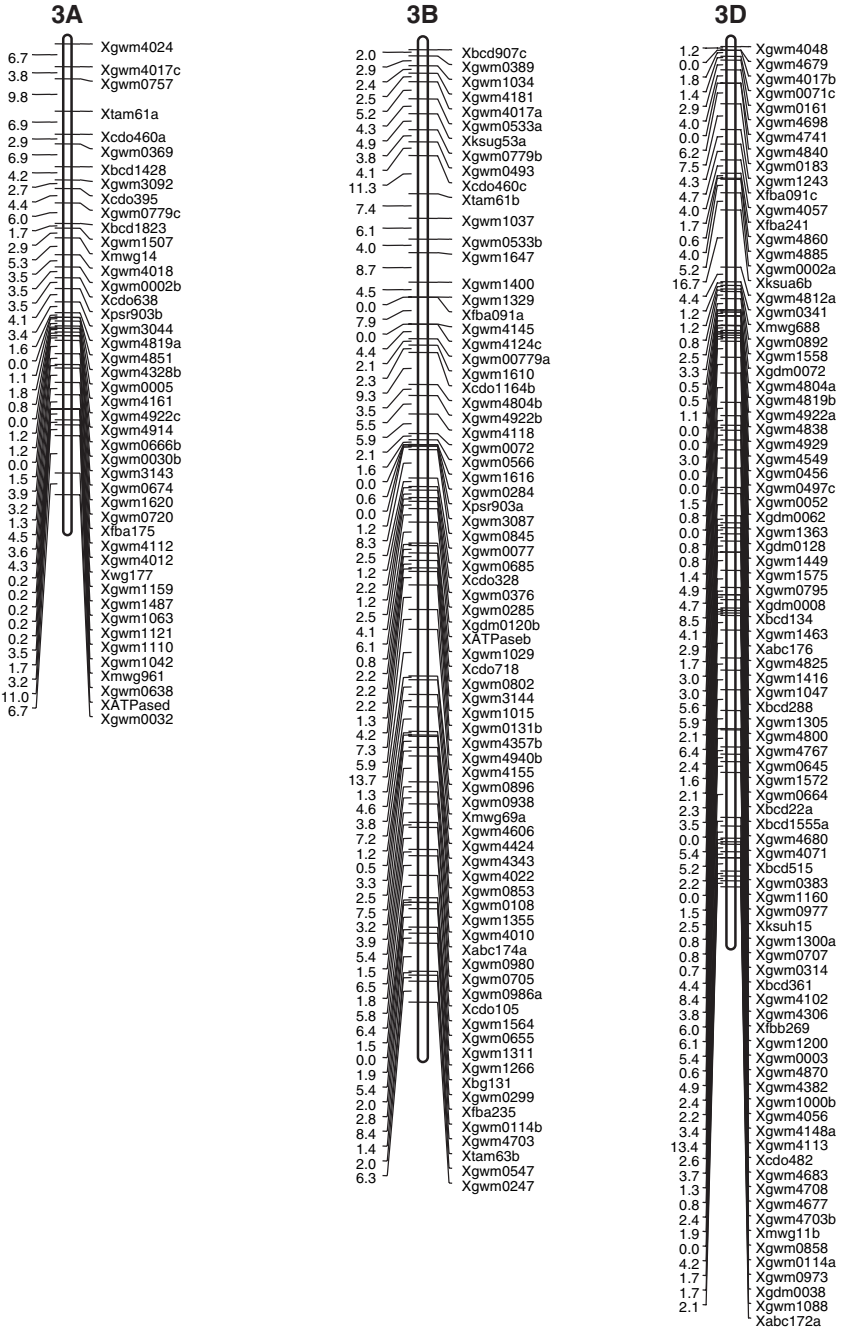


Figure 1. (Continued)

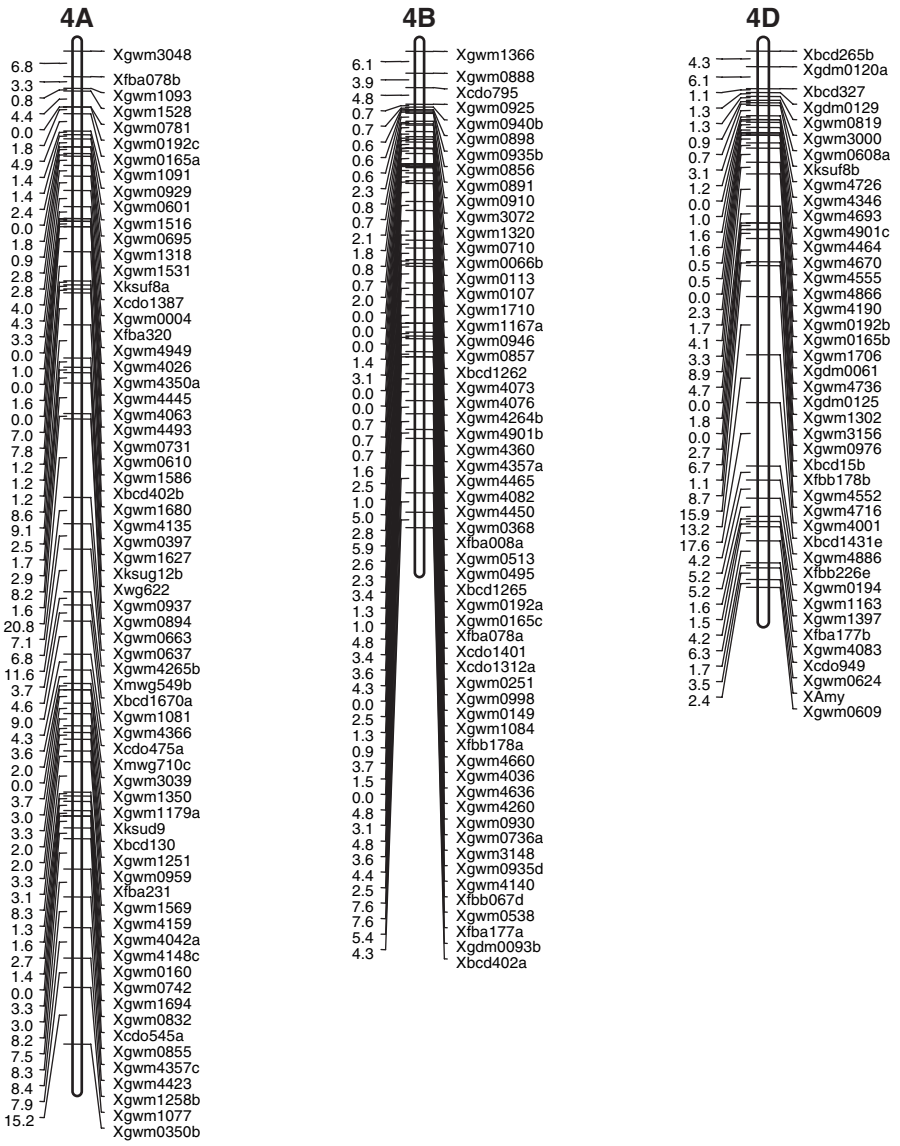


Figure 1. (Continued)

advantages of wheat microsatellite markers during marker-assisted selection and backcrossing of important traits into elite material and the development of new varieties (Koebner and Summers, 2003; Powell and Langridge, 2004). Specific examples for such applications are also described in other chapters of this book.

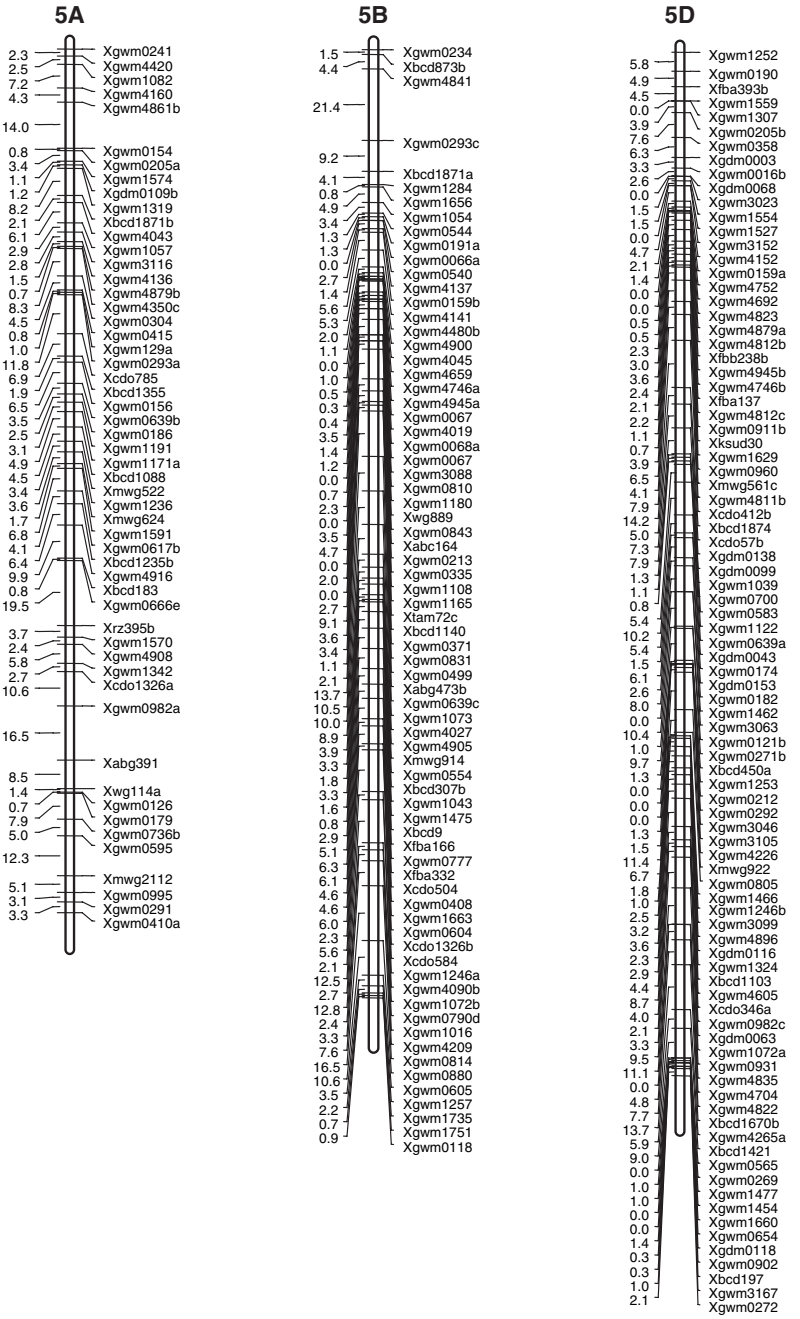


Figure 1. (Continued)

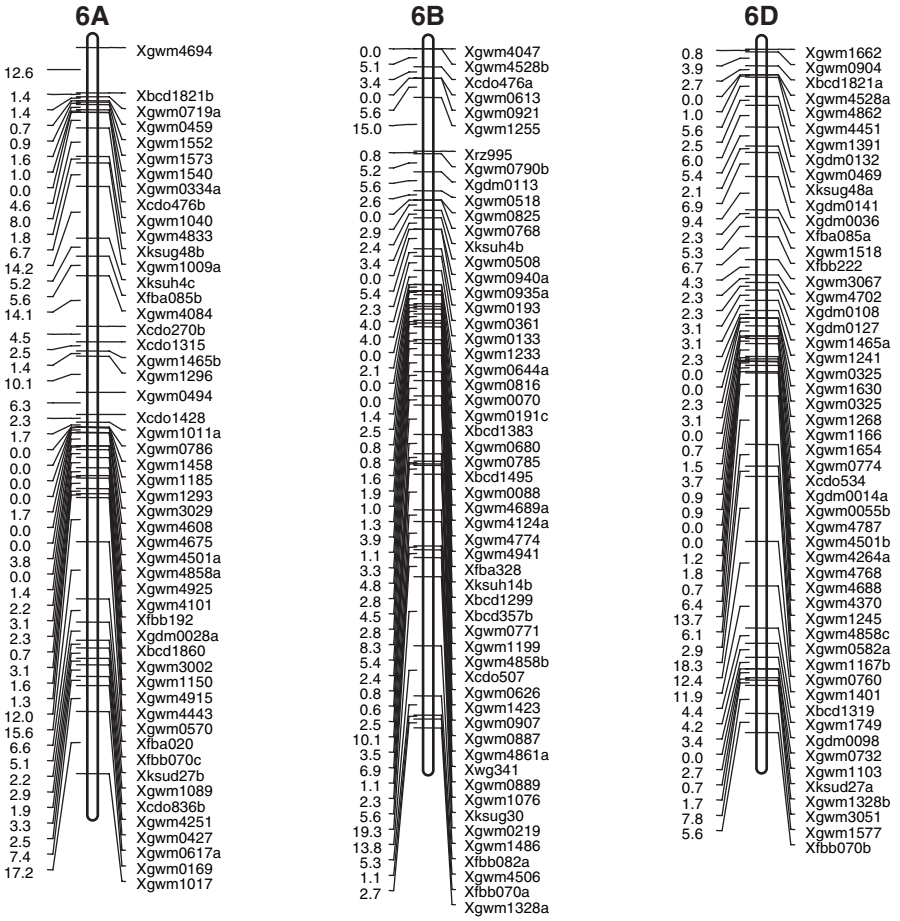


Figure 1. (Continued)

3. SINGLE NUCLEOTIDE POLYMORPHISMS OR SNP MARKERS

SNP markers are based on the variation of a specific nucleotide at a given sequence position between individuals. Predominantly, such variation occurs as biallelic alternative bases or as insertion/deletions of individual or small numbers of nucleotides. SNPs have in the last years gained considerable interest due to the fact that they are the smallest unit of genetic variation and being the basis of most genetic variation between individuals, they occur in virtually unlimited numbers. SNPs in coding sequences create furthermore the possibility of changes in the amino acid sequence within a protein (if they are not silent) and might have an effect on protein function and thus monogenic or polygenic traits associated with the expression of such genes (Johnson et al., 2001). SNP analysis has been spearheaded in human genome

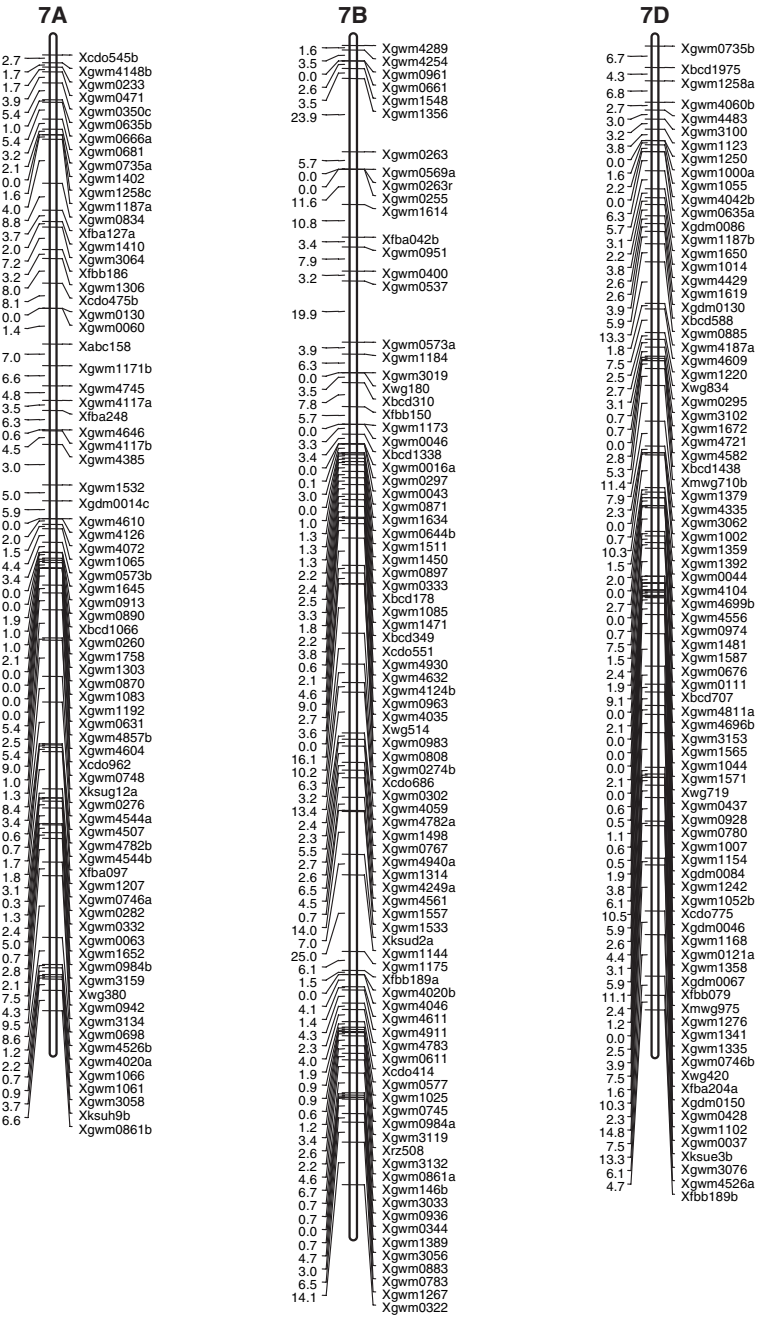


Figure 1. (Continued)

analysis, where meanwhile more than 2 million SNPs have been identified which represent in their entirety the majority of the genetic variation within the single copy and expressed part of the human genome (Sachidanandam et al., 2001).

The interest in SNP markers has further accelerated through the development of cost-efficient, high-throughput multiplexing (many markers being analyzed simultaneously) analysis techniques based on chips or other array techniques. Up to several million SNPs can now be analyzed simultaneously in the human genome with costs of a few cents per individual locus. Furthermore, individual SNPs can also be analyzed through a variety of technologies that permit the determination of individual genotypes in an unsurpassed speed and accuracy such as, for example through the use of fluorochrome-based analysis technologies that create data in real-time (e.g. Taqman, Ampliflour, Invader) and/or in a quantitative fashion (Gut 2001, 2004; Kwok, 2001).

Further hopes regarding the use of SNPs in routine analyses have been raised through the identification of haplotypes. Haplotypes are closely linked SNPs which occur along a chromosome in clearly defined structures or patterns (alleles) that extend over hundreds of base pairs or even several kilobases. The haplotype structure of SNPs alleviates the problem of scoring an extremely large number of SNPs by not requiring the analysis of each individual SNP in a genome but only of a limited number for genome coverage. Considerable efforts have been put into the identification of the haplotype pattern of the human genome (http://hapmap.org) but at present it is still not clear how efficient the analysis of haplotyopes is for the identification of quantitative traits (Johnson et al., 2001; Foster and Sharp, 2004).

In plants, large scale SNP development and analysis project have been performed predominantly in diploid crop plants such as maize (Ching et al., 2002), barley and soybean (Zhu et al., 2003) where meanwhile more than 1.000 genes with SNPs were identified each. These data have demonstrated that SNPs are present in large numbers in crop plants and that they share similar features (e.g. presence as haplotypes) as in other eukaryotic species (Rafalski 2002a,b; Kahl et al., 2005).

3.1. SNP Identification in Wheat

SNP markers are usually identified through the comparative sequencing of individual lines or varieties or the bioinformatics analysis of EST data generated from a variety of lines (Rafalski, 2002a,b). With the advent of complete genome sequences for a number of plants such as in the model organisms *Arabidopsis thaliana* or rice (*Oryza sativa*), SNPs can be readily identified in basically unlimited numbers in single-copy DNA once such information is available.

Due to its large genome sequence, complete genome sequencing of the wheat genome is still at least five years and probably more years away, so that these resources will not be available in the near future. Furthermore, SNP identification through comparative sequencing is not easily possible in the hexaploid wheat genome. The main reason is the fact that with a normal primer pair usually the locus is being simultaneously amplified from the three different genomes so that even

when only a single fragment is identified after PCR amplification, the sequencing of that product is composed of three different sequences which make the analysis of the data at least extremely difficult. In many cases, the sequence is impossible to determine since one of the three genomes harbours insertions/deletions pushing the sequence out of frame for that genome. With these problems, direct sequencing is not feasible on a routine basis in wheat and other more complex approaches for SNP identification have to be used.

The most direct way of identifying SNPs in the wheat genome is the bioinformatic mining of wheat ESTs that are available in the respective EST databases. Approximately 500.000 ESTs from the wheat genome have been deposited in the EST databases making the wheat genome one of the best-sampled plant genomes (e.g. Lazo et al., 2004). Through concerted efforts of the international wheat community, these ESTs have been sampled from a variety of lines or cultivars so that SNPs between these lines could be identified. One of the pitfalls of EST sequence data is that the sequence quality is usually not better than 99% or a Phred-score of 20 meaning that on average 1 base out of 100 is incorrect. Thus, comparing individual sequences to each other for SNP identification is not easily possible and further complicated by the presence of highly related sequences from the other two genomes. This problem can be circumvented by the clustering of all sequence data generated from a wheat gene. EST clusters with a sufficiently deep sequence cover (several copies of ESTs from each of the three genomes) permit the identification of a consensus sequence from each of the three genomes and these sequences can be compared to similar clusters for the same gene derived from other wheat accessions. Since the sequences of the three genomes can be readily discriminated in such a comparison due to their higher sequence variation, SNPs in the individual copies derived from the three (A, B and D) genomes can be identified in a quite reliable fashion. Several years ago, an initiative has been started to analyze sequence clusters of more than 1.000 wheat genes in that way. However, until now beside the results of a pilot study, no detailed data regarding the outcome of that study and the level of polymorphism have been published probably due to the complexity of such an approach (Somers et al., 2003).

Another approach towards the identification of SNPs is the reduction of the hexaploid wheat genome into a diploid by the use of genome-specific primers. The average sequence difference between genes in the three individual wheat genomes is in the range of a few percent. Genome-specific primers permit the amplification of PCR products from only one specific wheat chromosome and thus make comparative sequencing of PCR products from individual lines feasible. Furthermore, through its inbreeding nature, such sequences from individual genomes are readily analyzable since only one allele should be amplified. Genome-specific primers can be generated from clustered EST data through the generation of primers with genome-specific 3'-ends. Since the coding sequences of wheat genes are highly conserved, frequently one of the two genome-specific primers is being derived from the less-conserved 3'-non-coding end of the respective gene. The pitfall of this approach is however, that not all potentially genome-specific primers do in

fact amplify a genome-specific product. Usually, several primer pairs have to be tested for obtaining one good genome-specific PCR product, thus making the approach quite laborious through the requirement of large-scale bioinformatic data processing and primer testing with a yield of usually less than 50% of functional genome-specific primers. One advantage of wheat in this approach is that through the use of nullitetrasomic lines, it is quite easy to confirm the genome-specific nature of an amplification product and simultaneously assign this product to a specific wheat chromosome. Currently, this approach is being used by the wheat HapMap-project (http://wheat.pw.usda.gov/SNP), where large numbers of potentially genome-specific primers have been generated and tested for SNP identification in hexaploid wheat and its diploid and tetraploid ancestors. At TraitGenetics, we have also used this approach in a pilot study towards the identification of SNPs in hexaploid bread wheat where we have generated more than 200 confirmed genome-specific primer pairs (unpublished results).

Further methods for the identification of SNPs in wheat are also used. For the analysis of specific candidate genes in a number of lines or varieties with respect to the occurrence of SNPs, it is possible to add a cloning step after the amplification of the respective PCR product from the three different genomes for each line and then determine the DNA sequence of a representative number of clones from each line. This approach creates probably the most complete data set for a given gene without the need of generating genome-specific primers but it is not easily amenable to high-throughput SNP detection in large numbers of genes. Due to its low level of SNPs in wheat varieties that will be described below, approaches to identify wheat genes that contain SNPs within breeding germplasm will probably gain more interest in the future. Such approaches could be, for example, the identification of single feature polymorphisms (SFPs) through the use of chip-technologies by means of comparative analysis of wheat cultivars or the use of heteroduplex analyses techniques (e.g. nuclease treatment or denaturing HPLC) for a first screening (Martins-Lopes et al., 2001). Although these techniques will require sophisticated bioinformatics to discriminate polymorphisms between the three wheat genomes and SNPs between lines, they might in the future provide an important tool for the pre-selection of genes with useful SNPs in wheat, however, without alleviating the problem of generating genome-specific PCR products at a later time. Another way of identifying SNPs in genome-specific sequences has been the sequencing of amplified microsatellite-flanking regions. This approach has recently also enabled the identification of a number of SNPs adjacent to microsatellite sequences (Ablett et al., 2006).

3.2. SNP Polymorphism in Wheat

The level of DNA polymorphism in hexaploid bread wheat is quite low when looking at fragment length polymorphism via RFLP analysis. Since point mutations are also the basis of SNPs, it can be expected that the frequency of SNPs is also low in hexaploid bread wheat. First data on SNPs have demonstrated that the SNP

frequency in the hexaploid wheat germplasm is in the order of 1 SNP per 540 base pairs (Somers et al., 2003). With that, the SNP frequency in bread wheat is at least five times lower than in maize (Ching et al., 2002; Rafalski, 2002a). Comparable levels of SNP polymorphism are being found, for example, in tomato and soybean (Zhu et al., 2003) both being crops that have passed through severe bottlenecks during domestication and breeding. For wheat, a reason for a low level of SNPs is also that hexaploid bread wheat is a recently generated polyploid with less than 10.000 years of divergence since its generation. An SNP frequency of one SNP per 540 base pairs would mean that on average in each sequenced gene fragment from an EST approximately one SNP would be located. This is however not the case since in many cases SNPs are clustered in 3'-untranslated regions and introns, so that a considerable amount of sequenced amplicons does not contain SNPs. Furthermore, the currently published data are based on germplasm that has been collected from all over the world and also contains highly polymorphic lines such as Chinese Spring or synthetic wheat lines which have added diversity to the wheat germplasm through the recreation of hexaploid wheat through the artificial hybridization of *Triticum turgidum* with *Aegilops tauschii* (Caldwell et al., 2004).

For applications in plant breeding, the used molecular markers have to be polymorphic in the respective breeding germplasm. European, North American or Australian wheat germplasm does only contain a fraction of the entire genetic variation of hexaploid bread wheat. Thus, for the use of SNP markers, it is important to determine the actual level of sequence polymorphism in such breeding germplasm. Since SNPs are usually biallelic markers it is necessary for practical purposes that the identified SNPs do show a high allele frequency. The allele frequency of the minor allele for an SNP should be at least 0.2 or 20% in order to be useful in actual wheat marker analysis since SNP occurring in only one or a few lines are not generally useful. We have performed a pilot study regarding the level of SNP polymorphism in a number of wheat lines predominantly representing well characterized European germplasm through the comparative sequencing of 202 genome-specific amplicons. These data have demonstrated that only 75 (37%) of the sequenced genome-specific amplicons showed at least one SNPs in a set of 12 wheat lines of which 4 were non-European wheat lines. An analysis of the eight European wheat varieties that were selected to represent the range of the European wheat germplasm based on microsatellite data, demonstrated that approximately two thirds of the identified SNPs were also present in the European wheat breeding material. Many of the identified SNPs were not identified in only one line but in at least 2 lines indicating a relatively high allele frequency of these SNPs in the wheat germplasm. This is also substantiated by the fact that SNPs that were found in the four exotic wheat lines were quite frequently also present in the wheat varieties suggesting that with respect to SNPs at least the European wheat breeding material and varieties do cover a considerable amount of the nucleotide polymorphism that is found in the entire hexaploid wheat gene pool. The comparative sequencing approach has also shown that in case of more than 1 SNP occurring in an amplicon, these are usually present in the form of well defined haplotypes.

In most cases only two haplotypes were observed but in some cases up to four haplotypes were observed.

In summary, in only one out of 4 sequenced amplicons, one or more SNPs were identified in European wheat germplasm indicating that a large number of genome-specific amplicons need to be investigated in order to cover the wheat genome at a reasonable density. For example, if a density of on average one informative SNP per 10 centiMorgan interval between two representative European wheat lines should be needed, one would need to have access to a large number of sequenced gene fragments. With an estimated polymorphism level of 20% between two average wheat European wheat lines and a genome size of approximately 3.500 to 4.000 centiMorgan for the three genomes together, this would require the availability of sequence data from nearly 10.000 genome-specific amplicons to have a sufficiently high number of genes with SNPs available.

3.3. Detection Methods for SNPs in Wheat

Another point that is important concerning SNP analysis in wheat is the mode of detecting individual SNPs or large numbers of SNPs through multiplexed systems. Currently, there are large numbers of individual systems available that can be used for SNP analysis in plants (e.g. Lee et al., 2004; Giancola et al., 2006). With respect to wheat, basically no publications regarding SNP detection on a large scale exist so that it is not entirely clear what system would be best-suited. Furthermore, the hexaploid wheat genome will also make SNP analysis quite difficult. Most SNP assay systems are based on the identification of individual SNPs either in individual or multiplexed systems with two major technological approaches for that purpose (Gut 2001, 2004; Kwok, 2001). One approach requires the amplification of the respective locus with a set of specific primers followed by the actual assay by means of various detection systems. For example, the frequently used TaqMan system is employing for this a fluorescent dye and a linked quencher, that are separated from each other during the detection process. In Pyrosequencing, a minisequencing procedure is performed with a detection primer that is a few bases upstream or downstream of the actual SNP. Primer extension requires an assay primer that is directly adjacent to the SNP which will be extended by incorporating a fluorescently labelled didesoxynucleotide. The other approach does not require the amplification of specific fragments via PCR. Such techniques are mostly used in a multiplex fashion in order to detect large numbers of SNPs at many different loci. Examples for these are chip-based systems that require the hybridization of genomic DNA onto oligonucleotides which are capable of identifying the two alleles of an SNP through specific hybridization to one or the other allele. The oligo-ligation assay (OLA) system catalyzes the allele-specific ligation of two oligonucleotides. Again another technique that does not require prior amplification is the nuclease-based Invader-technology.

Both technological approaches will have their problems with regard to SNP analysis due to the fact that the respective SNPs have to be detected in the wheat

genome that contains three highly related genomes. Specifically, SNP systems that do not require a prior PCR step have the problem that usually all three wheat genomes directly adjacent to a given SNPs are identical. This means that the analysis primer will bind to more than one genome. As a result of that a codominant scoring of SNPs will be impossible or require an extremely quantitative system that in the worst case needs to be able to discriminate a 6:0 from a 5:1 and from a 4:2 ratio (6:0 - hybridization to all three genomes with the investigated genome homozygous for allele 1; 5:1 – hybridization to all three genomes with one genome heterozygous; 4:2 – hybridization to all three genomes with one genome homozygous for allele 2). At present, it is not clear whether this can be achieved with sufficient accuracy so that in most cases SNPs detected with these non-PCR-based systems could only be scored as presence/absence and thus providing only 50% of the actual information compared to a codominant marker. SNP detection methods that require a PCR amplification step prior to the actual SNP assay do not face this problem because for the amplification of the actual locus, genome-specific primers could be used. In that way only one of the genomes will be assayed and only the three allelic states also observed in a diploid organisms will be present so that codominant scoring will be possible. In principle, genome-specific primers already exist in case of SNP identification via comparative sequencing but the amplified fragments for comparative sequencing are usually large in order to obtain as much sequencing information as possible and for SNP analysis fragments should be amplified for most technologies that are less than 100 base pairs since larger fragments result frequently in unequal amplification and require the optimization of the amplification conditions for each individual marker making high-throughput analysis difficult.

4. OUTLOOK TOWARDS THE USE OF SNPS IN WHEAT BREEDING

The future use of SNPs in wheat breeding has to be regarded from different angles. The general use of SNPs in wheat breeding as a replacement of other markers (especially microsatellite markers) is still far away. Based on the first data towards SNP identification and SNP frequencies in wheat breeding material, it will need quite some time until a sufficiently large number of SNPs with a good allele frequency will be identified. This will require the coordinated efforts of many different laboratories since, for a reasonable genome coverage a minimum of 10.000 genome-specific amplicons have to be investigated for the presence of SNPs. Although first efforts have been made to identify SNPs in wheat on a large scale through the wheat HapMap project (http://wheat.pw.usda.gov/SNP), it is very likely that such an effort will require several years until completion. Furthermore, it is at present not clear what technology could be used for assaying SNPs on a large and cost efficient scale in hexaploid wheat since all currently available technologies have problems with SNP analysis in hexaploid species. Thus it is very likely that over the next five years genome-wide SNP analysis will not be used widely in wheat breeding efforts.

In the short term, SNP markers will however gain significantly more importance for the analysis of individual genes. With established techniques for the generation of genome-specific primers and comparative sequencing, SNP analysis with a variety of techniques that involve a pre-amplification step with genome-specific primers will certainly be used for the analysis of specific genes that have a known influence on specific traits or have to be considered as interesting candidate genes. Examples for this have already been published such as the SNP analysis of the grain hardiness locus (Giroux and Morris, 1998; Huang and Röder, 2005), the analysis of storage protein loci which control aspects of milling and baking quality (Zhang et al., 2003; Ravel et al., 2006) and other genes (Ellis et al., 2002; Yanagisawa et al., 2003; Blake et al., 2004). It is certain that we will see more progress being made in the near future towards the SNP analysis of such genes in wheat breeding since such individual analyses can be performed in a cost-efficient way on large numbers of individual plants once such assays have been developed and optimized. The same will be true for markers tightly linked traits of interest in wheat breeding which are converted to SNP markers for high-throughput analysis.

ACKNOWLEDGEMENTS

We thank all former and current members of the laboratory at the IPK that were involved in the various aspects of the wheat research. At TraitGenetics, the contribution of Dr. Hartmut Luerßen and his staff is acknowledged for the unpublished work concerning SNP development. Funding for the SNP work at TraitGenetics was provided in part through a grant from the BMBF (03i0605).

REFERENCES

Ablett G, Hill H, Henry RJ (2006) Sequence polymorphism discovery in wheat microsatellite flanking regions using pyrophosphate sequencing. Mol Breed 17:281–289

Akhunov ED, Akhunova AR, Linkiewicz AM, Dubcovsky J, Hummel D, Lazo G, Chao S, Anderson OD, David J, Qi L et al (2003a) Synteny perturbations between wheat homeologous chromosomes caused by locus duplications and deletions correlate with recombination rates along chromosome arms. Proc Natl Acad Sci USA 100:10836–10841

Akhunov ED, Goodyear AW, Geng S, Qi LL, Echalier B, Gill BS, Miftahudin, Gustafson JP, Lazo G, Chao S et al (2003b) The organization and rate of evolution of wheat genomes are correlated with recombination rates along chromosome arms. Genome Res 13:753–763

Anderson JA, Stack RW, Liu S, Waldron BL, Fjeld AD, Coyne C, Moreno-Sevilla B, Fetch JM, Song QJ, Cregan PB, Frohberg RC (2001) DNA markers for *Fusarium* head blight resistance QTLs in two wheat populations. Theor Appl Genet 102:1164–1168

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

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

Bai G, Shaner G (2004) Management and resistance in wheat and barley to *Fusarium* head blight. Annu Rev Phytopathol 42:135–161

Blake NK, Sherman JD, Dvorak J, Talbert LE (2004) Genome-specific primer sets for starch biosynthesis genes in wheat. Theor Appl Genet 109:1295–1302

Börner A, Schumann E, Fürste A, Cöster H, Leithold B, Röder M, Weber W (2002) Mapping of quantitative trait loci determining agronomic important characters in hexaploid wheat (*Triticum aestivum* L.). Theor Appl Genet 105:921–936

Bryan GJ, Collins AJ, Stephenson P, Orry A, Smith JB, Gale MD (1997) Isolation and characterisation of microsatellites from hexaploid wheat. Theor Appl Genet 94:557–563

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

Caldwell KS, Dvorak J, Lagudah ES, Akhunov E, Luo MC, Wolters P, Powell W (2004) Sequence polymorphism in polyploidy wheat and their D-genome diploid ancestor. Genetics 167: 941–947

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

Christiansen MJ, Andersen SB, Ortiz R (2002) Diversity changes in an intensively bred wheat germplasm during the 20th century. Mol Breed 9:1–11

del Blanco IA, Frohberg RC, Stack RW, Berzonsky WA, Kianian SF (2003) Detection of QTL linked to *Fusarium* head blight resistance in Sumai 3-derived North Dakota bread wheat lines. Theor Appl Genet 106:1027–1031

Devos KM, Bryan GJ, Collins AJ, Gale MD (1995) Application of two microsatellite sequences in wheat storage proteins as molecular markers. Theor Appl Genet 90:247–252

Devos KM, Ma J, Pontaroli AC, Pratt LH, Bennetzen JL (2005) Analysis and mapping of randomly chosen bacterial artificial chromosome clones from hexaploid bread wheat. Proc Natl Acad Sci USA 102:19243–19248

Donini P, Law JR, Koebner RMD, Reeves JC, Cooke RJ (2000) Temporal trends in the diversity of UK wheat. Theor Appl Genet 100:912–917

Ellis MH, Spielmeyer W, Gale KR, Rebetzke GJ, Richards RA (2002) 'Perfect' markers for the *Rht-B1b* and *Rht-D1b* dwarfing genes in wheat. Theor Appl Genet 105:1038–1042

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

Feldman M, Levy AA (2005) Allopolyploidy – a shaping force in the evolution of wheat genomes. Cytogenet Genome Res 109:250–258

Foster MW, Sharp RR (2004) Beyond race: towards a whole-genome perspective on human populations and genetic variation. Nat Rev Genet 5:790–796

Gale MD, Devos KM (1998) Comparative genetics in the grasses. Proc Natl Acad Sci USA 95:1971–1974

Gao IF, Jing RL, Huo NX, Li Y, Li XP, Zhou RH, Chang XP, Tang JF, Ma ZY, Jia JZ (2004) One hundred and one new microsatellite loci derived from ESTs (EST-SSRs) in bread wheat. Theor Appl Genet 108:1392–1400

Giancola S, McKhann HI, Bérard A, Camilleri C, Durand S, Libeau P, Roux F, Reboud X, Gut IG, Brunel D (2006) Utilization of the three high-througput SNP genotyping methods, the GOOD assay, Amplifluor and TaqMan, in diploid and polyploid plants. Theor Appl Genet 112:1115–1124

Giroux MJ, Morris CF (1998) Wheat grain hardness results from highly conserved mutations in the friabilin components puroindoline A and B. Proc Natl Acad Sci USA 95:6262–6266

Gupta PK, Balyan HS, Edwards KJ, Isaac P, Korzun V, Röder M, Gautier M-F, Joudrier P, Schlatter AR, Dubcovscy J et al (2002) Genetic mapping of 66 new microsatellite (SSR) loci in bread wheat. Theor Appl Genet 105:413–422

Gut IG (2001) Automation in genotyping single nucleotide polymorphisms. Hum Mutat 17:475–492

Gut IG (2004) An overview of genotyping and single nucleotide polymorphisms (SNPs). In: Rapley R, Harbon S (eds) Molecular analysis and genome discovery. Wiley, Chichester, pp 43–64

Guyomarc'h H, Sourdille P, Charmet G, Edwards KJ, Bernard M (2002) Characterisation of polymorphic microsatellite markers from *Aegilops tauschii* and transferability to the D-genome of bread wheat. Theor Appl Genet 104:1164–1172

Hancock JM (1999) Microsatellites and other simple sequences: genomic context and mutational mechanisms. In: Goldstein DB, Schlötterer C (eds) Microsatellites: evolution and applications. Oxford University Press, Oxford, pp 1–6

Holton TA, Christopher JT, McClure L, Harker N, Henry RJ (2002) Identification and mapping of polymorphic SSR markers from expressed gene sequences of barley and wheat. Mol Breed 9:63–71

Huang XQ, Röder MS (2005) Development of SNP assays for genotyping the puroindoline b gene for grain hardness in wheat using pyrosequencing. J Agric Food Chem 23:2070–2075

Huang XQ, Börner A, Röder MS, Ganal MW (2002) Assessing genetic diversity of wheat (*Triticum aestivum* L.) germplasm using microsatellite markers. Theor Appl Genet 105:699–707

Huang XQ, Cöster H, Ganal MW, Röder MS (2003) Advanced backcross QTL analysis for the identification of quantitative trait loci alleles from wild relatives of wheat (*Triticum aestivum* L.). Theor Appl Genet 106:1379–1389

Huang XQ, Kempf H, Ganal MW, Röder MS (2004) Advanced backcross QTL analysis in progenies derived from a cross between a German elite winter wheat variety and a synthetic wheat (*Triticum aestivum* L.). Theor Appl Genet 109:933–943

Johnson GC, Esposito L, Barratt BJ, Smith AN, Heward J et al (2001) Haplotype tagging for the identification of common disease genes. Nat Genet 29:233–237

Kahl G, Mast A, Tooke N, Shen R, van den Boom D (2005) Single nucleotide polymorphisms: detection techniques and their potential for genotyping and genome mapping. In: Mekshem K, Kahl G (eds) The handbook of plant genome mapping: genetic and physical mapping. Wiley-VCH, Weiheim, pp 75–107

Keller B, Feuillet C, Yahiaoui N (2005) Map-based isolation of disease resistance genes from bread wheat: cloning in a supersize genome. Genet Res Camb 85:93–100

Koebner RMD, Summers RW (2003) 21st century wheat breeding: plot selection or plate selection. Trends Biotechnol 21:59–63

Koebner RMD, Powell W, Donini P (2001) Contribution of DNA molecular marker technologies to the genetics and breeding of wheat and barley. Plant Breed Rev 21:181–220

Kwok P-Y (2001) Methods for genotyping single nucleotide polymorphisms. Ann Rev Genomics Hum Genet 2:235–258

Lazo GR, Chao S, Hummel DD, Edwards H, Crossman CC, Lui N, Matthews DE, Carollo VL, Hane DL, You FM et al (2004) Development of an expressed sequence tag (EST) resource for wheat (*Triticum aestivum* L.): EST generation, unigene analysis, probe selection and bioinformatics for a 16,000-locus bin-delineated map. Genetics 168:585–593

Lee S-H, Walker DR, Cregan PB, Boerma HR (2004) Comparison of four flow cytometric SNP detection assays and their use in plant improvement. Theor Appl Genet 110:167–174

Manifesto MM, Schlatter AR, Hopp HE, Suarez EY, Dubcovsky J (2001) Quantitative evaluation of genetic diversity in wheat germplasm using molecular markers. Crop Sci 41:682–690

Martins-Lopes P, Zhang H, Koebner R (2001) Detection of single nucleotide mutations in wheat using single strand conformation polymorphism gels. Plant Mol Biol Rep 19:159–162

Peng JH, Lapitan NLV (2005) Characterization of EST-derived microsatellites in the wheat genome and development of eSSR markers. Funct Integr Genomics 5:80–96

Perretant MR, Cadalen T, Charmet G, Sourdille P, Nicolas P, Boeuf C, Tixier MH, Branlard G, Bernard S, Bernard M (2000) QTL analysis of bread-making quality in wheat using a doubled haploid population. Theor Appl Genet 100:1167–1175

Pestsova E, Röder MS (2002) Microsatellite analysis of wheat chromosome 2D allows the reconstruction of chromosomal inheritance in pedigrees of breeding programmes. Theor Appl Genet 106:84–91

Pestsova E, Ganal MW, Röder MS (2000) Isolation and mapping of microsatellite markers specific for the D genome of bread wheat. Genome 43:689–697

Plaschke J, Ganal MW, Röder MS (1995) Detection of genetic diversity in closely related bread wheat using microsatellite markers. Theor Appl Genet 91:1001–1007

Powell W, Langridge P (2004) Unfashionable crop species flourish in the 21st century. Genome Biol 5:233

Qi LL, Echalier B, Chao S, Lazo GR, Butler GE, Anderson OD, Akhunov ED, Dvorak J, Linkiewicz AM, Ratnasiri A et al (2004) A chromosome bin map of 16,000 expressed sequence tag loci and distribution of genes among the three genomes of polyploidy wheat. Genetics 168:701–712

Rafalski JA (2002a) Application of single nucleotide polymorphisms in crop genetics. Curr Opin Plant Biol 54:357–374

Rafalski JA (2002b) Novel genetic mapping tools in plants: SNPs and LD-based approaches. Plant Sci 162:329–333

Ravel C, Nagy IJ, Martre P, Sourdille P, Darevet M, Balfourier F, Pont C, Giancola S, Praud S, Charmet G (2006) Single nucleotide polymorphism, genetic mapping, and expression of genes coding for the DOF wheat prolamin-box binding factor. Funct Integr Genomics 6:310–321

Röder MS, Plaschke J, König SU, Börner A, Sorrells ME, Tanksley SD, Ganal MW (1995) Abundance, variability and chromosomal location of microsatellites in wheat. Mol Gen Genet 246:327–333

Röder MS, Korzun V, Gill B, Ganal MW (1998a) The physical mapping of microsatellite markers in wheat. Genome 41:278–283

Röder MS, Korzun V, Wendehake K, Plaschke J, Tixier MH, Leroy P, Ganal MW (1998b) A microsatellite map of wheat. Genetics 149:2007–2023

Röder MS, Wendehake K, Korzun V, Bredemeijer G, Laborie D, Bertrand L, Isaac P, Rendell S, Jackson J, Cooke RJ et al (2002) Construction and analysis of a microsatellite-based database of European wheat varieties. Theor Appl Genet 106:67–73

Röder MS, Huang X-Q, Ganal MW (2004) Wheat microsatellites: potential and implications. In: Lörz H, Wenzel G (eds) Biotechnology in agriculture and forestry, Vol. 55. Molecular marker systems, Springer, Berlin Heidelberg, pp 255–266

Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD et al (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 409:928–933

Somers DJ, Kirkpatrick R, Maniwa M, Walsh A (2003) Mining single-nucleotide polymorphisms from hexaploid wheat ESTs. Genome 49:431–437

Somers DJ, Isaac P, Edwards K (2004) A high-density microsatellite consensus map for bread wheat (*Triticum aestivum* L.). Theor Appl Genet 109:1105–1114

Song QJ, Fickus EW, Cregan PB (2002) Characterization of trinucleotide SSR motifs in wheat. Theor Appl Genet 104:286–293

Song QJ, Shi JR, Singh S, Fickus EW, Costa JM, Lewis J, Gill BS, Ward R, Cregan PB (2005) Development and mapping of microsatellite (SSR) markers in wheat. Theor Appl Genet 110:550–560

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

Stein N, Feuillet C, Wicker T, Schlagenhauf E, Keller B (2000) 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 97:13436–13441

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

Torada A, Koike M, Mochida K, Ogihara Y (2006) SSR-based linkage map with new markers using an intraspecific population of common wheat. Theor Appl Genet 112:1042–1051

Van Deynze AE, Nelson JC, Yglesias ES, Harrington SE, Braga DP, McCouch SR, Sorrells ME (1995) Comparative mapping in grasses: Wheat relationships. Mol Gen Genet 248:744–754

Varshney RK, Graner A, Sorrells ME (2005) Genic microsatellite markers in plants: features and applications. Trends Biotechnol 23:48–55

Wicker T, Stein N, Albar L, Feuillet C, Schlagenhauf E, Keller B (2001) Analysis of a contiguous 211 kb sequence in diploid wheat (*Triticum monococcum* L.) reveals multiple mechanisms of genome evolution. Plant J 26:307–316

Yanagisawa T, Kiribuchi-Otobe C, Hirano H, Suzuki Y, Fujita M (2003) Detection of single nucleotide polymorphism (SNP) controlling the waxy character in wheat by using a derived cleaved amplified polymorphic sequence (dCAPS) marker. Theor Appl Genet 107:84–88

Yu JK, Dake TM, Singh S, Benscher D, Li W, Gill B, Sorrells ME (2004) Development and mapping of EST-derived simple sequence repeat markers for hexaploid wheat. Genome 47:805–818

Zhang W, Gianibelli MC, Ma W, Rampling L, Gale KR (2003) Identification of SNPs and development of allele-specific PCR markers for gamma-gliadin alleles in *Triticum aestivum*. Theor Appl Genet 107:130–138

Zhang LY, Ravel C, Bernard M, Balfourier F, Leroy P, Feuillet C, Sourdille P (2006) Transferable bread wheat EST-SSRs can be useful for phylogenetic studies among the *Triticeae* species. Theor Appl Genet 113:407–418

Zhou W, Kolb FL, Bai G, Shaner G, Domier LL (2002) Genetic analysis of scab resistance QTL in wheat with microsatellite and AFLP markers. Genome 45:719–727

Zhu YL, Song QJ, Hyten DL, van Tassell CP, Matukumalli K, Grimm DR, Hyatt SM, Fickus EW, Young ND, Cregan PB (2003) Single-nucleotide polymorphisms in soybean. Genetics 163:1123–1134

CHAPTER 2

MOLECULAR MARKERS AND QTL ANALYSIS FOR GRAIN QUALITY IMPROVEMENT IN WHEAT

DOMENICO LAFIANDRA[1,*], MARIA CORINNA SANGUINETI[2], MARCO MACCAFERRI[2] AND ENZO DEAMBROGIO[3]

[1]*Department of Agrobiology and Agrochemistry, University of Tuscia, Viterbo, Italy*
[2]*Department of Agroenvironmental Sciences and Technology, University of Bologna, Bologna, Italy*
[3]*Società Produttori Sementi Bologna, Divisione Ricerca, Argelato (BO), Italy*

Abstract: Molecular marker technology is playing an increasingly important role in the selection of wheat lines with improved quality attributes. This is due to the identification of molecular markers tightly linked to chromosome regions involved in the control of important quality characteristics such as dough properties, grain hardness, semolina and flour colour, grain protein content and starch composition, which strongly influence wheat end use, and its nutritional and market value. Marker assisted selection (MAS) will increase the efficiency of the breeding process, particularly when phenotyping requires laborious and time-consuming analyses, performed in advanced generations because of the relatively large amount of grain required. Moreover, the implementation of MAS allows the selection of individuals carrying the favourable alleles at the target loci, and also the pyramiding of favourable QTL alleles from different sources and for different traits. This not withstanding, the progress obtained until now in applying MAS to quality characteristics has been slow compared to other traits.

1. INTRODUCTION

Durum and bread wheats are the major foods for much of the human population and are mainly consumed as processed products because of the unique functional properties they confer to the derived foods. Improving end-use quality has become of increasing importance to wheat breeders over the past few decades with an emphasis on developing cultivars (cvs.) of durum and bread wheat for specific applications such as bread (leavened, flat and steamed), other baked goods (cakes,

*Author for corresspondence: lafiandr@unitus.it

R.K. Varshney and R. Tuberosa (eds.), Genomics Assisted Crop Improvement: Vol. 2: Genomics Applications in Crops, 25–50.

cookies, crackers, etc.), pasta and noodles, and a wide range of other products of restricted regional uses. Processing and end-use quality of wheat-based products are influenced by several factors such as protein content and composition, starch, kernel hardness and lipids.

Some of these traits are complex (i.e. controlled by several genes and influenced by environmental factors and management procedures), while others are simply inherited. The selection of superior breeding lines with improved quality characteristics, as well as the dissection of their complexity, are being greatly favoured by the establishment and exploitation of biochemical and molecular markers (Gale 2005; Howitt et al. 2006). The ultimate aim is to tailor wheat cvs. with improved quality and to satisfy the specific end-use requirements of the market and consumers.

2. GRAIN PROTEIN CONTENT

Grain protein content (GPC) is one of the major quality traits in durum and bread wheat strongly associated to end-use performance of derived products. In bread wheat, the existence of a linear relationship between protein content and loaf volume have been reported (Finney and Barrimore 1948); similarly, in durum wheat it has been clearly established that high protein content and strong gluten are major factors for producing pasta with superior cooking properties, better cooking firmness and tolerance to overcooking (D'Egidio et al. 1990; Marchylo et al. 1998). The variation induced by environmental factors and the negative association with grain yield have hindered progress in increasing grain protein content by traditional breeding approaches, but the availability of genetic stocks such as aneuploids (see Snape et al. 1995 for advantages in locating QTL's using single chromosome recombinant lines) and large gene pools together with marker-mediated approaches are leading to the improvement of this trait.

Genes to improve protein content have been sought among wheat related species and promising materials have been detected in the tetraploid wild wheat progenitor *T. turgidum* ssp *dicoccoides* with values ranging from 14 to 29% protein (Avivi 1978; Grama et al. 1984). In particular, the accessions FA15-3 of *T. turgidum* ssp *dicoccoides* identified by Avivi (1978) has been used to introduce the high protein content trait into durum and bread wheats. Joppa and Cantrell (1990) developed a set of disomic substitution lines in the durum wheat cv. Langdon (LDN); in each line a pair of chromosomes of LDN were replaced by their homeologues derived from the *T. dicoccoides* accession FA15-3. The analysis of a population of recombinant substitution lines (RSLs) obtained from the cross between the high GPC line LDN(DIC-6B) and the cv. LDN led to the identification of a quantitative trait locus (QTL) for GPC on the short arm of chromosome 6B in the *Xmwg79–Xabg387* interval (Joppa et al. 1997). This QTL, designated *QGpc.ndsu.6Bb*, accounted for approximately 66% of the variation in protein content in this cross, suggesting that a major gene or a closely linked group of genes were segregating in this population.

Additionally, the significant increase in GPC was associated with an increase in the pasta quality (Joppa et al. 1991), while not significantly affecting grain yield or

kernel weight (Cantrell and Joppa 1991). Following these results, Khan et al. (2000) developed PCR-based markers to facilitate the introgression of the *QGpc.ndsu.6Bb* region into commercial tetraploid and hexaploid varieties: the microsatellite markers *Xgwm508* and *Xgwm193* proved particularly useful to monitor the introgression of the *T. dicoccoides* chromatin containing the high GPC gene(s). More recently, using field trials with ten replications and a large set of secondary RSLs obtained from the cross between LDN and RSL65, in a line carrying the complete *T. dicoccoides* segment flanked by the *Xgwm508* and *Xgwm193* markers, the source of GPC variation was mapped as a single Mendelian locus within a 2.7-cM region encompassed by the restriction fragment length polymorphism (RFLP) markers *Xcdo365* and *Xucw67*. This locus has been designated *Gpc-B1* (Olmos et al. 2003). To further delimit the location of *Gpc-B1*, the microcolinearity between rice and wheat present in this region was used by Distelfeld et al. (2004, 2006). This approach resulted in narrowing the position of the *Gpc-B1* locus to a 0.3-cM region flanked by PCR markers *Xucw79* and *Xucw71* and in identifying a codominant PCR marker (*Xuhw89*) tightly linked to the *Gpc-B1* locus. A candidate gene (a transcription factor regulating senescence) has now been cloned (Uauy et al. 2006), thus providing the possibility of using perfect markers in marker-assisted selection (MAS) programmes.

Mesfin et al. (1999) used three hard red spring wheat recombinant lines to identify genomic regions associated with high protein content inherited from *T. turgidum* ssp *dicoccoides.* A single region associated to high GPC was detected with five RFLP markers located near the centromere on chromosome 6B. It is possible that this high protein content QTL corresponds to the locus already described by Joppa et al. (1997); in fact, both research groups used the same *T. turgidum* ssp *dicoccoides* accession as donor of high GPC.

Khan et al. (2000) compared the *T. dicoccoides* accession FA15-3 and Glupro, a hexaploid high GCP cv. obtained from a three-way cross involving two bread wheat cvs. and FA15-3 (donor of the GPC gene/s), to determine the size of the *dicoccoides* chromosome 6B segment transferred to the bread cv. Using microsatellites, amplified fragment length polymorphism (AFLPs) and RFLPs, the authors showed that two separate segments from the *dicoccoides* accession were transferred to Glupro chromosome 6B, one in the distal region of the long arm and the other in the proximal region of the short arm, this latter encompassing the GPC QTL with the highest LOD identified by Joppa et al. (1997).

In a different study, Blanco et al. (1996) tested a set of 65 recombinant inbred lines developed by single seed descent from the cross between the durum wheat cv. 'Messapia' (low protein content) and the wild tetraploid *T. dicoccoides* accession MG 4343 (high protein content) to locate QTLs controlling GPC. One QTL was identified on each of the chromosome arms 4BS, 5AL, 6BS and 7BS, and two distinct QTLs were identified on chromosome 6A. According to the environment considered, the six QTLs explained from 49.2 to 56.4% of the phenotypic variation for GPC. Only some of the markers were found to be negatively associated with grain yield and/or seed weight in one or two environments. An additional locus

was detected on chromosome 7A (Blanco et al. 2002) when the same lines were cultivated in different environments and years.

Moreover, Blanco et al. (2006) used the backcross inbred lines approach to introgress *T. dicoccoides* high GPC alleles in durum wheat and to identify molecular markers linked to chromosome regions controlling high GPC. Using SSRs and AFLPs, three QTLs with major effects on GPC were detected on chromosome arms 2AS, 6AS and 7BL, identified by the markers *Xcfa2164*, *XP39M37(250)* and *Xgwm577*, respectively.

Prasad et al. (1999) tested a set of RILs derived from a cross between two bread wheat genotypes PH132 (high GPC) and WL711 (low GPC) and identified a sequence-tagged microsatellite site (STMS) *Xwmc41* associated to a QTL for protein content (designated *QGpc.ccsu-2D.1)*, located on the chromosome arm 2DL. The same research group identified additional molecular markers linked to QTLs for high GPC: a significant association was found between *Xwmc415* and a QTL designated *QGpc.ccsu-5A.1* (Harjit-Singh et al. 2001).

Additional studies evidenced several major GPC QTLs on different chromosomes in the cultivated and the wild wheat gene pools and proved allelic relationships between some of the detected loci (Snape et al. 1995; Sourdille et al.1999; Perretant et al. 2000; Turner et al. 2004).

3. GLUTEN PROTEINS

It is now accepted that the amount and composition of gluten proteins (the viscoelastic mass remaining after washing the dough with salt solutions) is mainly responsible for qualitative differences, such as pasta-making and bread-making properties, among durum and bread wheat cvs. (see Shewry et al. 2003, for a comprehensive review on gluten proteins). Gluten proteins are composed of two fractions, gliadins and glutenins, mainly differing for their capability to form polymers: gliadins are monomeric molecules, while glutenins are polymers containing different subunits connected by intermolecular disulphide bonds. When glutenin polymers are reduced and released components separated using sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE), two major groups are identified which have been designated as High Molecular Weight Glutenin Subunits (HMW-GS) and Low Molecular Weight Glutenin Subunits (LMW-GS). Technological properties of both durum and bread wheat doughs are positively correlated with the molecular size of glutenin polymers, which are highly variable and can reach millions of daltons.

The gliadin fraction includes proteins which are subdivided into α/β-, γ- and ω-types according to their N-terminal amino acid sequences, where disulphide bonds, if present, are intramolecular. Genes encoding gliadin components are located on the short arm of the homoeologous groups 1 and 6 chromosomes of the A, B and D genomes, at the *Gli-1* and *Gli-2* loci.

HMW-GS are encoded by genes present at the *Glu-1* loci on the long arms of the homoeologous group 1 chromosomes. Each *Glu-1* locus contains two tightly linked

genes encoding for two types of subunits (x- and y-type characterised by high and low molecular weight, respectively) differing in structural characteristics such as the number of cysteine residues, and the size and composition of the repetitive domain. The y-type gene present at the *Glu-A1* locus is always silent in tetraploid and hexaploid cultivated wheat, whereas the x-type gene at the same locus and the y-type at the *Glu-B1* locus are expressed only in some cvs.; this leads to variation in the number of subunits from three to five in bread wheat and from two to three in durum wheat. SDS-PAGE of seed proteins has been extensively used to assess HMW-GS composition of bread wheat cvs. Allelic variation has been detected at each *Glu-1* locus and the analysis of crosses between lines with different HMW-GS composition allowed the identification of subunits affecting breadmaking properties of different bread wheat cvs. In particular, the subunit pair 5+10 and 2+12 encoded by genes present at the *Glu-D1* locus have been correlated with good and poor breadmaking properties, respectively. The superiority of the pair 5+10 may be due to the additional cysteine residue present in subunit 1Dx5 at the beginning of the repetitive domain; this residue would promote the formation of glutenin polymers of larger size, thus endowing dough with increased strength (Greene et al. 1988; Lafiandra et al. 1993).

Similar differential effects of allelic subunits, encoded by the *Glu-A1* and *Glu-B1* loci have also been reported (Payne 1987).

LMW-GS are encoded by multigene families located on the homoeologous group 1chromosomes at the *Glu-3* loci (*Glu-A3*, *Glu-B3 and Glu-D3*) tightly linked to the *Gli-1* loci. As compared to the HMW-GS , the LMW-GS proved to be more difficult to study and characterise due to their heterogeneity and a range of different analytical tools have been used to elucidate their complexity and role in quality aspects. LMW-GS have been subdivided into the B, C and D types according to their biochemical characteristics (Jackson et al. 1983). Moreover, the B type subunits can be divided into three groups (LMW-m, LMW-s and LMW-i) based on the first amino acid residue (Met, Ser and Ile, respectively) of the mature protein. The LMW-i type has been detected more recently than the others and seems to be mainly encoded by genes present at the *Glu-A3* locus (Cloutier et al. 2001; Ikeda et al. 2002; Zhang et al. 2004). Additionally, this last group of subunits shows striking structural differences compared to the LMW-m and LMW-s groups as they lack an N-terminal region and all the cysteines are localized in the C-terminal region, though sharing the same number of cysteine residues (eight) with the other LMW-GS types. This difference in cysteine distribution might impact glutenin polymer formation and, more in general, gluten interactions and be responsible for quality differences. The C and D type subunits are composed mainly of proteins related to α/β-,γ- and ω-gliadins which have variable numbers of cysteine residues. It has been proposed that they are incorporated into the polymeric network by virtue of unpaired cysteines (D'Ovidio and Masci 2004).

Several studies have indicated that the allelic variation at the *Glu-3* loci is also associated to dough quality both in bread and durum wheat, and ranking of different alleles has been reported (see Juhász and Gianibelli 2006 for a review).

The availability of gene sequence data for both LMW- and HMW-GS made it possible to generate molecular markers, known as functional or perfect markers, closely linked or located within the gene of interest (Andersen and Lubberstedt 2003; Varshney et al. 2005), thus making available powerful tools to incorporate superior alleles in breeding materials.

In durum wheat, LMW-GS encoded by genes present at the complex *Glu-B3* locus have been associated to both pasta cooking properties and bread making properties (Boggini and Pogna 1989). The two major alleles (*LMW-1* and *LMW-2*) detected at this locus are related to poor and good gluten viscoelastic properties, respectively (Payne et al. 1984; Pogna et al. 1990).

In the attempt to develop molecular markers capable of discriminating durum wheat genotypes with the *LMW-1* or *LMW-2* allele and thus providing an alternative approach to electrophoretic techniques, D'Ovidio (1993) developed a pair of oligonucleotide primers that produced two main PCR amplification products in all the analysed genotypes. The PCR fragment of lower size, common to all the durum genotypes tested, was not assigned to any specific chromosome, whereas the other was polymorphic between the lines with different allelic LMW-GS. Subsequently, D'Ovidio and Porceddu (1996) developed a new set of primers yielding a single amplification product, ranging in size between 780 (*LMW-1*) and 830 bp (*LMW-2*), capable of differentiating durum wheat cvs. with poor or good pasta making properties.

More recently, efforts from different research groups resulted in the development of bread wheat specific primers that permitted to amplify, isolate and study several LMW-GS genes associated to specific *Glu-3* loci, thereby contributing to the detailed description of their structural features. Moreover, these primers are capable of differentiating homo- and homoeoalleles thus making available molecular markers suitable to follow the introgression of superior alleles in breeding programmes (Zhang et al. 2004; Ikeda et al. 2006; Long et al. 2006; Zhao et al. 2006). In particular, Zhang et al. (2004) isolated and characterised the complete coding sequence of one LMW-GS gene for each of the seven alleles (from *Glu-A3a* to *Glu-A3g*) present in different bread wheat cvs. All sequences were classified as i-type LMW-GS genes based on the presence of an N-terminal isoleucine residue and eight cysteine residues located within the C-terminal domain of the predicted, mature amino acid sequence. Comparison of gene sequences obtained from different alleles showed a wide range of sequence identity between the genes, with between 1 and 37 single nucleotide polymorphisms and between one and five insertion/deletion events. Allele-specific PCR markers were designed based on the DNA polymorphisms identified between the LMW-GS genes, and these markers were validated against a group of bread wheat cvs. containing different *Glu-A3* alleles. Using the same approach Zhao et al. (2006) developed PCR markers for the identification of a few *Glu-D3* alleles.

Although SDS-PAGE has been invaluable for studying HMW-GS, it has some limitations, such as co-migration of certain subunits or difficulty in detecting differences in expression levels, resulting in incorrect identification of alleles with

different functional effects. Additionally, glutenin composition analysis can be carried out only on the mature grains. To overcome these limitations, allele specific PCR markers, based on DNA polymorphism present among the glutenin subunit genes, have been obtained to assess allelic variation at the *Glu-1* loci and to provide high-throughput analysis of different alleles in breeding materials. Differently from SDS-PAGE, PCR can be performed on leaf samples collected during the vegetative growth stages; this in turn allows for the selection of the genotypes with the desired allele composition before flowering.

D'Ovidio and Anderson (1994) used the presence of some differences in the nucleotide sequences of the 1Dx2 and 1Dx5 glutenin genes to develop two oligonucleotides specific for the subunit 1Dx5. Similarly, in the attempt to identify bread wheat lines carrying the superior pair of subunits 5+10, a codominant PCR marker was developed by Smith et al. (1994). Subsequently, primer specific for homo- and homoeo-allelic HMW-GS genes have been developed by several research groups (D'Ovidio et al. 1994, 1995; Lafiandra et al. 1997; Ahmad 2000; De Bustos et al. 2000, 2001; Radovanovic and Cloutier 2003).

The usefulness of molecular markers appears particularly relevant to discriminating different alleles for HMW-GS associated to the *Glu-B1* locus. In a study carried out on several Canadian bread wheat cvs., Marchylo et al. (1992) reported that, though allelic subunits 1Bx7 and 1Bx7* could not be distinguished on the basis of their elution times, when analysed by reversed phase high performance liquid chromatography, the proportion of the first subunit was significantly higher than the latter. D'Ovidio et al. (1997) suggested that the results of Marchylo et al. (1992) could be due to a gene duplication involving the gene coding for the 1Bx7 subunit. Recent electrophoretic and chromatographic analyses of wheat cvs. and landraces from around the world revealed the large diffusion of the overexpressed 1Bx7 allele designated *Glu-B1al* and its positive impact on dough technological properties (Butow et al. 2003, 2004; Vawser and Cornish 2004). Butow et al. (2004) utilised a co-dominant PCR marker, designed to amplify the 1Bx MAR region 750 bp upstream of the coding region of the *Glu-B1* gene, to discriminate bread wheat materials with the over-expressed subunit Bx7 from those with the subunit 1Bx7*. All the cvs. and landraces showing increased 1Bx7 expression (i.e. with the *Glu-B1al* allele) produced a fragment longer than that of the other cvs. (~563 vs. 520 bp), indicating the presence of a 43 bp insertion in the MAR of the *Glu-B1al* allele. A marker for the negative selection of the poor quality allele Bx6 was developed by Schwarz et al. (2004) using the DNA polymorphisms between the coding sequences of the x-type HMW-GS alleles 1Bx6, 1Bx7 and 1Bx17, whereas a set of dominant and codominant markers have been developed by Lei et al. (2006) for specific HMW-GS genes encoding 1By-type subunits. These markers resulted in an enhanced discrimination of alleles at the *Glu-B1* locus and particularly between alleles with contrasting effects on quality and difficult to screen by SDS-PAGE.

A further breakthrough in the use of PCR-based markers for the selection of HMW-GS alleles is represented by the introduction of the multiplex approach, in which more than one target sequence is amplified using more than one primer

pair, in order to score simultaneously for the presence of different homo- or homoeo-allelic variants (Ahmad 2000; Ma et al. 2003). Salmanowicz and Moczulski (2004) developed a multiplex PCR method to identify bread wheat genotypes with specific HMW-GS allele composition at the *Glu-1* complex loci (*Glu-A1*, *Glu-B1* and *Glu-D1*) by capillary electrophoresis (CE) with laser-induced fluorescence (LIF) detection. DNA fragments amplified using two triplex primer sets were well separated both by agarose slab-gel electrophoresis and CE-LIF; minor differences among the sequences of 1Ax2, 1Ax null, 1Bx6, 1Bx7, 1Bx17 and 1Dx5 genes have been revealed. CE–LIF can thus be an efficient alternative to standard procedures for early selection of useful wheat genotypes with good bread-making quality. In fact, compared to gel electrophoresis, CE-LIF requires samples of smaller volumes and results can be obtained in less than 20 minutes.

Novel alleles at the *Glu-3* and *Glu-1* loci have been identified in germplasm collections of tetraploid and hexaploid wheat and in related wild species, thus providing variation suitable for exploitation in wheat quality improvement (Shewry et al. 2006; D'Ovidio and Masci 2004).

4. STARCH

Starch is the major constituent of flour and semolina and accounts for 65–75% of wheat grain dry weight. Starch is composed of two types of polymers (amylose and amylopectin) whose relative amounts and structures are responsible for its unique physical and chemical properties with strong influences on functional properties of flour or semolina and on its specific uses in the food and manufacturing industries (Zeng et al. 1997; Yoo and Jane 2002). Amylose is an essentially linear α-1,4 glucan and contributes about 20–30% to the total starch, while amylopectin is a branched α-1,4 glucan containing about 5% α-1,6 branch points and constitutes the remaining 70–80% of the total starch. A number of enzymes are involved in starch synthesis with at least five known isoforms of starch synthases (SS, as reviewed by James et al. 2003). Four of these isoforms are involved only in amylopectin synthesis, with two forms of branching and debranching enzymes. The granule bound starch synthases (GBSSI or waxy proteins) are the sole starch synthases responsible for amylose synthesis in storage tissues (Nakamura et al. 1993). In bread wheat there are three waxy proteins with a molecular weight ranging from 59 to 60 kDa (Murai et al. 1999), designated Wx-A1, Wx-D1 and Wx-B1 encoded by three genes (*Wx-A1*, *Wx-D1* and *Wx-B1*) located on chromosome arms 7AS, 7DS and 4AL, respectively. The latter was originally located on chromosome 7BS before a translocation occurred between 7BS and 4AL chromosomes during wheat evolution (Miura and Tanii 1994; Yamamori et al. 1994). The different effects of the three waxy genes on the amylose content has also been assessed, with the *Wx-B1* gene showing the highest effect, followed by the *Wx-D1* and the *Wx-A1* gene (Miura et al. 1994, 1999).

Electrophoretic studies of bread and durum wheats led to the identification of partial waxy mutant lines, characterised by the lack of one or two waxy proteins.

Null *Wx-A1* and *Wx-B1* alleles have been found in Asian, European and North American wheat cvs. In particular, null alleles at the *Wx-A1* locus are fairly common in bread wheat from Japan, Korea and Turkey, while null alleles at the *Wx-B1* locus are very common in bread wheat from Australia and India (Yamamori et al. 1994, 1998). On the contrary, null alleles at the *Wx-D1* locus seem to occur more rarely (Yamamori et al. 1994; Boggini et al. 2001; Monari et al. 2005). Crossing of selected genotypes has permitted the combination of the different null alleles detected with the production of the entire set of partial and total waxy lines (Nakamura et al. 1995; Urbano et al. 2002). These genetic materials are of interest due to their impact on quality aspects and to their possible use in non-food applications. In fact, flour with reduced amylose content extends the shelf life of various baked products (Graybosch et al. 1998; Lee et al. 2001) and produces higher quality Asian noodles (Miura et al. 1994). In this respect, partial waxy wheat have been shown to be particularly suitable for the production of certain Asian noodles such as the Japanese Udon or white salted noodles, typically produced using medium-protein soft wheat. Wheat suitable for the production of Udon noodles typically has starch or flour with high swelling volumes and high peak pasting viscosity which have been associated with reduced amylose content and presence of GBSS null alleles, and in particular with the null *Wx-B1*. These results have been obtained by SDS-PAGE analysis of granule bound starch proteins, an approach which is particularly time consuming, as starch isolation is required before electrophoretical analysis, thus limiting the number of samples to be processed. In addition, electrophoretical results show poor resolution of different protein components and can only be carried out on mature grains. Thus, great efforts were devoted to the development of several PCR markers to study and assist the introgression of the waxy trait in advanced breeding lines.

A recessive PCR marker was developed by Briney et al. (1998) to identify lines null at the *Wx-B1* locus; it was thus possible an accurate starch quality identification using small amounts of leaf tissue or single seeds, enabling the screening of large segregating populations. Shariflou et al. (2001) developed a perfect codominant marker to discriminate the normal and mutant null alleles at the *Wx-D1* locus originated from the Chinese landrace ‘Baihoumai’. Moreover, five primer pairs were designed by McLauchlan et al. (2001) using the available gene sequences to target each of the three waxy homoeoallele loci. Multiplex PCR has been used by Nakamura et al. (2002) to screen a large number of wheat samples from different countries in order to study waxy mutations and trace their origin.

Similarly to waxy proteins, mutant lines lacking one of the three possible Starch Synthase II (SSII) proteins have been identified by Yamamori and Quynh (2000) and bread lines lacking all these three proteins have been produced (Yamamori and Quynh 2000). In these lines, the apparent amylose content is significantly higher than that of the wild type (Yamamori et al. 2006). High amylose wheat flours show lower swelling and lower peak viscosity than waxy and normal wheat flours. Products made using SSII-deficient wheat, with a 35–40% amylose content,

have slightly increased levels of resistant starch. Resistant starch (RS) refers to the portion of starch and starch products that resist digestion as they pass through the gastrointestinal tract.

At present, there is a lot of interest in resistant starches because they have a low glycaemic index and reduce the risk of type 2 diabetes and obesity. They also act as a functional fiber, being fermented in the large bowel by the gut bacteria that produce short-chain fatty acids (especially butyric acid) which are responsible for the beneficial effects for the human gut by reducing the risk of colorectal cancer. Additionally, high-amylose starch is low in energy and in total carbohydrate content and seems to slightly improve cooked pasta firmness (Brouns et al. 2002).

Shimbata et al. (2005) characterised the mutations occurring at each SSII locus and identified allele-specific PCR-based markers for each locus. The markers developed were capable of distinguishing among heterozygous, homozygous null and homozygous wild type plants thus facilitating the introgression of mutated alleles in elite cvs.

5. KERNEL HARDNESS

Kernel hardness refers to the texture of the endosperm. Generally, the grain of bread wheat is identified as being either soft or hard. Grain hardness is a very important trait affecting milling, baking and end-use quality of wheat. Soft wheat kernels fracture more easily, release numerous intact starch granules and produce fine-textured flours with reduced starch damage (Giroux and Morris 1997). During milling of hard wheat, fractures yield clean, well-defined particles (Maghirang and Dowell 2003) with larger mean particle size, forming coarser-textured flours with a higher level of starch damage compared to soft wheat. The tetraploid durum wheat has the highest grain hardness and therefore the highest starch damage after milling. Since damaged starch granules absorb more water than undamaged granules, hard wheat is preferred for yeast-leavened products, whereas soft wheat is preferred for the production of cookies and cakes.

Among the different methods used to measure grain hardness, the two most popular with breeders are the Near-Infrared Reflectance Method, and the Single-Kernel Characterization System (AACC 2003a) because they are easy to use and produce reliable results.

Studying the inheritance of grain hardness, Symes (1965) showed that the difference between hard and soft wheat was due to a single major gene and to modifying minor genes. The *Hardness* locus (*Ha*) is located on the short arm of chromosome 5D (Mattern et al. 1973). Although the locus is named *Hardness*, the dominant trait is softness.

As far as the biochemical basis of hardness is concerned, Greenwell and Schofield (1986) found a 15-kDa protein named “friabilin” on the surface of water-washed wheat starch. The friabilin in water-washed starch is present in a higher amount in soft wheat than in hard wheat and is absent in durum wheat (very hard). According to recent studies, the 15-kDa protein complex consists of at least five components:

puroindoline-a, puroindoline-b, grain softness protein (respectively coded by the *Pina-D1*, *Pinb-D1* and *Gsp-1* genes) and two alpha-amylase inhibitors (Clarke and Rahman 2005). The *Pina-D1*, *Pinb-D1* and *Gsp-1* genes have been found to be tightly linked to the *Ha* locus at the distal end of chromosome 5DS (Giroux and Morris 1997; Tranquilli et al. 2002). The wild-type puroindoline genes are associated with soft kernel texture (Lillemo and Morris 2000). On the other hand, it has been demonstrated that in all the hard wheat examined there was a mutation either in *Pina-D1* or *Pinb-D1* , but not in *Gsp-1* (Giroux and Morris 1997; Lillemo and Morris 2000; Morris et al. 2001).

Although it is well documented that the *Ha* locus on the chromosome arm 5DS is the main determinant of kernel hardness (Sourdille et al. 1996; Campbell et al. 1999; Perretant et al. 2000; Igrejas et al. 2002), it was also shown that this locus alone and the tightly-linked genes *Pina-D1* and *Pinb-D1* cannot explain all the phenotypic variation of this trait in the different mapping populations. Many authors used QTL analysis to genetically dissect the complexity of the hardness trait. As a result, other QTLs with minor effect on grain hardness were detected on different chromosomes: 2A, 2D, 5B, 6D with single-factor effects and 5A, 6D, 7A with interacting effects (Sourdille et al. 1996), 3A (Campbell et al. 1999), 1A and 6D (Perretant et al. 2000), 2B and 6B (Galande et al. 2001), 5D in a position not coincident with the *Ha* locus (Igrejas et al. 2002) and 3BL (Narasimhamoorthy et al. 2006). The QTLs detected using bi-parental mapping populations do not provide the complete information as to the genetic control of a complex trait. Moreover, the QTL effect is estimated comparing two alleles in one genetic background only. From the breeders' standpoint it is much more interesting to characterise a range of different alleles in a germplasm collection including accessions derived from different crosses involving more than just two parents, as occurs in an ordinary breeding programme. Recently, Crepieux et al. (2005) presented results of a QTL mapping study carried out in a real wheat breeding population. Using the approach of the variance component (VC) based on the identity by descent (IBD), they detected two QTLs for grain hardness: one is likely to be associated to the *Ha* locus on the 5DS chromosome arm, the other is on the 1D chromosome close to the *Glu-D1* locus. However, in this latter case, they could not rule out the possibility of an artefact caused by the storage protein. Applying the "mixed-model analysis" to the same set of data, Arbelbide and Bernardo (2006) detected two markers strongly associated with kernel hardness in the multiple-marker analysis. One of the markers was very close to the *Pinb-D1* gene and therefore to the *Ha* locus. The second marker was close to the end of the 1AS chromosome arm, where the *Glu-A3* locus is located. A marker close to the QTL on 1D detected by Crepieux et al. (2005) was significant only in the single-marker analysis.

The QTL analysis is a first step towards the identification of the genes underlying each QTL. For a more efficient wheat breeding approach, the availability of reliable markers tightly linked to the QTL(s) or, even better, to the gene(s) of interest is very useful.

The puroindoline proteins a and b form the molecular basis of a large part of the wheat grain hardness (Morris et al. 2001). The availability of gene-specific primers allows the full-length amplification of the *Pina-D1* and *Pinb-D1* genes showing the sequence differences between alleles, which might have a different role in wheat breeding (Massa et al. 2004).

It has been demonstrated that the different *Pina-D1* and *Pinb-D1* alleles have a great influence on milling properties and also affect bread baking traits (Martin et al. 2001; Eagles et al. 2006). It is also known that, in synthetic hexaploid wheat, puroindoline alleles from *Aegilops tauschii* produce an endosperm softer than the one of soft common wheat (Gedye et al. 2004). A decrease in kernel hardness in substitution lines carrying additional copies of puroindoline genes from *Triticum monococcum* was found by Tranquilli et al. (2002), who also demonstrated that the simultaneous deletion of both puroindoline loci in common wheat increased its hardness to the level typical of durum wheat. They also showed that *Gsp-1* genes do not have a critical role on grain texture. There is also the possibility that new alleles of the puroindoline genes will be discovered, increasing the amount of tools available to modulate the hardness trait. On the other hand, it is well known that a part of the variation in kernel texture is not caused by the puroindoline genes.

As previously discussed, QTLs affecting grain textures were found on different chromosomes analysing recombinant inbred, double haploid lines (Sourdille et al. 1996; Campbell et al. 1999; Perretant et al. 2000; Galande et al. 2001; Igrejas et al. 2002) and backcross lines (Narasimhamoorthy et al. 2006). There is the possibility that even more QTLs and markers will be discovered using a novel approach that allows the analysis of lines deriving from different crosses, as happens in conventional breeding programmes (Crepieux et al. 2005; Arbelbide and Bernardo 2006).

The information already available on QTLs, genes and markers related to grain hardness is going to increase in the near future offering the breeders a wealth of tools enabling them to modulate the hardness trait taking also into account its influence on milling and baking quality. It is also possible to extend the range of variation of the trait in both bread and durum wheat resulting in the possibility to obtain supersoft or superhard varieties with novel end uses.

6. FLOUR/ SEMOLINA COLOUR

Another important quality parameter of bakery and pasta products is colour. In fact, natural colour is an important feature of bread, noodle and durum wheat pasta. In particular, the presence of yellow colour in the wheat flour is detrimental for bakery products made with common wheat, while it is desirable for the production of yellow alkaline noodles and for the majority of the durum wheat-based products. Thus, breeders have to apply different selection criteria depending on the wheat quality class being selected. In this respect, durum wheat genotypes with kernels containing a high level of yellow pigment and a low level of undesired components producing darker colours are today being selected.

Lepage and Sims (1968) pointed out that the yellow colour in durum wheat extracts is attributable to carotenoids, mainly to the xanthophyll lutein and its fatty acid esters (*ca*. 85% free lutein, 10% lutein monoesters and 5% lutein diesters). Wildfeuer and Acker (1968) reported that carotene accounts for 1% of the total carotenoid content in durum wheat semolina. Hentschel et al. (2002), analysing eight durum wheat cvs., showed that the fraction of carotenoids (only luteins and small amounts of zeaxanthin) accounts for 30–50% of the total yellow pigment content so they argued that other compounds of still unknown structure contribute considerably to the colour of the grains. Conversely, in a similar study Fratianni et al. (2005) evidenced the prevailing role of carotenoids, with lutein being the main compound, followed by zeaxanthin and β-carotene, with only a small fraction attributable to other compounds.

High levels of natural yellow colour in wheat flour or in semolina and thus in the end-use products can be primarily reached by selection for the natural carotenoid pigments accumulated throughout the inner and outer kernel layers. However, yellow pigments in final products such as noodles and pasta can be substantially broken down, during the manufacturing process, by the action of lipoxygenases (LOX activity), as well as peroxidases and polyphenoloxidases, the latter being mainly responsible for the formation of undesired brown colour components (Porceddu 1995; Dexter and Marchylo 2000; Peña and Pfeiffer 2005). Polyphenol compounds are also responsible for the red colour present in the kernel outer layers of some bread and durum wheat cvs. (Himi et al. 2005), but usually these compounds are not present in flour/semolina. It has been reported that in some durum wheat genotypes, a low LOX activity is more relevant to obtain yellow pasta as compared to the pigment content of the kernels (Borrelli et al. 1999).

Many analytical techniques (AACC 2003b), mainly based on chemical pigment extraction (e.g. HPLC) or light reflectance, have been proposed to measure the yellow pigment content in whole meals, semolina and pasta; it is thus possible to choose the methodology to be used according to the aim, the level of precision and the speed required. Results obtained on durum wheat samples by means of an automatic reflectance instrument (Minolta Chroma Meter CR-200) based on the color-space system (CIE 1986), the HPLC procedure and the standard water-saturated butanol (WSB) method have recently been compared by Fratianni et al. (2005). The authors found highly significant correlations among the measurements obtained from semolina samples using the three different techniques, while the relationships between the yellow index (b* parameter) and the pigment content as determined by HPLC on the whole meals were strongly influenced by the sample characteristics. It was thus suggested that the reflectance determinations on whole meal samples, although fast and safe, provide only relative and not absolute values. This not withstanding, due to its relatively low-cost and easy management, the Minolta instrument is widely used by breeders to measure the yellow index (b*) of several hundreds of whole meal/semolina samples per day. Further, it is suitable to analyse small flour samples, thus allowing breeders to select for high yellow pigment in early segregating generations.

Notable sources of genetic variation suitable for increasing yellow colour in durum wheat kernels have been found, in addition to the primary gene pool, mainly in *Hordeum chilense* (Martin et al. 2000; Ballesteros et al. 2005) and especially *T. monococcum* which is characterised by high amount of luteins (Abdel-Aal et al. 2002). Recently released durum cvs. generally meet the yellow quality requirements and often can be used to produce semolina stocks suitable to enhance the semolina quality; this notwithstanding a wealth of important landmark durum cvs. that have been/are still largely used in breeding programmes, are characterised by low yellow pigment content, like cv. Creso in Italy and other cvs. derived from Altar 84. A high genetic variability for yellow pigment content in the whole meal has been found when analysing a wide collection including 325 durum wheat accessions representing the most important improved durum gene pools (Maccaferri et al., data not published).

Although several environmental factors have been shown to influence the yellow pigment content (Borrelli et al. 1999) and the presence of genotype by environment interaction, this quality trait can be considered highly heritable (Parker et al. 1998) with the heritability coefficient reaching values even equal to 0.90 in durum wheat; further, the genetic control of this traits is mainly due to additive effects (Nachit et al. 1995).

Studies on the genetic control of yellow pigment content in wheat, and especially in durum wheat, were scarce until a few years ago. In recent years, the development of DNA-based molecular markers has provided powerful tools to identify the genetic factors underpinning the variation of quantitatively inherited traits (Lee 1995; Tuberosa et al. 2002). Several QTL studies have now addressed the genetic dissection of yellow pigment content in wheat and some information are available on-line in the MAS-dedicated web site at http://maswheat.ucdais.edu. A detailed summary of the results so far obtained as to the QTLs for yellow pigment content in semolina/wheat flour and for lypoxygenase (LOX) activity affecting yellow colour and brightness of the end-use products (noodles and pasta) is reported in Table 1. Overall, it can be underlined that the natural variation present in the wheat elite germplasm (both tetraploid and hexaploid) is mainly controlled by a relatively low number of genetic factors: in durum wheat, the regions located in the groups 5, 6 and 7L appear to account for a large portion of variation across homoeologous chromosomes and populations. Only chromosome regions located in group 3 seem to play an important role in bread wheat while not influencing the yellow pigment content in durum wheat. Moreover, it is worth mentioning that QTLs for LOX activity, which have a significant impact on the end-product quality colour (yellowness and brightness), were co-locating with the corresponding enzyme-coding genes.

In principle, once QTLs for the trait of interest 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; Young 1999). However, to date only a few successful applications of MAS for the improvement of quantitative traits have been described (Hu et al. 1997; Ragot

Table 1. Summary of the major chromosome regions harbouring QTLs for (A) yellow pigment content in semolina/wheat flour and (B) for lypoxygenase (LOX) activity affecting yellow colour and brightness of the end-use products. QTLs most probably identified across homoeologous regions and/or different mapping populations have been listed within dashed horizontal lines

Chrom.arm	Mapping population [(1)]	Parent with plus allele	Markers/chromosome regions	R^2(%)
A				
1BS	K/UC	–	QTL interval: 10–30 cM from the chrom top	10
2BL	L/P	–	*Xgwm382-Xwmc167*	–
2D	CD87/K		*Xwmc25a-P44/M54-3* [(2)]	12
3AS	CD87/K	–	*XksuB8-Xwmc50-Xpsr598-P44/M61-4-P40/M61-3*	17
3AS	S/Y	Schomburgk	*Xbcd828*	13
3BS	C/H	–	*Xgwm285-P40/M42-2-P40/M44-2-Xcdo583*	–
3BS	S/T	–	*Xgwm285-P40/M42-2-P40/M44-2-Xcdo583*	20
4AS	J/C	–	*XPaccMcga-3*	9
4BL	USA/S	–	*Xgwm1084*	9
5AS	L/P	–	*Xgwm293*	–
5BS	L/P	–	*Xwmc149-Xgwm234-Xgwm443*	–
5AL	J/C	–	*Xbcd926*	10
5BL	S/T	–	*Xstm286-Xgwm499-P36/M40-2-P36/M40-1Xgwm639B*	12
5D	C/H	–	–	–
6A	CD87/K	–	–	13
6AS	K/UC		QTL peak at 30 cM from the chrom. top	19
6AL	L/P	–	*XDuPw167*	–
6AL	USA/S	–	*Xgwm1150*	7
7BS	USA/S	–	*Xgwm573-Xgwm1184-Xgwm3019*	13
7AL	Om2*/*T. dic.*	Omrabi5	*XmcaaEacg198*	6
7AL	S/Y	Schomburgk	*Xcdo347-Xwg232, plus AFLP marker: Xwua2,16, 18, 26, 35, 39, 56STS from AFLP Xwua26*	54
7AL	C/H	–	*Xfba349-Xcdo347-Xgwm344-128-Xpsr121-Xpsr680a-XksuH9c, plus 7 AFLP markers*	–
7AL	K/UC	Kofa	QTL peak interval at +170 cM from the chrom. top	40
7AL	L/P	–	*Xgwm332-Xgwm282-Xgwm344-Xbarc267*	
7AL	Om2*/*T. dic.*	Omrabi5	*Xgwm63e*	13
7BL	CD87/K	–	*Xpsr680*	10
7BL	K/UC	Kofa	QTL peak interval at +150 cM from the chrom. top	16
7BL	K/W	Kofa	*TdPsy1* + *SSR* and *DArT markers*	–
7BL	L/P	–	*Xbarc267-Xbarc1073-XDuPw398*	–

(Continued)

Table 1. (Continued)

Chrom.arm	Mapping population [1]	Parent with plus allele	Markers/chromosome regions	R^2(%)
7BL	Om2*/*T. dic.*	Omrabi5	*Xgwm344*	53
7EL	7A-7EL Ag1-23/ Aconchi// 3* UC1113	*Lophopyrum ponticum*	*Xwg420-Lr19-STSLr19-130-Xmwg2062-Xpsr148-Xpsr680-Xpsr121-Xpsr687-Y*	
$7H^{ch}$	Tritordeum	*Hordeum chilense*		–
$2H^{ch}$	Hordeum chilense			–
B				
4BS	J/C	Jennah Khetifa	Polymorphism at the *Lpx-B1* locus	–
4BS	K/UC	Kofa	QTL peak coincident with the *Lpx-B1*locus	–
2BS	J/C	–	*Xrz444*	–

– : data not rep orted.

[1]Mapping population

Bread wheat:

S/Y = Schomburgk/Yarralinka (150 F4:F5 lines); Parker et al. 1998; Parker and Langridge 2000

C/H = Cranbrook/Halberd (163 Double haploid – DH – lines); Mares and Campbell 2001

CD87/K = CD87/Katepwa (180 DHs); Mares and Campbell 2001

S/T = Sunco/Tasman (163 DHs Mares and Campbell 2001)

Durum wheat:

J/C = Jennah Khetifa/Cham1 (110 Recombinant inbred lines – RILs –); Hessler et al. 2002; Nachit et al. 2001

L/P = Latino x Primadur (RILs); Somma et al. 2004

USA/S = USA x Svevo (249 RILs); Jurman et al. 2004

Om2*/*T. dic.* = Omrabi5/*T. dicoccoides 600545*//Omrabi5 (114 Backcross inbred lines – BDL –); Elouafi et al. 2001

K/W = Kofa/W9262-D063 (155 DHs); Pozniak et al. 2006

K/UC = Kofa/UC1113 (RILs); Zhang et al. 2006

Genetic stocks:

7A-7EL Ag1-23/Aconchi//3*UC1113 (7^a-7EL Recombinant backcross lines); Zhang et al. 2005

$7H^{ch}$ (Addition lines); Alvarez et al. 1998

$2H^{ch}$ (*Hordeum chilense* lines) Atienza et al. 2004

[2]Chromosome region associated also to polyphenol oxidase (PPO) activity influencing noodle brightness in the S/T population.

et al. 2000) due mainly to weak associations (in terms of genetic distance) between markers and target QTLs, unpredictable QTL effects across different background and/or the high costs of MAS (Salvi et al. 2001; Koebner 2003; Peleman and van der Voort 2003). Certainly, a more encouraging picture for MAS emerges considering single-gene traits such as disease resistance (Bus et al. 2000; Witcombe and Hash 2000) or even major QTLs, accounting for a sizeable portion of phenotypic variation and validated throughout different elite genetic backgrounds. In this respect, yellow pigment content, thanks also to the abundance of polymorphic locus-specific SSR

markers developed by the private and public wheat community in the latest years (Somers et al. 2004; Song et al. 2005), appears as a very interesting candidate for MAS applied to breeding programmes, especially in durum wheat.

As an alternative to traditional QTL studies involving one or more mapping populations, association mapping (AM) studies involving a large number (at least 100–200) of possibly unrelated accessions provides the opportunity to uncover the most important QTLs regulating the variation for the trait of interest within the germplasm of a particular crop (Rafalski and Morgante 2004; Breseghello and Sorrells 2006). In simple terms, the primary objective of AM is to detect correlations between genotypes and phenotypes, based on linkage disequilibrium (LD), i.e. the non-random combination of alleles at two genetic loci. This approach appears to be highly suited for the study of the yellow pigment content, thanks to a) the high level of variation present in the wheat germplasm, b) the usually high heritability values of the trait, and c) the good knowledge of the genetic bases of the carotenoids and xanthophyll biosynthesis pathways. In the elite durum wheat germplasm accessions, the large extent of LD (at the cM level, within 5–10 cM on average, Maccaferri et al. 2005, 2006) suggests that whole-genome search for chromosome regions harbouring QTLs can already be attained by using a high number (several hundreds) of neutral molecular probes, already available in the public domain (e.g. SSR markers; http://wheat.pw.usda.gov/GG2/index.shtml).

Alternatively, as already explored especially in maize, single genes known to play a crucial role in the carotenoid pathway can be searched for single feature polymorphisms/marker haplotypes associated to variation in yellow pigment content. In this respect, the genes and enzymes included in the carotenoid biosynthesis pathway are well known (Cunningham and Gantt 1998), while this is only partially true for the regulation mechanisms (von Linting et al. 1997; Gallagher et al. 2004; Cervantes-Cervantes et al. 2006).

In maize, Wong et al. (2003) found major QTLs controlling the accumulation level of different carotenoids and suggested that the genetic variation at key precursors in a common biosynthetic pathway/regulatory region may have quantitative effects on more than one compound. Moreover, they evidenced co-locations between QTLs for carotenoid accumulation in maize kernels and strong candidate genes, namely phytoene synthase and carotene desaturase. Subsequently, exploiting panels of genetically diverse accessions suitable for AM, Palaisa et al. (2004) elucidated the pattern of diversity and LD at *Y1* gene by exploring a large sequence window of several hundreds of kb up- and down-stream of the phytoene synthase gene and showed the presence of significant associations with the trait and extensive as well as selective sweeps caused by selection.

The isolation of candidate genes from the carotenoid biosynthesis pathway have also been attempted in wheat: Cenci et al. (2004) isolated BAC clones (from the tetraploid wheat Langdon) containing the three major genes (phytoene synthase, *PSY*, phytoene desaturase, *PSD*, and zeta-carotene desaturase, *ZDS*) and mapped some of the isolated clones using deletion stocks: *PSY* mapped on group 5, *PDS* on group 4 and *ZDS* on group 2. Only recently, Pozniak et al. (2006) identified a second

copy of *PSY* in the durum wheat genome mapping in the 7AL/7BL homoeologous regions harbouring the major yellow pigment content QTL in wheat. These wheat sequences are good candidates for more in depth analyses of genetic diversity and LD analysis in the wheat germplasm.

The alternative approach of searching for co-localizations between known QTLs and putative candidate genes/sequences by exploiting the conserved rice-wheat synteny is also feasible (Francki et al. 2004). However, because in a first attempt this approach failed to identify the presence of a sequence related to wheat *PSY* in the hortologous region of rice, careful in evaluating the patterns of co-linearity at small interval levels and in identifying the putative candidate genes/sequences should be used.

7. CONCLUSIONS AND OUTLOOK

The development of lines combining the desired expression of qualitative traits, such as those herein discussed, and acceptable yield levels still represents a major challenge for all the breeders involved in wheat selection programmes.

The availability of molecular markers closely linked to the QTL/genes of interest will allow for the implementation of MAS not only to select the individuals carrying the favourable alleles at the target loci, but also to pyramid favourable QTL alleles from different sources. MAS offers the opportunity for improving the efficiency of the breeding process, particularly when phenotyping alone does not guarantee the identification of the desired genotype or, alternatively, when labourious and time-consuming analyses are required to perform the phenotypic evaluation. Moreover, quick DNA extraction protocols coupled with high-throughput genotyping based on the scoring of markers which do not require the use of gels (Salvi et al. 2001) will streamline MAS at a lower cost.

REFERENCES

AACC (2003a) 39–70A Near-infrared reflectance method for hardness determination in wheat; 55–31 Single-Kernel characterization system for wheat kernel texture. In: Approved methods of the American association of cereal chemists – 10th edition – Including 2001, 2002 and 2003 Supplements, St. Paul, MN

AACC (2003b) 14–10 Pekar color text (Slick Test). 14–30 Agtron color test for flour; 14–50 determination of pigments. In: Approved methods of the American association of cereal chemists – 10th edition – Including 2001, 2002 and 2003 Supplements, St. Paul, MN

Abdel-Aal ESM, Young JC, Wood PJ, Rabalski I, Hucl P, Falk D, Fregeau-Reid J (2002) Einkorn: a potential candidate for developing high lutein wheat. Cereal Chem 79:455–457

Ahmad M (2000) Molecular marker-assisted selection of HMW glutenin alleles related to wheat bread quality by PCR-generated DNA markers. Theor Appl Genet 101:892–896

Alvarez JB, Martin LM, Martin A (1998) Chromosomal localization of genes for carotenoid pigments using addition lines of *Hordeum chilense* in wheat. Plant Breed 117:287–289

Andersen JR, Lubberstedt T (2003) Functional markers in plants. Trends Plant Sci 8:554–560

Arbelbide M, Bernardo R (2006) Mixed-model QTL mapping for kernel hardness and dough strength in bread wheat. Theor Appl Genet 112:885–890

Atienza SG, Ramirez CM, Hernandez P, Martin A (2004) Chromosomal location of genes for carotenoid pigments in *Hordeum chilense*. Plant Breed 123:303–304

Avivi L (1978) High protein content in wild tetraploid *Triticum dicoccoides* Korn. In: Ramanujam S (ed) Proceedings of the 5th international wheat genetics symposium, New Delhi, India. Indian Society of Genetics and Plant Breeding (ISGPB), pp 372–380

Ballesteros J, Ramirez MC, Martinez C, Atienza SG, Martin A (2005) Registration of HT621, a high carotenoid content *Tritordeum* germplasm line. Crop Sci 45:2662–2663

Blanco A, De Giovanni C, Laddomada B, Sciancalepore A, Simeone R, Devos KM, Gale MD (1996) Quantitative trait loci influencing grain protein content in tetraploid wheat. Plant Breed 115: 310–316

Blanco A, Pasqualone A, Troccoli A, Di Fonzo N, Simeone R (2002) Detection of grain protein content QTLs across environments in tetraploid wheat. Plant Mol Biol 48:615–623

Blanco A, Simeone R, Gadaleta A (2006) Detection of QTLs for grain protein content in durum wheat. Theor Appl Genet 112:1195–1204

Boggini G, Pogna NE (1989) The breadmaking quality and storage protein composition of durum wheat. J Cereal Sci 9:131–138

Boggini G, Cattaneo M, Paganoni C, Vaccino P (2001) Genetic variation for waxy proteins and starch properties in Italian wheat germplasm. Euphytica 119:111–114

Borrelli GM, Troccoli A, Di Fonzo N, Fares C (1999) Durum wheat lipoxygenase activity and other quality parameters that affect pasta color. Cereal Chem 76:335–340

Breseghello F, Sorrells ME (2006) Association analysis as a strategy for improvement of quantitative traits in plants. Crop Sci 46:1323–1330

Briney A, Wilson R, Potter RH, Barclay I, Crosbie G, Appels R, Jones MGK (1998) A PCR-based marker for selection of starch and potential noodle quality in wheat. Mol Breed 4:427–433

Brouns F, Kettlitz B, Arrigoni E (2002) Resistant starch and "the butyrate revolution". Trends Food Sci Technol 13:251–261

Bus V, Ranatunga C, Gardiner S, Bassett H, Rikkerink E, Geibel M, Fischer C (2000) Marker assisted selection for pest and disease resistance in the New Zealand apple breeding programme. Acta Horticulturae 538:541–547

Butow BJ, Ma W, Gale KR, Cornish GB, Rampling L, Larroque O, Morell MK, Bekes F (2003) Molecular discrimination of Bx7 alleles demonstrates that a highly expressed high molecular weight glutenin allele has a major impact on wheat flour dough strength. Theor Appl Genet 107:1524–1532

Butow BJ, Gale KR, Ikea J, Juhász A, Bedö Z, Tamás L, Gianibelli MC (2004) Dissemination of the highly expressed Bx7 glutenin subunit (*GluB1al* allele) in wheat as revealed by novel PCR markers and RP-HPLC. Theor Appl Genet 109:1525–1535

Campbell KG, Bergman CJ, Gualberto DG, Anderson JA, Giroux MJ, Hareland G, Fulcher RG, Sorrels ME, Finney PL (1999) Quantitative trait loci associated with kernel traits in a Soft × Hard wheat cross. Crop Sci 39:1184–1195

Cantrell RG, Joppa LR (1991) Genetic analysis of quantitative traits in wild emmer (*Triticum turgidum* L. var. *dicoccoides*). Crop Sci 31:645–649

Cenci A, Somma S, Chantret N, Dubcovsky J, Blanco A (2004) PCR identification of durum wheat BAC clones containing genes for coding for carotenoid biosynthesis enzymes and their chromosome localization. Genome 47:911–917

Cervantes-Cervantes M, Gallagher CE, Zhu C, Wurtzel ET (2006) Maize cDNAs expressed in endosperm encode functional farnesyl diphosophate synthase with geranylgeranyl synthase activity. Plant Physiol 141:220–231

CIE (1986) Colorimetry. Publication 15.2, 2nd ed, Central Bureau of the Commission Internationale de l'Eclairage, Wien, Austria

Clarke B, Rahman S (2005) A microarray analysis of wheat grain hardness. Theor Appl Genet 110: 1259–1267

Cloutier S, Rampitsch C, Penner GA, Lukow OM (2001) Cloning and expression of a LMW-i glutenin gene. J Cereal Sci 33:143–154

Crepieux S, Lebreton C, Flament P, Charmet G (2005) Application of a new IBD-based QTL mapping method to common wheat breeding population: analysis of kernel hardness and dough strength. Theor Appl Genet 111:1409–1419

Cunningham FX, Gantt E (1998) Genes and enzymes of carotenoid biosynthesis in plants. Annu Rev Plant Physiol Plant Mol Biol 49:557–583

De Bustos A, Rubio P, Jouve N (2000) Molecular characterisation of the inactive allele of the gene *Glu*-A1 and the development of a set of ASPCR markers for HMW glutenins of wheat. Theor Appl Genet 100:1085–1094

De Bustos A, Rubio P, Soler C, Garcia P, Jouve N (2001) Marker assisted selection to improve HMW-glutenins in wheat. Euphytica 119:69–73

D'Egidio MG, Mariani BM, Nardi S, Novaro P, Cubadda R (1990) Chemical and technological variables and their relationships: a predictive value equation for pasta cooking quality. Cereal Chem 67:275–281

D'Ovidio R (1993) Single-seed PCR of LMW glutenin genes to distinguish between durum wheat cultivars with good and poor technological properties. Plant Mol Biol 22:1173–1176

D'Ovidio R, Anderson OD (1994) PCR analysis to distinguish between alleles of a member of a multigene family correlated with wheat bread-making quality. Theor Appl Genet 88:759–763

D'Ovidio R, Masci S (2004) The low-molecular weight glutenin subunits of wheat gluten. J Cereal Sci 39:321–339

D'Ovidio R, Porceddu E (1996) PCR-based assay for detecting 1B-genes for low molecular weight glutenin subunits related to gluten quality properties in durum wheat. Plant Breed 115:413–415

D'Ovidio R, Porceddu E, Lafiandra D (1994) PCR analysis of genes encoding allelic variants of high-molecular-weight glutenin subunits at the *Glu-D1* locus. Theor Appl Genet 88:175–180

D'Ovidio R, Masci S, Porceddu E (1995) Development of a set of oligonucleotide primers specific for genes at the *Glu-1* complex loci of wheat. Theor Appl Genet 91:189–194

D'Ovidio R, Masci S, Porceddu E, Kasarda DD (1997) Duplication of the Bx7 high-molecular-weight glutenin subunit gene in bread wheat (*Triticum aestivum* L.) cultivar Red River 68. Plant Breed 116:525–531

Dexter JE, Marchylo BA (2000) Recent trends in durum wheat milling and pasta processing: impact on durum wheat quality requirements. In: Abecassis J, Autran JC, Feillet P (eds) International workshop on durum wheat, semolina and pasta quality, Montpellier, France, November 27, (Colloques de l'INRA, Inst Natl Recherche Agronomique: Paris), pp 139–164

Distelfeld A, Uauy C, Olmos S, Schlatter AR, Dubcovsky J, Fahima T (2004) Microcolinearity between a 2-cM region encompassing the grain protein content locus Gpc-6B1 on wheat chromosome 6 and a 350-kb region on rice chromosome 2. Funct Integr Genom 4:59–66

Distelfeld A, Uauy C, Fahima T, Dubcovsky J (2006) Physical map of the wheat high-grain protein content gene Gpc-B1 and development of a high-throughput molecular marker. New Phytol 169: 753–763

Eagles HA, Cane K, Eastwood RF, Hollamby GJ, Kuchel H, Martin PJ, Cornish GB (2006) Contributions of glutenin and puroindoline genes to grain quality traits in southern Australian wheat breeding programs. Aust J Agric Res 57:179–186

Elouafi I, Nachit MM, Martin LM (2001) Identification of a microsatellite on chromosome 7B showing a strong linkage with yellow pigment in durum wheat (*Triticum turgidum* L. var. *durum*). Hereditas 135:255–261

Finney KF, Barrimore MA (1948) Loaf volume and protein content of hard winter and spring wheat. Cereal Chem 25:291–312

Francki M, Carter M, Ryan K, Hunter A, Bellgard M, Appels R (2004) Comparative organization of wheat homoeologous group 3S and 7L using wheat–rice synteny and identification of potential markers for genes controlling xanthophylls content in wheat. Funct Integr Genom 4:118–130

Fratianni A, Irano M, Panfili G, Acquistucci R (2005) Estimation of color of durum wheat. Comparison of WSB, HPLC, and reflectance colorimeter measurements. J Agric Food Chem 53:2373–2378

Galande AA, Tiwari R, Ammiraju JSS, Santra DK, Lagu MD, Rao VS, Gupta VS, Misra BK, Nagarajan S, Ranjekar PK (2001) Genetic analysis of kernel hardness in bread wheat using PCR-based markers. Theor Appl Genet 103:601–606

Gale KR (2005) Diagnostic DNA markers for quality traits in wheat. J Cereal Sci 41:181–192

Gallagher CE, Matthews PD, Li F, Wurtzel ET (2004) Gene duplication in the carotenoid biosynthetic pathway preceded evolution of the grasses. Plant Physiol 135:1776–1783

Gedye KR, Morris CF, Bettge AD (2004) Determination and evaluation of the sequence and textural effects of the puroindoline a and puroindoline b genes in a population of synthetic hexaploid wheat. Theor Appl Genet 109:1597–1603

Giroux MJ, Morris CF (1997) A glycine to serine change in puroindoline b is associated with wheat grain hardness and low levels of starch-surface friabilin. Theor Appl Genet 95:857–864

Grama A, Gerechter-Amiti ZK, Blum A, Rubenthaler GL (1984) Breeding bread wheat cultivars for high protein content by transfer of protein genes from *Triticum dicoccoides*. In: Cereal grain improvement. Int. Atomic Energy Agency, Series 681-E, Vienna, pp 145–153

Graybosch RA, Peterson CJ, Hansen LE, Rahman S, Hill AS, Skerritt JH (1998) Identification and characterisation of US wheat carrying null alleles at the *wx* loci. Cereal Chem 75:162–165

Greene FC, Anderson OD, Yip RE, Halford NG, Malpica Romero J-M, Shewry PR (1988) Analysis of possible quality related sequence variations in the 1D glutenin high-molecular-weight subunit genes of wheat. In: Miller TE, Koebner RMD (eds) Proceedings of the 7th international wheat genetics symposium. Bath Press, Bath, UK, pp 699–704

Greenwell P, Schofield JD (1986) A starch granule protein associated with endosperm softness in wheat. Cereal Chem 63:379–380

Harjit-Singh PM, Prasad M, Varshney RK, Roy JK, Baylan HS, Dhaliwal HS, Gupta PK (2001) STMS markers for grain protein content and their validation using near-isogenic lines in bread wheat. Plant Breed 120:273–278

Hentschel V, Kranl K, Hollmann J, Lindhauer MG, Bohn V, Bitsch R (2002) Spectrophotometric determination of yellow pigment content and evaluation of carotenoids by high-performance liquid chromatography in durum wheat grain. J Agric Food Chem 50:6663–6668

Hessler TG, Thomson MJ, Benscher D, Nachit MM, Sorrells ME (2002) Association of a lipoxygenase locus, *Lpx-B1*, with variation in lipoxygenase activity in durum wheat seeds. Crop Sci 42:1695–1700

Himi E, Nisar A, Noda K (2005) Color genes (R and Rc) for grain and coleoptile upregulate flavonoid biosynthesis genes in wheat. Genome 48:747–754

Howitt CA, Gale KR, Juhász A (2006) Diagnostic markers for auality. In: Wrigley CW, Békés F, Bushuk W (eds) Gliadin and glutenin: the unique balance of wheat quality. Am Assoc Cereal Chem, St. Paul, MN, USA, pp 243–280

Hu X, Ribaut JM, Gonzales de Leon D (1997) Development of PCR-based markers to facilitate large-scale screening in molecular maize breeding. Maize Genet Coop Newsl 71:61–62

Igrejas G, Leroy P, Charmet G, Gaborit T, Marion D, Branlard G (2002) Mapping QTLs for grain hardness and puroindoline content in wheat (*Triticum aestivum* L.). Theor Appl Genet 106:19–27

Ikeda TM, Nagamine T, Fukuoka H, Yano H (2002) Identification of new low-molecular-weight glutenin subunit genes in wheat. Theor Appl Genet 104:680–687

Ikeda TM, Araki E, Fujita Y, Yano H (2006) Characterization of low-molecular-weight glutenin subunit genes and their protein products in common wheat. Theor Appl Genet 112:327–334

Jackson EA, Holt LM, Payne PI (1983) Characterisation of high-molecular weight gliadin and low-molecular-weight glutenin subunits of wheat endosperm by two-dimensional electrophoresis and the chromosomal localisation of their controlling genes. Theor Appl Genet 66:29–37

James MG, Denyer K, Myers AM (2003) Starch synthesis in the cereal endosperm. Curr Opin Plant Biol 6:215–222

Joppa LR, Cantrell RG (1990) Chromosomal location of genes for grain protein content of wild tetraploid wheat. Crop Sci 30:1059–1064

Joppa LR, Hareland GA, Cantrell RG (1991) Quality characteristics of the Langdon durum-*dicoccoides* chromosome substitution lines. Crop Sci 31:1513–1517

Joppa LR, Du C, Hart GE, Hareland GA (1997) Mapping gene(s) for grain protein in tetraploid wheat (*Triticum turgidum* L.) using a population of recombinant inbred chromosome lines. Crop Sci 37:1586–1589

Juhász A, Gianibelli MC (2006) Low-molecular-weight glutenin subunits: insight into this abundant subunit group present in glutenin polymers. In: Wrigley CW, Békés F, Bushuk W (eds) Gliadin and glutenin: the unique balance of wheat quality. Am Assoc Cereal Chem, St. Paul, MN, USA, pp 171–212

Jurman I, Castelluccio MD, Wolf M, Olivieri A, De Ambrogio E, Morgante M (2004) Construction of an SSR based linkage map and associated QTLs in durum wheat. Proceedings of the XLVIII SIFV-SIGA Joint Meeting, Lecce, Italy, September 15–18, ISBN 88-900622-5-8 Oral communication abstract 405

Khan IA, Procunier JD, Humphreys DG, Tranquilli G, Schlatter AR, Marcucci-Poltri S, Frohberg R, Dubcovsky J (2000) Development of PCR based markers for a high grain protein content gene from *Triticum turgidum* ssp. *dicoccoides* transferred to bread wheat. Crop Sci 40:518–524

Koebner R (2003) MAS in cereals: green for maize, amber for rice, still red for wheat and barley. In: Marker assisted selection: a fast track to increase genetic gain in plant and animal breeding. Turin, Italy, October 17–18, FAO, Rome. http://www.fao.org/biotech/docs/Koebner.pdf

Lafiandra D, D'Ovidio R, Porceddu E, Margiotta B, Colaprico G (1993) New data supporting high *M*r glutenin subunit 5 as determinant of quality differences among the pairs 5+10 vs 2+12. J Cereal Sci 18:197–205

Lafiandra D, Tucci GF, Pavoni A, Turchetta T, Margiotta B (1997) PCR analysis of x- and y-type genes present at the complex *Glu-A1* locus in durum and bread wheat. Theor Appl Genet 94:235–240

Lee M (1995) DNA markers and plant breeding programs. Adv Agron 55:265–344

Lee MR, Swanson BG, Baik BK (2001) Influence of amylase content on properties of wheat starch and breadmaking quality of starch and gluten blends. Cereal Chem 78:701–706

Lei ZS, Gale KR, He ZH, Gianibelli MC, Larroque O, Xia XC, Butow BJ, Ma W (2006) Y-type gene specific markers for enhanced discrimination of high-molecular weight glutenin alleles at the *Glu-B1* locus in hexaploid wheat. J Cereal Sci 43:94–101

Lepage M, Sims RPA (1968) Carotenoids of wheat flour: their identification and composition. Cereal Chem 45:600–604

Lillemo M, Morris CF (2000) A leucine to proline mutation in puroindoline b is frequently present in hard wheat from Northern Europe. Theor Appl Genet 100:1100–1107

Long H, Wei YM, Yan ZH, Baum B, Nevo E, Zheng YL (2006) Classification of wheat low-molecular-weight glutenin subunit genes and its chromosome assignment by developing LMW-GS group-specific primers. Theor Appl Genet 111:1251–1259

Ma W, Zhang W, Gale KR (2003) Multiplex-PCR typing of high molecular weight glutenin alleles in wheat. Euphytica 134:51–60

Maccaferri M, Sanguineti MC, Noli E, Tuberosa R (2005) Population structure and long-range linkage disequilibrium in a durum wheat elite collection. Mol Breed 15:271–289

Maccaferri M, Sanguineti MC, Natoli V, Araus Ortega JL, Ben Salem M, Bort J, Chenenaoui C, De Ambrogio E, Garcia del Moral L, De Montis A, El-Ahmed A, Maalouf F, Machlab H, Moragues M, Motawaj J, Nachit M, Nserallah N, Ouabbou H, Royo C, Tuberosa R (2006) A panel of elite accessions of durum wheat (*Triticum durum* Desf.) suitable for association mapping studies. Plant Gen Res 4:79–85

Maghirang EB, Dowell FE (2003) Hardness measurements of bulk wheat by single-kernel visible and near-infrared reflectance spectroscopy. Cereal Chem 80:316–322

Marchylo BA, Lukow OM, Kruger JE (1992) Quantitative variation in high molecular weight glutenin subunit 7 in some Canadian wheat. J Cereal Sci 15:29–37

Marchylo BA, Dexter JE, Clarke JM, Ames N (1998) Effects of protein content on CWAD quality. In: Fowler DB, Geddes WE, Johnston AM, Preston KR (eds) Wheat protein production and marketing. Proceedings of the wheat protein symposium, Saskatoon, Canada, March. University Extension Press, University of Saskatchewan, Saskatoon, pp 53–62

Mares DJ, Campbell AW (2001) Mapping components of flour and noodle colour in Australian wheat. Aust J Agric Res 52:1297–1309

Martin A, Cabrera A, Hernandez P, Ramirez MC, Rubiales D, Ballesteros J (2000) Prospect for the use of *Hordeum chilense* in durum wheat breeding. In: Royo C, Nachit MM, Di Fonzo N, Araus JL

(eds) Proceedings of the seminar on "Durum wheat improvement in the Mediterranean region: new challenges", Zaragoza, Spain, April 12–14, Options Méditerranéennes, Serie A, Séminaires Méditerranéens, 40:111–115

Martin JM, Frohberg RC, Morris CF, Talbert LE, Giroux MJ (2001) Milling and bread baking traits associated with puroindoline sequence type in hard red spring wheat. Crop Sci 41:228–234

Massa AN, Morris CF, Gill BS (2004) Sequence diversity of puroindoline-a, puroindoline-b, and the grain softness protein genes in *Aegilops tauschii* coss. Crop Sci 44:1808–1816

Mattern PJ, Morris R, Schmidt JW, Johnson VA (1973) Locations of genes for kernel properties in the wheat variety 'Cheyenne' using chromosome substitution lines. In: Sears ER, Sears LMS (eds) Proceedings of the 4th international wheat genetics symposium, August 6–11, University of Missouri, Columbia, MO, pp 703–707

McLauchlan A, Ogbonnaya FC, Hollingsworth B, Carter M, Gale KR, Henry RJ, Holten TA, Morell MK, Rampling LR, Sharp PJ, Shariflou MR, Jones MGK, Appels R (2001) Development of robust-PCR-based DNA markers for each homeoallele of granule-bound starch synthase and their application in wheat breeding programs. Aust J Agric Res 52:1409–1416

Mesfin A, Frohberg RC, Anderson JA (1999) RFLP markers associated with high grain protein from *Triticum turgidum* L. var. *dicoccoides* introgressed into hard red spring wheat. Crop Sci 39:508–513

Miura H, Tanii S (1994) Endosperm starch properties in several wheat cultivars preferred for Japanese noodles. Euphytica 72:171–175

Miura H, Tanii S, Nakamura T, Watanabe N (1994) Genetic control of amylose content in wheat endosperm starch and differential effects of three *Wx* genes. Theor Appl Genet 89:276–280

Miura H, Araki E, Tarui S (1999) Amylose synthesis capacity of the three *Wx* genes of wheat cv Chinese spring. Euphytica 108:91–95

Monari AM, Simeone MC, Urbano M, Margotta B, Lafiandra D (2005) Molecular characterization of new waxy mutants identified in bread and durum wheat. Theor Appl Genet 110:1481–1489

Morris CF, Lillemo M, Simeone MC, Giroux MJ, Babb SL, Kidwell KK (2001) Prevalence of puroindoline grain hardness genotypes among historically significant North American spring and winter wheat. Crop Sci 41:218–228

Murai J, Taira T, Oht D (1999) Isolation and characterisation of the three *Waxy* genes encoding the granule-bound starch synthase in hexaploid wheat. Gene 234:71–79

Nachit MM, Baum M, Impiglia A, Ketata H (1995) Studies on some grain quality traits in durum wheat grown in Mediterranean environments. In: Di Fonzo N, Kaan F, Nachit M (eds) Proceedings of the seminar on "Durum wheat improvement in the Mediterranean region", Zaragoza, Spain, November 17–19, 1993, Options Méditerranéennes, Série A, Séminaires Méditerranéens, 22:181–187

Nachit MM, Elouafi I, Pagnotta MA, El Saleh A, Iacono E, Labhilili M, Asbati A, Azrak M, Hazzam H, Benscher D, Khairallah M, Ribaut JM, Tanzarella OA, Porceddu E, Sorrells ME (2001) Molecular linkage map for an intraspecific recombinant inbred population of durum wheat (*Triticum turgidum* L. var *durum*). Theor Appl Genet 102:177–186

Nakamura T, Yamamori M, Hirano H, Hidaka S (1993) Identification of three Wx proteins in wheat *(Triticum aestivum* L.). Biochem Genet 31:75–86

Nakamura T, Yamamori M, Hirano H, Hidaka S, Nagamine T (1995) Production of waxy (amylose-free) wheat. Mol Gen Genet 248:253–259

Nakamura T, Vrinten P, Saito M, Konda M (2002) Rapid classification of partial waxy wheat using PCR-based markers. Genome 45:1150–1156

Narasimhamoorthy B, Gill BS, Fritz AK, Nelson JC, Brown-Guedira GL (2006) Advanced backcross QTL analysis of a hard winter wheat × synthetic wheat population. Theor Appl Genet 112:787–796

Olmos S, Diestelfeld A, Chicaiza O, Schatter AR, Fahima T, Echenique V, Dubcovsky J (2003) Precise mapping of a locus affecting grain protein content in durum wheat. Theor Appl Genet 107:1243–1251

Palaisa K, Morgante M, Tingey S, Rafalski A (2004) Long-range patterns of diversity and linkage disequilibrium surrounding the maize *Y1* gene are indicative of an asymmetric selective sweep. Proc Natl Acad Sci USA 101:9885–9890

Parker GD, Langridge P (2000) Development of a STS marker linked to a major locus controlling flour colour in wheat (*Triticum aestivum* L.). Mol Breed 6:169–174

Parker GD, Chalmers KJ, Rathjen AJ, Langridge P (1998) Mapping loci associated with flour colour in wheat (*Triticum aestivum* L.). Theor Appl Genet 97:238–245

Payne PI (1987) Genetics of wheat storage proteins and the effect of allelic variation on break making quality. Ann Rev Plant Physiol 38:141–153

Payne PI, Jackson EA, Holt LM (1984) The association between γ-gliadin 45 and gluten strength in durum wheat varieties: a direct causal effect or the result of genetic linkage? J Cereal Sci 2:73–81

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

Peña RJ, Pfeiffer WH (2005) Breeding methodologies and strategies for durum wheat quality improvement. In: Royo C, Nachit MM, Di Fondo N, Araus JL, Pfeiffer WH, Slafer GA (eds) Durum wheat breeding. Current approaches and future strategies. Food Product Press, Binghamton, NY, pp 663–703

Perretant MR, Cadalen T, Charmet G, Sourdille P, Nicolas P, Boeuf C, Tixier MH, Branlard G, Bernard S, Bernard M (2000) QTL analysis of bread-making quality in wheat using a doubled haploid population. Theor Appl Genet 100:1167–1175

Pogna NE, Autran JC, Mellini F, Lafiandra D, Feillet P (1990) Chromosome 1B encoded gliadins and glutenin subunits in durum wheat: genetics and relationship to gluten strength. J Cereal Sci 11: 15–34

Porceddu E (1995) Durum wheat quality in the Mediterranean countries. In: Di Fonzo N, Kaan F, Nachit M (eds) Proceedings of the seminar on "Durum wheat quality in the Mediterranean region", Zaragoza, Spain, November 17–19, 1993, Options Méditerranéennes, Série A, Séminaires Méditerranéens, 22:11–21

Pozniak CJ, Suprayogi Y, Knoz RE, Clarke FR, Clarke JM (2006) Genetic mapping of a phytoene synthase gene family from durum wheat. Plant & animal genome XIV conference, January 14–18, San Diego, CA, Poster P350, p 189

Prasad M, Varshney RK, Kumar A, Balyan HS, Sharma PC, Edwards KJ, Singh H, Dhaliwal HS, Roy JK, Gupta PK (1999) A microsatellite marker associated with a QTL for grain protein content on chromosome arm 2DL of bread wheat. Theor Appl Genet 99:341–345

Radovanovic N, Cloutier S (2003) Gene-assisted selection for high molecular weight glutenin subunits in wheat doubled haploid breeding programs. Mol Breed 12:51–59

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

Ragot M, Gay G, Muller JP, Durovray J (2000) Efficient selection for adaptation to the environment through QTL mapping and manipulation in maize. In: Ribaut JM, Poland D (eds) Molecular approaches for the genetic improvement of cereals for stable production in water-limited environments. Workshop held at CIMMYT, El Batán, Mexico, June 21–25, 1999, pp 128–130

Ribaut JM, Hoisington D (1998) Marker-assisted selection: new tools and strategies. Trends Plant Sci 31:236–239

Salmanowicz BP, Moczulski M (2004) Multiplex polymerase chain reaction analysis of *Glu-1* high-molecular-mass glutenin genes from wheat by capillary electrophoresis with laser-induced fluorescence detection. J Chromatogr A 1032:313–318

Salvi S, Phillips RL, Tuberosa R (2001) Development of PCR-based assays for allelic discrimination in maize by using the 5′-nuclease procedure. Mol Breed 8:169–176

Schwarz G, Felsenstein FG, Wenzel G (2004) Development and validation of a PCR-based marker assay for negative selection of the HMW glutenin allele *Glu-B1-1d (Bx-6)* in wheat. Theor Appl Genet 109:1064–1069

Shariflou MR, Hassani M, Sharp PJ (2001) A PCR-based DNA marker for detection of mutant and normal alleles of the *Wx-D1* gene of wheat. Plant Breed 120:121–124

Shewry PR, Halford NG, Lafiandra D (2003) The genetics of wheat gluten proteins. Adv Genet 49: 111–184

Shewry PR, Halford NG, Lafiandra D (2006) The high-molecular-weight subunits of glutenin. In: Wrigley CW, Békés F, Bushuk W (eds) Gliadin and glutenin: the unique balance of wheat quality. Am Assoc Cereal Chem, St. Paul, MN, USA, pp 143–169

Shimbata T, Nakamura T, Vrinten P, Saito M, Yonemaru J, Seto Y, Yasuda H (2005) Mutations in wheat *starch synthase II* genes and PCR-based selection of a SGP-1 null line. Theor Appl Genet 111:1072–1079

Smith RL, Schweder ME, Barnett RD (1994) Identification of glutenin alleles in wheat and triticale using PCR-generated DNA markers. Crop Sci 34:1373–1378

Snape JW, Hyne V, Aitken K (1995) Targeting genes in wheat using marker-mediated approaches. In: Li ZS, Xin ZY (eds) Proceedings of the 8th international wheat genetics symposium, Beijing, July 20–25, 1993, China Agric Scientech Press, Beijing, China, pp 749–759

Somers DJ, Isaac P, Edwards K (2004a) A high-density wheat microsatellite consensus map for bread wheat (*Triticum aestivum* L.). Theor Appl Genet 109:1105–1114

Somma S, Cenci A, Mangini G, Blanco A (2004b) Detection of QTL for carotenoid pigment content in durum wheat. Proceedings of the XLVIII SIFV-SIGA joint meeting Lecce, Italy, September 15–18, ISBN 88-900622-5-8 Poster abstract H.03

Song QJ, Shi JR, Singh S, Fickus EW, Costa JM, Lewis J, Gill BS, Ward R, Cregan PB (2005) Development and mapping of microsatellite (SSR) markers in wheat. Theor Appl Genet 110:550–560

Sourdille P, Perretant MR, Charmet G, Leroy P, Gautier MF, Joudrier P, Nelson JC, Sorrells ME, Bernard M (1996) Linkage between RFLP markers and genes affecting kernel hardness in wheat. Theor Appl Genet 93:580–586

Sourdille P, Perretant MR, Charmet G, Cadalen T, Tixier MH, Joudrier P, Gautier MF, Branlard G, Bernard S, Boeuf C, Bernard M (1999) Detection of QTL for bread making quality in wheat using molecular markers. In: Scarascia Mugnozza GT, Porceddu E, Pagnotta MA (eds) Genetics and breeding for crop quality and resistance. Kluwer, Netherlands, pp 361–366

Symes KJ (1965) The inheritance of grain hardness in wheat as measured by the particle-size index. Aust J Agric Res 16:113–123

Tranquilli G, Heaton J, Chicaiza O, Dubcovsky J (2002) Substitutions and deletions of genes related to grain hardness in wheat and their effect on grain texture. Crop Sci 42:1812–1817

Tuberosa R, Salvi S, Sanguineti MC, Landi P, Maccaferri M, Giulani S, Conti S (2002) Mapping QTLs regulating morpho-physiological traits and yield: case studies, shortcomings and perspectives in drought-stressed maize. Ann Bot 89:941–963

Turner AS, Bradburne RP, Fish L, Snape JW (2004) New quantitative trait loci influencing grain texture and protein content in bread wheat. J Cereal Sci 40:51–60

Uauy C, Distelfeld A, Fahima T, Blechl A, Dubcovsky J (2006) A NAC gene regulating senescence improves grain protein, zinc, and iron content in wheat. Science 314:1298–1301

Urbano M, Margiotta B, Colaprico G, Lafiandra D (2002) Waxy protein in diploid, tetraploid and hexaploid wheat. Plant Breed 121:1–5

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

Vawser MJ, Cornish GB (2004) Over-expression of HMW glutenin subunit *Glu-B1* 7x in hexaploid wheat varieties (*Triticum aestivum* L.). Aust J Agric Res 55:577–588

Von Lintig J, Welsch R, Bonk M, Giuliano G, Batschauer A, Kleinig H (1997) Light-dependent regulation of carotenoid biosynthesis occurs at the level of phytoene synthase expression and is mediated by phytochrome in *Sinapis alba* and *Arabidopsis thaliana* seedlings. Plant J 12:625–634

Wildfeuer I, Acker L (1968) Uber die Bestimmung von Carotenoiden in Teigwaren und deren Rohstoffen. Mitt Geb Lebensm Hyg 59:392–400

Witcombe JR, Hash CT (2000) Resistance gene deployment strategies in cereal hybrids using marker-assisted selection: gene pyramiding, three-way hybrids, and synthetic parent populations. Euphytica 112:175–186

Wong JC, Lambert RJ, Wuertzel ET, Rocheford TR (2004) QTL and candidate genes phytoene synthase and zeta-carotene desaturase associated with the accumulation of carotenoids in maize. Theor Appl Genet 108:349–359

Yamamori M, Quynh NT (2000) Differential effects of Wx-A1, -B1 and -D1 protein deficiencies on apparent amylose content and starch pasting properties in common wheat. Theor Appl Genet 100:32–38

Yamamori M, Nakamura T, Endo TR, Nagamine T (1994) Waxy protein deficiency and chromosomal location of coding genes in common wheat. Theor Appl Genet 89:179–184

Yamamori M, Nakamura T, Kiribuchi-Otobe C (1998) Waxy protein alleles in common and emmer wheat germplasm. Misc Publ Natl Inst Agrobiol Resour 12:57–104

Yamamori M, Fujita S, Hayakawa K, Matsuki J, Yasui T (2000) Genetic elimination of a starch granule protein, SGP-1, of wheat generates on altered starch with apparent high amylase. Theor Appl Genet 101:21–29

Yamamori M, Kato M, Yui M, Kawasaki M (2006) Resistant starch and starch pasting properties of a starch synthase IIa-deficient wheat with apparent high amylose. Aust J Agric Res 57:531–535

Yoo SH, Jane JL (2002) Structural and physical characteristics of waxy and other wheat starches. Carbohydr Polym 49:297–305

Young ND (1999) A cautiously optimistic vision for marker-assisted breeding. Mol Breed 5:505–510

Zeng M, Morris CF, Batty IL, Wrigley CW (1997) Sources of variation for starch gelatinization, pasting and gelation properties in wheat. Cereal Chem 74:63–71

Zhang W, Gianibelli MC, Rampling L, Gale KR (2004) Characterisation and marker development for low molecular weight glutenin genes from *Glu-A3* alleles of bread wheat (*Triticum aestivum* L.). Theor Appl Genet 108:1409–1419

Zhang W, Lukaszewski A, Kolmer J, Soria MA, Goyal S, Dubcovsky J (2005) Molecular characterization of durum and common wheat recombinant lines carrying leaf rust resistance (*Lr19*) and yellow pigment (Y) genes from *Lophopyrum ponticum*. Theor Appl Genet 111:573–582

Zhang W, Chao S, Manthey F, Carrera A, Echenique V, Cervigni G, Helguera M, Dubcovsky J (2006) QTLs mapping for semolina and pasta colour in durum wheat. Plant & Animal Genome XIV conference, January 14–18, San Diego, CA, Poster P321, p 181

Zhao XL, Xia XC, He ZH, Gale KR, Lei ZS, Appels R, Ma W (2006) Characterisation of three low-molecular-weight *Glu-D3* subunit genes in common wheat. Theor Appl Genet 113:1247–1259

CHAPTER 3

MOLECULAR APPROACHES AND BREEDING STRATEGIES FOR DROUGHT TOLERANCE IN BARLEY

MICHAEL BAUM[1,*], MARIA VON KORFF[1], PEIGUO GUO[1,2], BERHANE LAKEW[1], ALADDIN HAMWIEH[1], SAMER LABABIDI[1], SRIPADA M. UDUPA[1], HAITHAM SAYED[1], WAFA CHOUMANE[1], STEFANIA GRANDO[1] AND SALVATORE CECCARELLI[1]

[1]*Biodiversity and integrated Gene Management Program, International Center for Agricultural Research in the Dry Areas, Aleppo, Syria*

[2]*College of Life Science, Guangzhou University, Guangzhou 510006, China*

Abstract: Barley genotypes, in particular landraces and wild species, represent an important source of variation for adaptive traits that may contribute to increase yield and yield stability under drought conditions, and that could be introgressed into improved varieties. Traits that have been investigated include physiological/biochemical and developmental/ morphological traits. Yield performance under drought is particularly a complex phenomenon, and plants exhibit a diverse range of genetically complex mechanisms for drought resistance. Quantitative trait loci (QTL) studies with and without *H. spontaneum* have shown that developmental genes, notably those involved in flowering time and plant stature show pleiotropic effects on abiotic stress tolerance and ultimately determine yield. Problems associated with the hybridization of *H. spontaneum* such as alleles with deleterious effects on field performance could be best addressed in the advanced backcross (AB-) QTL analysis. It was interesting to see that in AB-QTL populations like in balanced populations major QTL overshadowed minor QTL-alleles. Nevertheless, crosses with *H. spontaneum*, AB-QTL populations and association studies with *H. spontaneum* have also identified new alleles and genes that are related to abiotic stress tolerance. In order to identify genes that are related to drought tolerance microarrays analysis to monitor gene expression profiles for plants exposed to limited water environment is performed. Several studies with rapid dehydration treatment have shown that osmotic-stress-inducible genes could explain the response to drought stress in plants. Another development is the identification and use of nucleotide polymorphisms (SNP) in genes related to abiotic stress tolerance. An understanding of the combined function and expression of genes involved in various abiotic stresses, could help identify candidate genes underlying QTL of interest.

*Author for correspondence: m.baum@cgiar.org

R.K. Varshney and R. Tuberosa (eds.), Genomics Assisted Crop Improvement: Vol. 2: Genomics Applications in Crops, 51–79.

1. INTRODUCTION

Drought continues to be a challenge to agricultural scientists in general and plant breeders in particular, despite many decades of research (Blum, 1996). Drought, or more generally, limited water availability is the main factor limiting crop production. Although it reaches the front pages of the media only when it causes famine and death, drought is a permanent constraint to agricultural production in many developing, and an occasional cause of crop losses in developed, countries. The development, through breeding, of cultivars with higher harvestable yield under drought conditions would be a major breakthrough (Ceccarelli and Grando, 1996). However, the ability of some plants to produce a higher economic yield under drought than others is a very elusive trait from a genetic point of view. This is because the occurrence, severity, timing, and duration of drought vary from location to location and in the same location from year to year. Cultivars successful in one dry year may fail in another, or cultivars resistant to terminal drought may not be resistant to intermittent drought, or to drought occurring early in the season (Turner, 2002). To make matters worse, drought seldom occurs in isolation; it often interacts with other abiotic (particularly temperature extremes) and biotic (root diseases and nematodes being the most obvious examples) stresses (Ceccarelli et al. 2004). In addition, areas with a high risk of drought (and/or other abiotic stresses) generally have low-input agriculture (Cooper et al., 1987). Thus, even though breeding for drought tolerance is complex, it is worsened by its interactions with the other stresses.

2. APPROACHES TO IMPROVEMENT IN DROUGHT TOLERANCE

2.1. Breeding Approaches

Two contrasting philosophies have been used to breed crops for drought tolerance. The first uses selection under optimum growing conditions, and is based on the assumption that an increased yield potential will have a carryover effect when the improved cultivars are grown under less favorable conditions. The second uses direct selection in the presence of drought within the target environment, and can take two forms: (i) selection for physiological or developmental traits (analytical breeding), and (ii) direct selection for grain yield (empirical or pragmatic breeding).

The first philosophy has failed to produce convincing results. In barley, scientists consistently found a negative relationship between yield potential and yield under stress conditions (Van Oosterom et al., 1993). The circumstantial evidence of the absence of a spillover effect is that drought continues to affect negatively agricultural production worldwide, despite the spectacular increases in crop yields obtained through breeding under optimum conditions.

Breeding based on traits associated with drought tolerance has been very popular. Traits that have been investigated include physiological/biochemical and developmental/morphological traits. The physiological/biochemical traits include

osmotic adjustment, osmotic potential, water soluble carbohydrates, proline content, stomatal conductance, epidermal conductance, canopy temperature, relative water content, leaf turgor, abscisic acid content, transpiration efficiency, water use efficiency, carbon isotope discrimination, and retranslocation. The developmental/morphological traits include leaf emergence, early growth vigor, leaf area index, leaf waxiness, stomatal density, tiller development, flowering time, maturity-rate, cell membrane stability, cell wall rheology, and root characteristics.

In barley, scientists at ICARDA have found that the traits more consistently associated with higher grain yield under drought are growth habit, early growth vigor, earliness, plant height under drought, peduncle length, and short grain-filling duration (Acevedo and Ceccarelli, 1989). The analytical approach has been very useful in understanding which traits are associated with drought tolerance and why. The approach has however been less useful in developing new cultivars that show improved drought tolerance under field conditions. This is because, as mentioned earlier, under field conditions, drought varies in timing, intensity, and duration. Therefore, it is the interaction among traits that determines the overall crop response to the variable nature of the drought stress, rather than the expression of any specific trait (Ceccarelli et al., 1991). Although breeding for drought tolerance based on direct selection for grain yield in the target environment (empirical or pragmatic breeding) appears intuitively as the most obvious solution, it has faced the criticism because the chances of progresses appears slow and remote. Two of the major problems with selection in stress environments are precision of selection and existence of several target environments. Each environment is characterized by its own specific type of drought and generally a combination of stresses.

The issue of precision of selection under moisture stress conditions has been addressed by several papers, which have focused attention on the magnitude of heritability (Ceccarelli, 1994). A review of literature (Ceccarelli, 1996a) has shown the absence of a consistent relationship between grain yield (as a measure of the intensity of the stress) and the magnitude of heritability. Recently, vom Brocke et al. (2002) confirmed that in pearl millet, it is possible to detect genetic differences even under severe moisture stress conditions. Furthermore, considerable progress has been made in both experimental design and statistical analysis, which improve the estimate of experimental error (Singh et al., 2003). Therefore, it is possible to combine precision and relevance by conducting trials in the target environment even when it is a stress environment.

The second issue is how to deal with the multitude of target environments. This is intimately associated with broad and specific adaptation, which has been debated in plant breeding since the early 1920s and is still highly controversial. One of the causes of controversy is the definition of stress environment, which is different for different crops. For example, in a country like Syria, with a large spatial variability of rainfall, bread wheat, durum wheat, and barley are grown within short distances in progressively drier environments with some overlapping. Therefore, a stress environment for bread wheat is moderately favorable for durum wheat, and a stress

environment for durum wheat is moderately favorable for barley and lentil. At the drier end, barley is the only rainfed crop, while the other cereals are only grown under irrigation.

Another cause of controversy is the confusion between adaptation over time and adaptation over space, even though the distinction is of fundamental importance. Adaptation over time (also called stability or dependability) refers to the performance of a cultivar in a given location over several years. If the cultivar performs consistently better than a reference cultivar, it is said to be stable. Adaptation over space refers to the performance of a cultivar in several locations. If the cultivar performs consistently better than a reference cultivar, it is said to be widely adapted. It can be argued that wide adaptation over time (also defined as stability) is more important to farmers than wide adaptation over space. The latter is, for obvious reasons, the major concern of seed producers.

Breeding for specific adaptation is particularly important in crops predominantly grown in unfavorable conditions, because unfavorable environments tend to be more different from each other than favorable environments. Breeding for specific adaptation to unfavorable conditions is often considered an undesirable breeding objective, because it is usually associated with a reduction of potential yield under favorable conditions. This issue has to be considered in its social dimension and in relation to the difference between adaptation over space and adaptation over time.

2.2. Choice of Germplasm

Landraces are still the backbone of agricultural production in many developing countries, and for crops grown in stress environments. The reasons farmers prefer to grow only landraces or continue to grow landraces even after partial adoption of modern cultivars are not well documented. These include quality attributes, seed storability, and the ability of landraces to produce some yield even in difficult conditions where modern varieties are less reliable.

The value of landraces in barley (Grando et al., 2001) and several other crops (Brush, 1999) is well documented in Syria. The comparison between barley landraces and modern cultivars in different conditions, ranging from severe stress (low input and low rainfall) to moderately favorable conditions (use of inputs and high rainfall) has given several indications (Table 1).

- Landraces produce more yield than the modern cultivars under low-input and stress conditions.
- The superiority of landraces is not associated with mechanisms to escape drought stress, as shown by their heading date.
- Within landraces, there is considerable variation in grain yield under low-input and stress conditions, but all the landraces-derived lines produce some yield whereas some modern cultivars do not.
- Landraces are responsive to inputs and rainfall, and the yield potential of some lines is high, though not as high as modern cultivars.

Table 1. Grain yields (kg/ha) of pure barley lines derived from Syrian barley landraces and modern cultivars at three levels of drought stress[a] in northern Syria (Ceccarelli 1996b)

Environment	Landraces (N = 44)	Modern (N = 206)	Difference	p[b]
Stress	1038	591	447	<0.01
Intermediate	3105	3291	–186	ns
Non-stressed	4506	6153	–1647	<0.01

[a]As defined by average precipitation and soil fertility
[b]Based on t-tests for groups of unequal size; ns = not significant

- It is possible to find modern cultivars that produce yields almost as high as landraces under low-input and stress conditions, but the frequency is very low.

Selection conducted only under high input conditions is likely to miss most of the lines that would have performed well under low-input conditions (Table 1).

2.3. Decentralized Selection

Selection for specific adaptation to drought stress in barley has been implemented at ICARDA initially by expanding the evaluation of early segregating populations in dry sites and by modifying the breeding method from a classical pedigree method to a modified bulk (Ceccarelli, 1996b). The most important results of this choice have been the re-evaluation of the role of landraces in breeding for drought tolerance (described earlier), and the discovery of the importance of the wild progenitor of cultivated barley, *Hordeum vulgare* ssp. *spontaneum*, as a source of tolerance to extreme levels of drought. Gas exchange observations made at anthesis in a wet site showed that *H. spontaneum* has widely open stomata, higher net photosynthesis, and lower pre-dawn leaf water potential at this stage of development than cultivated barley. The ability of some accessions of *H. spontaneum* to tolerate extreme levels of drought stress was evident during the severe drought of 1987. Two lines of *H. spontaneum* were the only survivors in the breeding nurseries at a site that received only 176 mm rainfall (Grando et al., 2001).

These lines had some photosynthetic activity early in the morning, even though six times less than in the absence of stress. The stomata were open and the pre-dawn leaf water potential was negative. At the same time, the stomata of the black-seeded local landrace Arabi Aswad, considered by farmers to be very resistant to drought, were closed even though the pre-dawn leaf water potential was slightly higher than in *H. spontaneum*. By midday, the stomatal conductance of *H. spontaneum* decreased, the net photosynthesis became negative, while the stomatal conductance of Arabi Aswad was zero.

Decentralized selection, i.e., selection in the target environment, is mostly based on grain yield. Grain yield is considered as the integrated response of a number of physiological, morphological, and developmental attributes. These

attributes allow the particular genotype or population to perform better in the particular type of drought encountered in a specific site and year, and in a combination of that particular type of drought with other stresses. The genotypes or populations selected under this specific set of conditions are then tested for a second season in a number of sites, and if again successful, for a third season.

The process makes use of the large annual variability, which characterizes stress locations. Thus, in a short period such as three years, and in seven to ten sites, there is a high probability that a genotype or population is exposed to a number of different types of drought in terms of occurrence, timing, duration, and severity. Therefore, the process progressively accumulates favorable alleles for performance under various conditions (Figure 1).

The process relies heavily on grain yield, but we also consider other attributes such as early vigor, habit of growth, flowering time, plant height, tillering, and, when extreme stress conditions occur, leaf rolling and senescence. The process combines the characteristics of the empirical and analytical approaches, and results in a slow but steady increase in yield under a combination of stresses. Figure 1 shows an example of this steady increase using a number of lines developed between 1985 and 2000. The yield increased from less than 0.96 t/ha in Harmal to 1.4 t/ha in Arta. Arta, however, has the major handicap of becoming very short under drought stress. A reduction of plant height is undesirable because it forces farmers to harvest by hand, which is much more expensive than mechanical harvesting. The use of *H. spontaneum* and one of the tallest selections from the black-seeded landrace, Zanbaka, has produced nearly 1.5 t/ha yield with a plant height that is only marginally lower than that of Zanbaka.

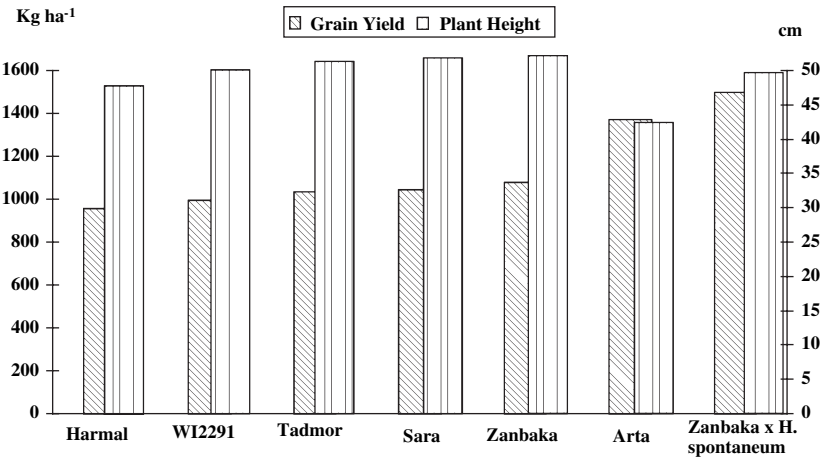


Figure 1. Grain yield and plant height of barely lines developed during the last 15 years (rainfall 197 mmm)

2.4. Conclusion

Barley breeding for stress environments is possible, provided it is conducted with strategies and methodologies that have little in common with those used in breeding for favorable environments. Adaptation over time can be improved by breeding for specific adaptation to a particular stress environment. This can be achieved by taking advantage of the temporal variability of stress environments, which permits exposure of the same breeding material to variable combinations of stresses over a (relatively) short period. We are aware that this is fundamentally different from the modern trend of plant breeding towards broad adaptation over space. The difference represents the contrasting interests of farmers and seed companies. Farmers are interested in cultivars that are consistently superior on their farm, regardless of how they perform at other locations or in other countries.

Recent advances in plant genomics have enabled the dissection of various molecular mechanisms involved in drought, cold and salt stress tolerance and in identifying various genes involved in these stress tolerance. Information generated in genomics should be integrated into practical plant breeding. Various genes identified in both model and crop plants could be used for developing stress tolerant plants through either marker-assisted selection or direct gene transfer.

3. MOLECULAR APPROACHES TO DROUGHT TOLERANCE

The advent of molecular markers has revolutionized the genetic analysis of crop plants and provided valuable new tools to identify chromosome regions influencing tolerance to abiotic stresses. Marker technologies and saturated marker maps allow the location of genomic regions or quantitative trait loci (QTL) with significant effects on drought tolerance or yield stability under adverse environmental conditions. The identification of these QTL with linked markers allows the breeders to use marker-assisted selection as a complementary tool to traditional selection. In addition, QTL mapping is a first step towards unraveling the molecular basis of drought resistance, i.e., by map-based cloning (Frary et al., 2000), with the ultimate goal of assigning functions to genes.

Yield performance under drought is particularly a complex phenomenon, because yield itself is a quantitative trait, and plants exhibit a diverse range of genetically complex mechanisms for drought resistance, including mechanisms of drought escape, drought avoidance, and drought tolerance (e.g., osmotic adjustment). This complexity may explain why most QTL studies of field-grown barley have so far been performed in favorable conditions, while only few studies have performed QTL analyses for agronomic performance under drought stress (Teulat et al., 2001b; Baum et al., 2003; Talamé et al., 2004; Forster et al., 2004; Long et al., 2001). These studies have shown that developmental genes, notably those involved in flowering time and plant stature have pleiotropic effects on abiotic stress resistance and ultimately determine yield potential.

In order to reduce complexity, physiological traits that have an influence on yield performance under drought were investigated in several studies on osmotic

adjustment, relative water content and carbon isotope discrimination conducted on field-grown barley (Teulat et al., 2002) and under controlled conditions (Teulat et al., 1998, 2001a).

In addition, the application of expression analyses and cloning techniques to the analysis of stress response has allowed the detection of a number of candidate genes whose expression is affected under stress (Cattivelli et al., 2002; Diab et al., 2004; Choi et al., 2000, 2002).

The following paragraphs focus on QTL studies for plant height, flowering time, and yield in relation to drought stress. Special emphasis is on studies using exotic barley germplasm, as wild barley being adapted to drought environments represents a promising resource for the improvement of drought tolerance. Therefore, recent association tests with wild barley grown in the fertile crescent (Ivandic et al., 2002, 2003) and QTL studies on advanced backcross populations tested in stressed and favorable conditions are discussed (Pillen et al., 2003, 2004; Matus et al., 2003; Li et al., 2005; Li et al., 2006; Talamé et al., 2004; von Korff et al., 2004, 2006; Hori et al., 2005). For balanced populations, only QTL analyses conducted under drought stress are presented (Teulat et al., 2001b; Baum et al., 2003; Talamé et al., 2004; Forster et al., 2004; Long et al., 2003). In addition, studies using physiological traits as well as recent advances in the discovery of genes involved in stress tolerance are introduced.

3.1. Chromosome Regions for Drought Tolerance and Developmental Traits in Balanced Populations

Agronomic performance in balanced populations was tested under drought prone conditions of West Asia, North Africa, and Australia. Baum et al. (2003) and Grando et al. (2005) used 194 RILs, randomly selected from a population of 494 RILs, and tested them in the 1996–97 and 1997–98 cropping seasons at ICARDA's research stations near Tel Hadya and Breda in Syria. Total rainfall in Tel Hadya in 1996–97 and 1997–98 was 433.7 and 410.5 mm, respectively, whereas in Breda it was 230.8 and 227.4 mm, respectively. The map extended over 890 cM and contained 189 marker loci, including one morphological marker locus (*btr* = brittle rachis), 158 AFLP loci and 30 SSR loci. Teulat et al. (2001a,b, 2003), Forster et al. (2004), and Diab et al. (2004) used a progeny of 167 two-row barley recombinant inbred lines (RILs), and the two parents Tadmor (selected by ICARDA from Arabi Aswad) and ER/Apm (an ICARDA breeding line released in Tunisia, Morocco and Libya), which were grown in three Mediterranean sites (Montpellier, France; Meknes, Morocco and Le Kef, Tunisia). The trial in Montpellier was conducted under two water treatments in 1999; 360.5 and 458 mm. For the two other sites, the plants were grown under rainfed conditions, in Meknes in 2000 (254.6 mm) and 2001 (267.4 mm) and in Le Kef in 2001 (282.8 mm). One hundred and thirty-three markers covered 1500 cM of the genome. Diab et al. (2004) added more candidate genes and ESTs to the available map. Ellis et al. (2002) and Forster et al. (2004) tested the Derkado/B83-12/21/5 doubled haploid population consisting of

156 doubled haploid two row barley lines in Meknes and Kef additionally at two sites (irrigated and non-irrigated) in Egypt. The cross was originally designed to investigate malting quality, yield and disease resistance traits, but has now been used to investigate over 60 traits including abiotic stress tolerance (Ellis et al., 2002). Derkado/B83-12/21/5 carry the semi-dwarf gene *sdw1* and ari-*e*.GP. One hundred and fifty-two markers covered 1426 cM of the eight linkage groups (3Ha, 3Hb, Ellis et al. 2002). Long et al. (2003) tested the 110 recombinant inbred lines of the cross Mundah/Keel at different sites in South Australia in 1999–2001. The cross was between a parent with superior adaptation on sandy soils low in fertility (Mundah) and parent with poor adaptation (Keel). Fifty-four markers (14 AFLPs, 12 RFLPs, 28 SSRs) were mapped in this population.

The developmental trait, heading date, in barley is controlled by three genetically independent mechanisms of the photothermal response, i.e., day length (photoperiod), plant response to a sum of temperature over a period (earliness *per se*), and response to a period of low temperature at the initial stages of plant development (vernalization) (Ellis et al., 1990). The major genes for photoperiodic response are the *Eam1* or *Ppd-H1* location on chromosome 2H (bin 4, Laurie et al., 1995) and *Ppd-H2* on chromosome 1H (bin 14) (Laurie et al., 1995). Earliness *per se* (*eps*) genes are minor genes, which influence onset of flowering independent of temperature or light. They have been identified as *eps2S* on 2H (bin 6), *eps3L* on 3H (bin 13), *eps4L* on 4HL, *eps5L* on 5H (bin 6), *eps6L.1* and *eps6L.2* on 6H (bin 7, 13), *eps7S* on 7H (bin 3) and *eps7L* on 7H (bin 12) (Laurie et al., 1995). *Vrn-H1* on chromosome 5H (bin 11), *Vrn-H2* on 4H (bin 13), and *Vrn-H3* on 1H (bin 13) determine the vernalization requirement in barley (Laurie et al, 1995; Dubcovsky et al., 1998, Figure 2).

For days to heading a few major chromosomal locations have been repeatedly identified in the above mentioned populations (Figure 2). QTL have been identified on 7H (bin 2-3) (Baum et al. 2003; Long et al., 2003), 3H (bin 12) (Baum et al., 2003; Teulat et al., 2001), 2H (bin 3-7) (*Ppd-H1*, Baum et al., 2003) 2H (bin 11-13) (Baum et al., 2003; Teulat et al., 2001), and 5H (bin 6) (Baum et al., 2003, Figure 2). Under better rainfall conditions the locations on 3H (bin 12), 2H (bin 3-7) (*Ppd-H1*) and 2H (bin 13) were more important in the Arta/*H. spontaneum* 41-1 population whereas under drier conditions the 7H (bin 3) location was more important (Baum et al., 2003). At the 7H location, *H. spontaneum* 41-1 contributed the drought escape allele. In the Tadmor//ER/Apm population 2H (bin 13), 3H (bin 12), 5H (bin 11) and 7H (bin 9) are the important loci. Forster et al. (2004) pointed out that for the Tadmor//ER/Apm populations flowering time variation in the RILs under non-vernalization conditions may be due to the *Vrn-H2* locus on 4H, and this may account for a cluster of QTL in this region on chromosome 4H, even though no QTL for heading date was among the QTL. Tadmor, however, has vernalization requirements which might not always be met in North African sites.

One of the most important characters for the extreme drought conditions with annual rainfall less than 200 mm is the plant height (under drought stress). Here, alleles from *H. spontaneum* 41-1 become very useful to increase plant height. QTL

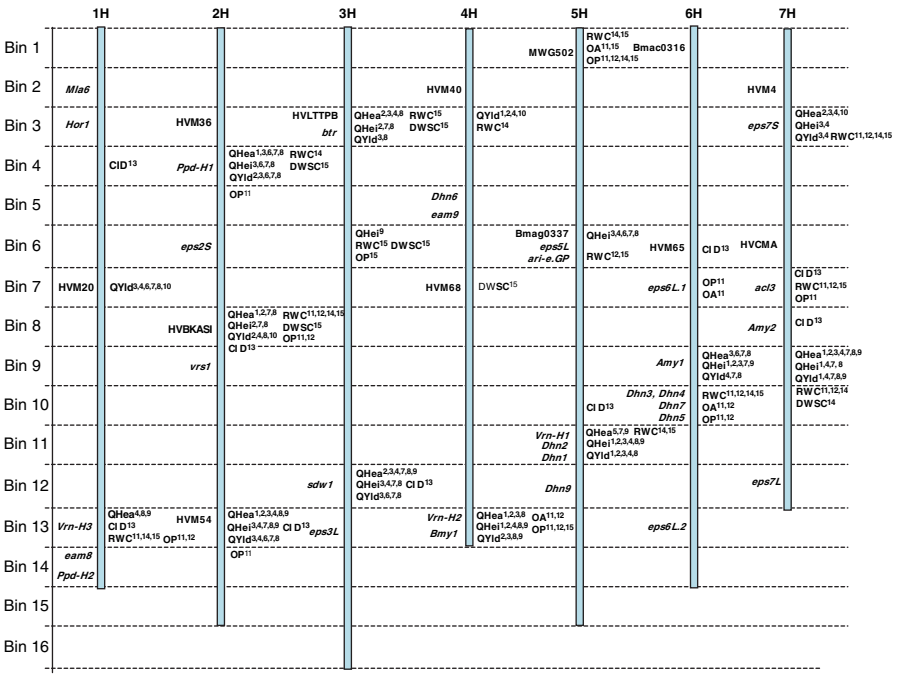


Figure 2. Genetic map of barley with locations for selected QTLs for agronomic and physiological traits. The genetic map is based on the Steptoe x Morex bin map by Kleinhofs and Graner (2001). Markers and loci have been assigned to the bin with the help of reference markers. Marker names are given on the left side of the chromosome bar. QTL for agronomic traits and physiological traits are presented on the right side of the bar. Selected QTL are shown for yield (QYld), heading date (QHea), plant height (QHei), physiological traits: relative water content (RWC), osmotic adjustement (OA), osmotic potential (OP), water-soluble carbohydrate concentration (WSC) and carbon isotope discrimination (CID).
References: 1 Pillen et al., (2003), 2 Pillen et al., (2004), 3 Baum et al., (2003), 4 Talame et al., (2004), 5 Hori et al., (2005), 6 Li et al., (2005), 7 Li et al., (2006), 8 von Korff et al., (2006), 9 Teulat et al., (2001b), 10 Long et al., (2003), 11 Teulat et al., (1998), 12 Teulat et al., (2001a), 13 Teulat et al., (2002), 14 Teulat et al., (2003), 15 Diab et al., (2004)

with major effects have been identified by Baum et al. (2003) on chromosomes 2H (bin 13), 3H (bin 12, *sdw*1), and 7H (bin 3). These might be more specific to *H. spontaneum* 41-1 background. In the Tadmor//ER/Apm population, major QTL for plant height were located on 2H (bin 13), 3H (bin 6), 4H (bin 13), 6H (bin 10) (Teulat et al., 2001). Plant height was not reported in the Mundah/Keel population. As expected, *sdw1* (3H) and *ari-e*.GP (5H) had major effects on plant height in the Drekado/B83-12/21/5 population. Both of them were also associated with most of the traits assessed, such as seedling establishment, plant vigour, peduncle length, thousand kernel weight and developmental rates between growth stages. Additionally, *ari-e*.GP (5H) was associated with shoot weight, root weight, shoot carbon isotope discrimination, and yield. *Sdw1* was associated with shoot carbon

isotope discrimination, root nitrogen isotope discrimination and grain nitrogen (Ellis et al., 2002; Forster et al., 2004).

In the Arta/*H. spontaneum* 41-1 population, QTL for grain yield and biological yield have been identified and located at the *btr* location on 3H (bin 3), and the *sdw1* (bin 12) locus in the part of the population with non-brittle lines. In both locations, pleiotropic effects for several other characters were identified. Other grain yield QTL were identified on 4H (bin 13), 5H (bin 11), and 7H (bin 3) (Baum et al., 2003). Teulat et al. (2001b) identified QTL for grain yield on chromosome 4H (bin 13); further on chromosome 7H (bin 9). In the Mundah/Keel population (Long et al., 2003) grain yield QTL were identified on chromosome 1H (bin 7), 2H (bin 8) and 4H (bin 3). In the Drekado/B83-12/21/5 population, *sdw1* was associated with grain yield in the three North African sites (Forster et al., 2004).

Cold and frost tolerance is a necessary requirement for barley production in the Fertile Crescent. In the Arta/*H. spontaneum* 41-1 population eight QTL for cold damage were identified. For the QTL with minor effects (2H, 4H and 6H), the allele from the *H. spontaneum* line showed the better protection. For all other QTL, the allele of Arta showed better ability to protect the plant from cold damage. Three of the four alleles with a high effect on cold damage were localized on chromosome 5H at the known locations for frost tolerance QTL (Pan et al., 1994; Reinheimer et al., 2004; Francia et al., 2004). Additionally, a QTL on 7H was identified. This is in agreement with the observation of Karsai et al. (2004), whose survey of H. *spontaneum* accessions of the Middle East origin showed some degree of vernalization response.

3.2. Genes Controlling Stress Tolerance

Cellular dehydration is common among the abiotic stresses and plants exhibit common molecular responses to them. Accumulation of a family of proteins known as dehydrins (*Dhns*) is one of the prominent components of this adaptive process (Close 1997). *Dhns* typically accumulate in different tissues in response to low temperature, drought, salinity, or abscisic acid (ABA) application and during the late phase of embryogenesis. Dehydrin genes have been mapped in the barley genome and some of them have been shown to co-segregate with stress tolerance QTL (Cattivelli et al., 2002). Three *Dhn* loci (*Dhn1/Dhn2* and *Dhn9*) are located in the same region of chromosome 5H where *Vrn-H1,* cold and salt tolerance QTL and ABA accumulation QTL have been mapped. Similarly, the *Dhn* cluster comprising *Dhn3*, *Dhn4*, *Dhn5* and *Dhn7* and the locus of *Dhn8/pAF93* on chromosome 6H are associated with drought tolerance QTL (Teulat et al., 2003). The recently identified *Dhn10* gene is located on 3H between saflp106 and ABG4 and *Dhn6* on chromosome 4H (Choi et al., 2000). Thirteen barley *Dhn* loci (Choi et al., 1999, 2000; Choi and Close 2000) have been identified and located on the different barley chromosomes.

Single-gene expression studies and transcriptome profiling in response to drought and cold revealed the induction of ‘effector’ genes responsive to a single stress

such as the barley Cor14b or Arabidopsis Rd22 and genes whose expression changed in response to both drought and cold (e.g. barley *Dhn* genes or Arabidopsis *Rd29a*). This suggests the existence of crosstalk between distinct stress-signaling pathways (Shinozaki et al., 2003) as quoted by Tondelli et al. (2006). Several transcription factors regulating stress inducible expression of the above mentioned effector genes via interaction with cis-acting elements in promoter regions have been isolated and characterized. The barley *HvCbf3* gene, ortholog of the A*rabidopsis CBF/DREB1* gene has been mapped to a position more proximal to the *Dhn1/Dhn2 – Dhn9* interval. Skinner et al. (2006) investigated and mapped homologs of the Arabidopsis low temperature regulatory genes CBFs to determine if there were candidates for winterhardiness-related quantitative trait loci (QTL). Twelve of the CBF genes were located on 5HL and 11 of them formed two tandem clusters. *AtFRY1* and *AtICE1* are regulators of CBF genes. The barley orthologs were mapped to chromosome 7H. TC147474 might be a positional candidate gene for the drought tolerance QTL located on chromosome 7H (Teulat et al., 2001). The *TC147474 (FRY1)* might co-locate with the QTL for kernel weight and earliness in the same population.

3.3. Putative Traits for Drought Tolerance: Physiological Traits

Breeding for drought tolerance based on traits associated with drought resistance, but easier to select for than grain yield, has been and still is very popular. Some physiological responses have been observed in plants induced by drought stress (Ludlow and Muchow, 1990).

Osmotic adjustment (OA) has been found to be one of the physiological mechanisms associated with plant tolerance to water deficit (Blum, 1988). Relative water content (RWC) is a measure of plant water status resulting from a cellular water deficit. RWC is an appropriate estimate of plant water status as affected by leaf water potential and OA. Carbon isotope discrimination is another method used to screen for drought tolerance. The correlation between water use efficiency and carbon isotope discrimination has been extensively studied in several crops including wheat (Farquhar and Richards 1984; Condon et al., 1987) and barley (Araus et al., 1997).

The chromosomal location of some traits related to drought tolerance is known. Differentially expressed sequence tags related to drought tolerance have been mapped in the Tadmor//ER/Apm RIL population by Diab et al. (2004). Osmotic potential, OP100 (is the osmotic potential under full turgor) and RWC are components of OA and they are highly correlated (Teulat et al., 1997). For that reason, OA is considered a drought-related trait (Teulat et al., 2001a). Several chromosomal regions related to variation in OA and water status were detected. Two QTL were identified for OA, one on chromosome 3H and one on 5H (Teulat et al., 2001; Diab et al., 2004) (Figure 1).

The genomic regions with QTL clusters for physiological traits were found on 1H (bin 13), 2H (bin 8), 5H (bin 1), 6H (bin 10) and 7H (bin 7) (see Figure 1). For example, QTL for RWC, OA and OP were all mapped to approximately the

same chromosomal location on 6H (Teulat et al., 1998, 2001a, 2003; Diab et al. 2004). This locus on 6H matches the location of the candidate genes *Dhn3*, *Dhn4*, *Dhn7*, *Dhn5*, which are expressed during drought and/or cold stress (Cattivelli et al.; 2002). Genomic regions with effects on OA, OP and RWC coincided with QTL for flowering time, height and yield as for example on 2H (bin 8) and 4H (bin 13) (Figure 1). The co-location of the QTL for DWSC100 (difference in water-soluble carbohydrate concentration between well-watered and stressed plants) and RWC in more than one genomic region, 2H (bin 4, bin 8) and 3H (bin 3), under water stress suggests that the accumulation of WSC (water-soluble carbohydrate concentration) may be important for plants to maintain their RWC. QTL for carbon isotope discrimination (CID) did also overlap with loci identified for RWC, OA and OP (1H, 2H, 7H). In addition, CID was located close to genomic regions with relevance for agronomic performance such as on 2H (bin 8) and 3H (bin 12).

3.4. Dormancy and Desiccation Tolerance

Seed dormancy in wild barley is an abiotic stress tolerance characteristic because in nature seed dormancy allows plants to escape drought and high temperatures in the warmer summer months (Snape et al., 2001). Drought, salinity, and freezing are stresses that lead to cellular dehydration through different mechanisms. The common component of water stress is evident in shared molecular responses to these stresses. Examples of such responses are common genes induced by all three types of stress and the importance played by the phytohormone abscisic acid (ABA). The coincidence of dormancy QTL with the abiotic stress tolerance QTL indicates that there might be a common physiological basis for dormancy (Zhang et al., 2005). One may infer that dormancy QTL are associated with those QTL that control abiotic stress tolerance. QTL have been identified on chromosomes 4, 5, and 7H in the Steptoe/Morex mapping population, chromosomes 1 and 5H in the Harrington/TR306 population, chromosomes 2H, 3H, and 5H in the Triumph/Morex population, and chromosomes 2H, 5H, and 7H in the Harrington/Morex population (Zhang et al., 2005). Zhang et al. (2005) identified QTL that are coincident with abiotic stress tolerance QTL in a cross with *H. spontaneum*: CD1H2 with a salt and drought tolerance QTL on chromosome 1H, CD2H with a drought tolerance QTL on chromosome 2H, CD4H with a salt tolerance QTL on chromosome 4H, CD5H1 and CD5H2 with a salt and a cold tolerance QTL, and CD7H3 with a drought tolerance QTL (Cattivelli et al., 2002).

3.5. Association Mapping for Drought Tolerance in Natural Populations

Wild barley from the Fertile Crescent shows a high phenotypic and genetic diversity and distinct changes in allele frequencies along climatic and geographic gradients. Ivandic et al. (2000) demonstrated that wild barley accessions sampled from a 100 m transect in Israel exhibited marked differences in response to water deficit. The same barley population with 52 different accessions was genotyped with 33

evenly distributed SSRs and phenotyped for developmental, morphological and yield-related traits under well watered and water stress conditions (Ivandic et al., 2003). The authors found significant associations for all studied traits, in particular at Bmac181 (4H) close to the dehydrin locus *Dhn6* and Bmac18 (6H) adjacent to the gene cluster *Dhn3*, *Dhn4*, *Dhn5*, *Dhn7*. In a previous association study, Ivandic et al. (2002) tested association of flowering time and adaptation to site of origin ecology and geography with 33 SSRs in 39 wild barley accessions from Israel, Turkey, and Iran. Flowering time was tested with and without vernalization and in long and short day. Significant associations for flowering were primarily detected close to known locations of earliness-per-se genes (*eps*) while none of the major photoperiod or vernalization response genes were detected in the wild barley germplasm. In addition, nine of the 14 significant associations with ecology and geography were also significant for flowering time, which underlines the importance of plant development in adaptation. The authors concluded that wild barley is characterised by different adaptation mechanisms and novel alleles as compared to cultivated barley. Association mapping in genetically diverse germplasm may thus form an effective strategy for detecting genomic regions and novel alleles from wild barley involved in adaptation to stressed environments.

4. ADVANCED BACKCROSS POPULATIONS

Wild barley has often been considered a promising resource for the improvement of agronomic and quality traits as well as stress tolerance. Ellis et al. (2000), for example, postulated that exotic barley being adapted to a wide range of environments offers the prospect of a goldmine of untapped genetic reserves. Nevo et al. (1992) demonstrated that wild barley harbours considerably more genetic variation than the cultivated species, and that many exotic alleles are associated with adaptation to specific environments with different abiotic stress conditions. As such, the challenge remains to identify allelic variation for these genes in exotic germplasm, with the most beneficial pleiotropic effects on abiotic stress resistance, and to introduce them into breeding programs. However, the improvement of agronomic performance with exotic germplasm is not straightforward, because wild barley harbours many alleles with deleterious effects on field performance.

With the advanced backcross quantitative trait locus (AB-QTL) mapping approach, Tanksley and Nelson (1996) developed a strategy, which allows a targeted transfer of favorable exotic alleles into elite breeding material. Through this approach, specific exotic alleles derived from the exotic donor are tagged with molecular markers and tested for association with agronomic traits. Since QTL detection is carried out in advanced backcross populations, problems associated with considerable phenotypic variation, linkage drag and epistasis in interspecific crosses are reduced. In addition, identified favourable QTL alleles can be rapidly transferred into QTL-NILs (near-isogenic lines) and used for breeding as well as for the analysis of gene function. In the following paragraphs, the performance of exotic alleles in advanced backcross populations is discussed for flowering time,

plant height and yield as relevant indicators for field performance under extreme and more favourable climatic conditions.

Several studies have employed the AB-QTL strategy to introgress exotic barley alleles into barley cultivars and examine agronomic performance, quality and disease tolerance (Pillen et al., 2003, 2004; Matus et al., 2003; Talamé et al., 2004; Forster et al., 2004; Li et al., 2005, 2006; Hori et al., 2005; von Korff et al., 2004, 2005, 2006; Yun et al., 2006, Figure 2). Whereas Pillen et al., (2003, 2004), Talamé et al., (2004), Forster et al., (2004), Li et al., (2005, 2006), von Korff et al., (2005, 2006) and Yun et al., (2006) concentrated on the analysis of phenotypic data from extensive field or greenhouse trials, Matus et al. (2003), Hori et al. (2004) and von Korff et al. (2004) focused more on the development of advanced backcross populations and detailed characterization of the genetic structures of these new genetic resources. Matus et al. (2003) developed recombinant chromosome substitution lines (140 BC2F6) derived from 'Harrington x Caesarea 24-26', analyzed them for agronomic, domestication-related traits and malting quality in California and conducted association test with 47 SSR markers. Hori et al. (2005) generated recombinant chromosome substitution lines, a BC3F1 (134 lines) and a derived DH (93 lines) population, from the cross Haruna Nijo x H602. The populations were phenotyped for seed dormancy and several agronomic traits in one environment and genotyped with 25 SSRs and 60 ESTs. Von Korff et al. (2004) developed the two BC2DH populations 'S42' from 'Scarlett x ISR42-8' (301 lines) and 'T42' from 'Thuringia x ISR42-8' (84 lines). The authors described the genomic architecture of two derived sets of candidate introgression lines as revealed with 98 and 65 SSRs.

Pillen et al. (2003, 2004) conducted the first thorough analysis of the agronomic performance of exotic barley germplasm. They genotyped two BC_2F_2 populations Apex x ISR101-23 (136 lines) and 'Harry x ISR101-23' (164 lines) with 45 and 50 SSRs, respectively. They field-tested them for agronomic traits and malting quality parameters in two consecutive years and at three different locations in Germany. A gronomic performance of the exotic germplasm under drought conditions was analyzed by Talamé et al. (2004). They examined a selected set of 123 DH lines derived from a BC_1F_2 Barke x HOR11508 in Italy, Morocco, and Tunisia (Forster et al., 2004). They studied the DH lines for agronomic traits and conducted a QTL analysis with 54 polymorphic AFLP markers and 59 SSRs.

Li et al. (2005) performed an AB-QTL analysis in 181 selected BC_3DH lines derived from the spring barley cultivar, Brenda, and the exotic accession, HS213. The population was tested in field trials in two consecutive seasons and at two different locations in Germany, and genotyped with 60 SSRs, revealing that only 61% of the lines carried a detectable introgression. Von Korff et al. (2005, 2006) phenotyped 301 BC_2DH lines of the population 'S42' for agronomic performance and disease resistance in two consecutive years and at four different locations in Germany. Li et al. (2006) tested 200 BC_3F_2 derived lines from the cross Brenda x HS584 at two locations over four years and conducted a QTL analysis with 108 SSRs. Yun et al. (2006) developed an AB-population ($BC_2F_{6:8}$) from the cross

Harrington x OUH602, phenotyped the population for seedling disease resistance, and conducted a QTL analysis with a set of 111 SSRs.

The exotic donors used in these studies were derived from Israel (Pillen et al., 2003, 2004; Li et al., 2005, 2006; Matus et al., 2003; von Korff et al., 2004), Greece (Talamé et al., 2004), and the Caspian Sea region (Hori et al., 2005). Their selection was primarily based on per se performance, origin, and passport data. QTL analyses, however, have shown that the phenotype of a plant is only a modest predictor of its genetic potential, especially with respect to quantitative traits (Tanksley et al., 1996). Accordingly, von Korff (2005) selected the donors based on agronomic performance of backcross progeny derived from crosses between ten barley cultivars and ten *H. vulgare* ssp. *spontaneum* accessions, rather than on per se performance of the wild barley accessions. Despite the rich collections of exotic barley, the challenge remains to develop strategies for effectively sampling novel and favourable exotic alleles and introducing them into elite germplasm.

The potential use of wild germplasm for the improvement of agronomic traits is still a matter of controversy. Matus et al. (2003) and Li et al. (2004) did not detect favourable exotic alleles for yield. They characterised the use of wild barley as disappointing from a plant breeding perspective, but suggested that exotic introgression libraries may be a valuable tool for gene discovery. In contrast, Pillen et al. (2003) found that at 34% of all QTL, the exotic allele improved agronomic performance. Similarly, von Korff et al. (2006) detected favorable exotic alleles at 36% of all QTL in the BC2DH population 'S42'.

Pillen et al. (2003) and von Korff et al. (2006) reported that the maximum average yield increase associated with an exotic QTL allele resulted in an average yield improvement of 7.7% and 7.1 %. Pillen et al. (2003) explained the lower favorable effect of exotic alleles on yield compared to the strong effect of exotic alleles in tomato (34%) (Fulton et al., 1997) and rice (17%, 18%) (Xiao et al. 1996, Xiao et al. 1998) with different breeding systems. In barley, homozygous backcross lines were analyzed (Pillen et al., 2003, 2004, Matus et al., 2003, Li et al., 2004). Fulton et al. (1997) and Xiao et al. (1998), however, compared hybrids in tomato and rice, where heterosis effects between elite and exotic germplasm may have contributed to the yield increase. Talamé et al. (2004), however, identified exotic QTL alleles in barley, causing an average yield increase of 17%, which is consistent with results in tomato and rice. This higher yield increase can possibly be explained by a selection against negative agronomic characters during the establishment of the AB-population. The exotic introgressions in AB-populations, however, are still relatively large, an average introgression size of 38.6 cM (Matus et al. 2003) and 35 cM (von Korff et al. 2004) was reported, and they contain thousands of genes. Therefore, selection against deleterious characters risks the elimination of linked alleles with a potentially beneficial effect. On the other hand, if major alleles with deleterious effects are removed, effects of minor alleles may be more easily detectable. Notorious examples here are the two complementary genes on chromosome 3H, *Brt1* and *Brt2* (Azhacuvel et al. 2006) with the dominant exotic

allele causing the formation of a brittle rachis with a strong negative effect on yield performance.

In addition, all these AB-QTL analyses reported a strong segregation of flowering time and plant stature in their respective populations. Wild barley is characterized by early and heterogeneous flowering and prostate growth behavior, which is presumably an adaptation strategy to drought-prone environments. Correspondingly, the majority of exotic alleles caused early flowering with the strongest effect consistently found at a QTL on 2HS (Pillen et al., 2003; Li et al., 2005, 2006; von Korff et al., 2006). This region harbors the photoperiod response gene *Ppd-H1*, which promotes early flowering under long day conditions (Laurie et al., 1995). At other loci, however, the exotic allele clearly delayed flowering, for example, at QTL close to the vernalization response genes *Vrn-H1* (Hori et al., 2005; Li et al., 2006), *Vrn-H3* (Talamé et al., 2004; von Korff et al., 2006) and *Vrn-H2* (Pillen et al., 2003, 2004; Korff et al., 2006). An effect of the exotic allele on heading date was also repeatedly found on 2H, bin 13 (Pillen et al., 2003, 2004; Talamé et al., 2004; von Korff et al., 2006), 3H, bin 3 close to the *btr* loci (Pillen et al., 2004; Talamé et al., 2004; von Korff et al., 2006), 3H, bin 12 and adjacent to the *sdw1* locus (Pillen et al., 2004; Talamé et al., 2004; von Korff et al., 2006; Li et al., 2006) and 7H, bin 9 (Pillen et al., 2003, 2004; Talamé et al., 2004; von Korff et al., 2006).

AB-populations also show a large variation for plant height. Talamé et al. (2004) found a maximum variation in plant height of between 88 and 144 cm in Morocco. Von Korff et al. (2006) reported an average plant height in the population 'S42' across eight environments of between 63 cm and 110 cm. Major plant height QTL were located on 2H, bin 8 (Pillen et al., 2004; von Korff et al., 2006; Li et al., 2006), 3H, bin 12 (Talamé et al., 2004; von Korff et al., 2006), 4H, bin 13 (Pillen et al., 2003, 2004; Talamé et al., 2004; von Korff et al., 2006) and 5H, bin 6 (Talamé et al., 2004; Li et al., 2005; von Korff et al., 2006; Li et al., 2006). Corresponding candidate genes are the semi-dwarf genes *sdw3*$_{[2H]}$ (Gottwald et al., 2004), *sdw1*$_{[3H]}$, and *ari-e.GP*$_{[5H]}$, and the flowering loci *Ppd-H1*$_{[2H]}$ and *Vrn-H2*$_{[4H]}$. At the majority of QTL-loci, the exotic allele increased plant height, in particular at QTL close to the candidate genes *sdw1* and *ari-e.GP*, but at the QTL on 2HS and 4HL the exotic allele consistently reduced plant height.

Under drought conditions, heading is negatively and plant height positively correlated with yield, indicating that tall early heading genotypes present a good yielding capacity under water limiting conditions. Indeed, the strongest favorable effects of the exotic germplasm on yield were found under drought conditions in Tunisia and Morocco (Talamé et al., 2004). The same authors, however, observed that the exotic alleles with a delay in flowering time showed a favorable effect on yield, indicating that the favorable effect on yield under conditions of limited water was not due to drought escape but to an increase in yield potential. Similarly, in the AB-population 'S42' the exotic introgression on 4HL showed a favorable effect on yield under drought conditions, although this QTL-allele postponed flowering (Hamam et al., in prep).

In AB-QTL studies, major QTL loci often showed pleiotropic effects on a number of different traits and resulted in a strong clustering of QTL in particular on 2HS, 3HL, 4HL, and 5HL. The exotic introgression at the semi-dwarf locus *sdw1*, for example, affected, next to flowering time, plant height, yield and thousand grain weight (Pillen et al., 2003; Li et al., 2005; von Korff et al., 2006; Li et al., 2006, Figure 2). Similarly, QTL close to the *Vrn-H2* locus influenced heading data, plant height, and yield (Pillen et al., 2003; Talamé et al., 2004; von Korff et al., 2006). Although the donor and recipient germplasm differed between the cited AB-QTL studies, the exotic alleles exhibited predominantly the same qualitative effect at these major QTL for heading date, plant height, and yield. The exotic alleles are thus often similar in their effects and clearly different from the elite alleles. Wild barley thus harbors novel genetic variability for these key loci. The AB-QTL strategy allows the selection of major genes/alleles from the exotic gene pool with the most beneficial pleiotropic effects, especially in stress environments, and introduces these into breeding programs while eliminating negative alleles such as brittleness.

Eshed and Zamir (1994) demonstrated that introgression lines in tomato are a powerful tool for map-based cloning (Frary et al., 2000) and the discovery of gene function by transcriptome and metabolome analysis (Schauer et al., 2006). In barley, advanced backcross populations enable the fast generation of such introgression lines as demonstrated by von Korff et al. (2004) and Hori et al. (2005). Von Korff et al. (2004) selected from each of the BC2DH populations 'S42' and 'T42' and Hori et al. (2005) from the BC_3F_1 population Haruna Nijo x H602, a minimal set with 49, 43, and 19 introgression lines, respectively, which cover a large percentage of the exotic genome in overlapping exotic segments. Further backcrossing and establishment of nearly isogenic lines generate a valuable resource for validating the effects of exotic QTL-alleles, introducing them into elite cultivars, map-based cloning of verified QTL, and ultimately for the study of gene function.

5. IDENTIFICATION OF FUNCTIONAL GENES RESISTANT TO DROUGHT BY MICROARRAY

A number of genes have been described that respond to drought at the transcriptional level (Bray 1993; Seki et al., 2002; Cheong et al., 2003; Liu and Baird, 2004). The functions of some genes have been predicted from sequence homology with known proteins and are thought to have a role in protecting the plant from water deficit. Some important stress-inducible genes have been used to improve the stress tolerance of plants by gene transfer (Xu et al., 1996; Pellegrineschi et al., 2002, 2004; Abebe et al., 2003; Umezawa et al., 2006). Although a large number of drought-induced genes have been identified in a wide range of plant species, a molecular basis for plant tolerance to water deficit remains far from being completely understood (Ingram and Bartels, 1996; Bruce et al., 2002).

There is widespread interest in using microarrays to monitor gene expression profiles for plants exposed to limited water environment. Several studies with rapid dehydration treatment have shown that osmotic-stress-inducible genes could explain

the response to drought stress in plants (Seki et al., 2001; Ozturk et al., 2002; Seki et al., 2002; Shinozaki et al., 2003; Talamé et al., 2003; Ueda et al., 2004; Zhang et al., 2004). It is also thought that these gene products may play important roles in the adaptation of plants.

It has become evident that multiple stress responses are necessary for plants to endure severe stress conditions (Nakashima and Yamaguchi-Shinozaki, 2006). It is possible for a single transcription factor to control the expression of many target genes through the specific binding of the transcription factor to *cis*-acting element in the promoters of their respective target genes. Nakashima and Yamaguchi-Shinozaki (2006) described the expression mechanisms of osmotic stress-responsive genes for the presence of multiple regulons in Arabidopsis. The dehydration-responsive element-binding protein (DREB1)/C-repeat (CRT)-binding factor (CBF) regulon is revolved in osmotic stress-responsive gene expression, and the other regulons including the DREB2 regulon are involved in osmotic stress-responsive gene expression. It is expected that biotechnological efforts, which strive to control gene expression within these regulons, will improve tolerance against multiple stresses in plants.

5.1. Responsive Genes to Osmotic-/Drought-Stress in Barley Under Seedling Stage

Compared with the model plant, barley, a major crop plant, is one of few rainfed crops in the dry areas. Thus, it is a good target for the study of the mechanisms of drought tolerance in crop plants. However, until now, there are only few reports on the use of barley cDNA micro-array for transcriptional analysis under dehydration (Ozturk et al., 2002) and osmotic stress (Ueda et al., 2004) conditions. Response to dehydration in barley (*Hordeum vulgare* L. cv. Tokak) was monitored by micro-array of 1463 DNA elements (Ozturk et al., 2002). Transcripts that showed significant up-regulation under dehydration stress were exemplified by jasmonate-responsive, metallothionein-like, late-embryogenesis-abundant (LEA) and ABA-responsive proteins. The most drastic down-regulation in a category was observed for photosynthesis-related functions. Under osmotic stress, Ueda et al. (2004) found that 52 genes were differentially expressed in a total of 460 non-redundant cDNA microarray, 22 of which were up-regulated, such as sucrose synthase, GAPDH, alanine aminotransferase, lipoxygenase, and PMP3. Thirty genes, including sugar transporter, HSP70, serine/threonine kinase, PEPC, actin, water channel 2, etc., were down-regulated. Although information regarding expression profiles in response to the development of abrupt and severe water stress is useful, studies that impose slower, more realistic rates of dehydration are needed.

In an attempt to meet this need, Talamé et al. (2003) compared gene expression patterns between several water stress treatments in young barley plants. From the tissues of treated and control plants, mRNA was extracted, labeled, and hybridized to micro-arrays containing 1,463 DNA elements derived from barley cDNA libraries (Ozturk et al., 2002). Micro-array analysis with mRNA extracted from water

shock and control tissues largely confirmed the changes in the expression level of transcripts as reported earlier by Ozturk et al. (2002). Nearly 15% of all transcripts were either up- or down-regulated under conditions of rapid drought stress. Transcripts that showed significant up-regulation included jasmonate-responsive, metallothionein-like, late-embryogenesis-abundant (LEA), and ABA-responsive proteins. In the second experiment, water-stressed plants showed significant changes in transcript abundance after the seventh day of water deficit when the relative water content was ca. 90–93%. The greatest change in expression was observed after eleven days of treatment. As in the water shock experiment, expression profiles were analyzed by cluster analysis. It was shown that the majority of treatment effects were due to transcript changes in genes whose putative function involved protein synthesis, and turnover early in the experiment followed by later induction of known stress-responsive transcripts.

5.2. Transcriptional Analysis Between Drought-Tolerance and Sensitive Genotypes at the Reproductive Stage

The grain yield of barley is likely to be reduced a little if drought stress occurs during the vegetative period. However, it is vulnerable to drought stress during the reproductive growth stage, which is usually characterized by terminal drought stress (Ceccarelli et al., 2004). Understanding the molecular mechanism of drought tolerance during the reproductive growth stage may facilitate identification of novel drought tolerance genes for genetic improvement of barley cultivars. Meanwhile, the different genotypes contrasting in drought tolerance should have different mechanisms in response to drought stress. Therefore, it is meaningful to investigate and identify the genes that are differentially expressed between drought-tolerant and -sensitive genotypes under water deficit conditions at the reproductive stage, and to gain insight into the processes involved in stress responses by analyzing their expression profiles.

Exploration for differentially expressed genes responsive to drought has been carried out at the International Center for Agricultural Research in the Dry Areas (ICARDA) using two barley genotypes of Tadmor and WI2291 as experimental material (Guo et al., 2005). A greenhouse experiment was conducted in an ICARDA temperature-controlled greenhouse with 16 hours daylight at 30°C and 8 hours dark period at 20°C with pots to estimate drought tolerance of two barley genotypes. Drought stress was imposed at the heading stage by reducing available water in the soil from 70% (non-stress condition) to 10% (severe drought stress). Two physiological traits chlorophyll content and Fv/Fm (maximal fluorescence (F_m)/variable fluorescence (F_v)) and grain yield were measured Tadmor showed significantly less yield loss than WI2291 under drought stresses, and there was almost no difference in the reduction of grain yield between moderate and severe drought stress in Tadmor, but there were pronounced differences in WI2291 (Figure 3). The ratio of F_v/F_m indicates the potential photochemical yield of PS II and quantum efficiency, and it is an important index for evaluating the photosynthesis efficiency (Figure 3).

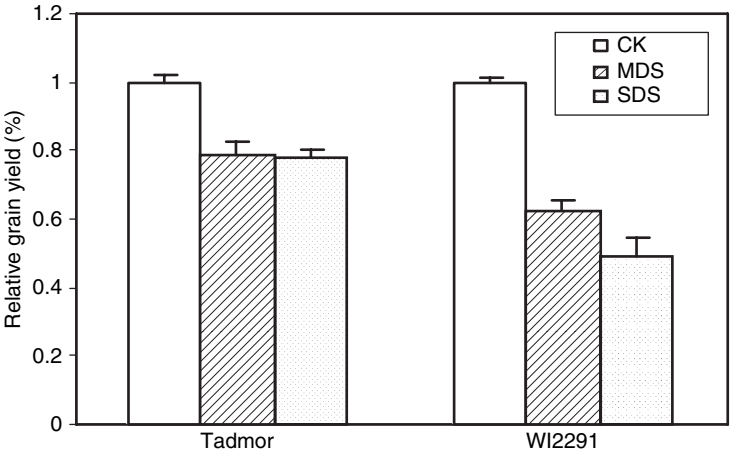


Figure 3. Grain yield of Tadmor and WI2291 under the three moisture regimes in the soil (means ± SE)
- CK = well-watered conditions, 70% available water in the soil
- MDS = moderate drought stress, 35% available water in the soil
- SDS = severe drought stress, 10% available water in the soil

In this study, the chlorophyll content and the ratio of F_v/F_m slightly decreased with time and almost remained constant in control leaves in both varieties through the 13-day period after treatment (Figure 4). However, the chlorophyll content and the ratio of F_v/F_m obviously decreased after 7 and 9 days in severe drought stress (SDS), 9 and 11 days in moderate drought stress (MDS) for Tadmor and after 5 and 7 days in SDS, 7 and 7 days in MDS for WI2291 after imposing water deficit. The chlorophyll content and the ratio of F_v/F_m decreased more in WI2291 than in Tadmor. These results confirm the field observation that Tadmor is more tolerant to water deficit stress than WI2291.

For gene expression analysis through 22K of Affymetrix GeneChip Barley 1 arrays, total RNAs were isolated from the flag leaves of Tadmor and WI2291 at the 3rd day of severe drought stress and non-stress treatment at post-anthesis stage. Sample processing, hybridization, and scanning of Affymetrix Barley Genome Arrays were performed through the standard Affymetrix protocol (Affymetrix GeneChip® Expression Analysis Technical Manual). Gene ontology analysis was conducted by HarvEST Barley 1.34 and GO functional categorization at TAIR (The Arabidopsis Information Resource). The results showed that some functional genes in both genotypes have similar response to water deficit stress in gene expression profiles, 77 of those genes were found to be significantly differentially expressed in Tadmor and WI2291 after drought stress. These genes may only be common genes responsive to drought stress since they appeared in both the tolerant and sensitive genotypes. The gene expression profiles were further analyzed for exploration of specific genes in response to drought stress through the comparison of gene expression profiles to drought stress between Tadmor and WI2291, 372 genes

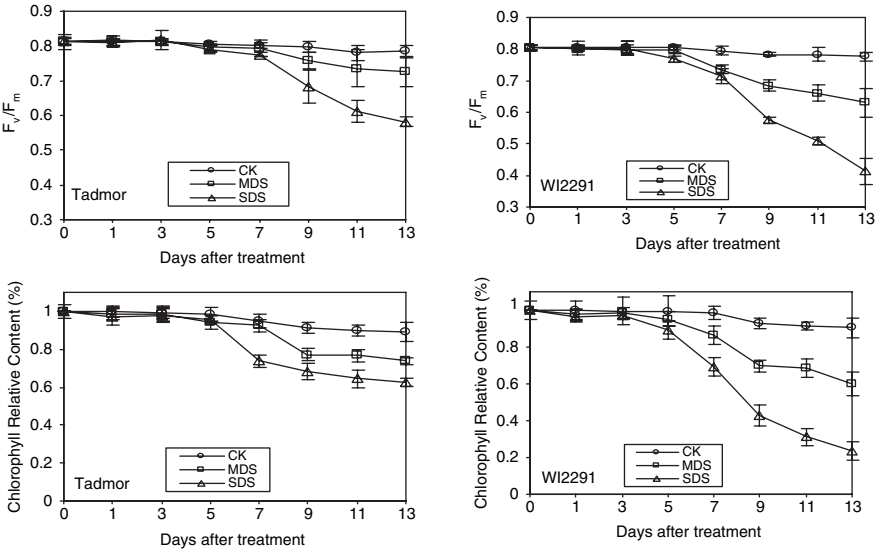


Figure 4. Chlorophyll fluorescence parameters (Fv/Fm) and chlorophyll contents of Tadmor and WI2291 under well-watered conditions (mean ± SE of five individual measurements)

were found to be significantly differentially expressed between two genotypes at post-anthesis drought.

Under drought stress different varieties of barley should respond to this stress with changing expression of sets of genes regardless of whether they are tolerant or sensitive to drought. Some of these differentially expressed genes may only be responsive/induced genes without contributing to the tolerance, while some of them may be specific genes that allow the plant to adapt to the stress. A number of genes showed in both lines, Tadmor and WI2291 a differential expression under drought. The results indicated that the genes like dehydration-responsive protein RD22, early drought-induced protein, heat shock protein-like, dehydrin 3, protein kinase HvPKABA1, protein kinase SPK-3, pyrophosphate-dependent phosphofructokinase, etc., increased significantly with severe drought stress of post-anthesis stage in both varieties. These genes could putatively only be responsive/induced genes to drought stress. The comparative analysis of gene expression profiles between Tadmor and WI2291 indicated that some genes were regulated by drought stress and could putatively be the specific/responsible genes for drought tolerance in Tadmor. These include serine/threonine kinase-like protein, hepatoma-derived growth factor, inositol-3-phosphate synthase, microsomal signal peptidase, chlorophyll a/b-binding protein 1A, ribulose-bisphosphate carboxylase activase, pathogenesis related protein-1 and 4, PRB1-2, PRB1-4, ABC transporter family protein, organic anion transporter, glutathione S-transferase, OsGSTT1, etc.. All these results may provide new insights for the response of plants to water deficit stress.

6. SNPS FOR THE SELECTION OF ABIOTIC STRESSES

Single nucleotide polymorphisms (SNP) result from a single base mutation, which substitutes one base for another, are biallelic in nature and less polymorphic than the multiallelic microsatellites. However, the extraordinary abundance of SNPs in the genome largely offsets the disadvantage of their being biallelic and makes them the most attractive molecular marker system developed so far (Gupta et al., 2001). The allelic frequencies of a given SNP may vary in different populations. On average, an SNP is believed to be present in every 100 to 300 bases in any genome (Gupta et al., 2001). Their wide distribution in the genome, coupled with the availability of several high throughput assay systems, makes SNPs good candidates as genetic markers for both mapping and association genetic studies in plants (Wilson et al., 2003).

The development of new varieties tolerant to abiotic and biotic stresses is an essential part of the continued barley improvement program. Rostoks et al. (2005) developed more than 2000 genome-wide barley SNPs from eight diverse barley accessions. The genes used for SNP discovery were selected based on their transcriptional response to a variety of abiotic stresses such as high salt, drought, low temperature, low nitrate, and water logging. More than 300 SNPs were found in three mapping populations (Oregon Wolfe Dominant x Oregon Wolfe Recessive, Steptoe x Morex and Lina x *Hordeum spontaneum* 92) in an integrated linkage map incorporating a large number of RFLP, AFLP, and SSR markers. The integrated map showed good agreement with the published maps in terms of marker order, and known abiotic stress QTL mapped in relevant crosses. A number of stress responsive genes mapped to chromosome 5H and could potentially be involved in salt and low temperature tolerance. These include cyclic nucleotide-gated ion channels (scsnp46137, scsnp14060, scsnp14759, scsnp22879), putative heat shock proteins (scsnp00892, scsnp01204, scsnp02771, scsnp6096), M-type thioredoxin (scsnp03171), putative protein kinase (scsnp14350), putative multidrug, and toxic compound extrusion efflux transporter (scsnp15296). An understanding of the combined function and expression of genes involved in various abiotic stresses, which could help identify candidate genes underlying the QTL of interest is important.

7. CONCLUSIONS

Numerous QTL studies for agronomic performance have been conducted in favorable environments using balanced populations. Although drought tolerance has been recognized as one of the prime breeding goals to increase food production in arid areas, only very few QTL studies are available which analyzed agronomic performance in drought stressed environments. For agriculture in the Mediterranean region *H. spontaneum* plays a major role as the main source of drought tolerance, and first results of QTL studies with exotic germplasm look promising. A number of favorable exotic alleles for agronomic performance were detected, in particular,

in the studies conducted under drought (Baum et al., 2003, Talame et al., 2004) (Figure 2). However, further QTL studies on drought stress using land races and exotic barley as donors are clearly needed to make efficient use of the vast collection of germplasm resources. QTL studies with and without *H. spontaneum* have shown that developmental genes, notably those involved in flowering time and plant stature show pleiotropic effects on abiotic stress tolerance and ultimately determine yield. Important loci in this regard are the semi-dwarf gene *sdw3* (bin 7 on 2H), the flowering locus *Ppd-H1* (2H, bin 4) and the vernalization requirement *Vrn-H2* (4H, bin 13) (Figure 2). It was shown that *H. spontaneum* offers novel allelic variance at these loci. These alleles from wild barley can be exploited for the improvement of drought tolerance in elite germplasm by selecting those exotic alleles with the most beneficial pleiotropic effects on drought resistance.

Problems associated with the hybridisation of *H. spontaneum* such as alleles with deleterious effects on field performance, i.e. the brittle rachis loci *btr*, could be best addressed in the advanced backcross QTL analysis. It was interesting to see that in AB-QTL populations like in balanced populations major QTL overshadowed minor QTL-alleles. Nevertheless, crosses with *H. spontaneum*, AB-QTL populations and association studies with *H. spontaneum* have also identified new alleles and genes that are related to abiotic stress tolerance. Their description and exploitation in breeding lines might lead to the availability of a wider gene pool for adaptation to drought tolerance.

In order to better understand which genes are involved in drought and other abiotic stresses in the dry areas two approaches are important. Appropriate genetic resources and populations using new donors such as exotic barley and landraces have to be generated. Here, first the establishment of introgression lines (mainly used in relation to introgressions of *H. spontaneum* into *H. vulgare*) and their testing in dry areas would give more insight into the effect of certain favorable alleles. Second, much larger populations need to be tested in order to be able to identify confounding effects through major QTL loci. Furthermore, as arid areas may be characterized next to drought by very different environmental conditions, a thorough selection of test environments and comprehensive phenotyping are indispensable.

ACKNOWLEDGEMENTS

We would like to acknowledge the technical support of: A. Sabbag, G. Bacchus, B. Edris, T. Istamboli, S. Tawkaz and M. Baalbaki. The authors' research was supported by grants to ICARDA from the German Federal Ministry of Economic Cooperation and Development (BMZ, Bonn, Germany).

REFERENCES

Abebe T, Guenzi AC, Martin B, Cushman JC (2003) Tolerance of mannitol-accumulating transgenic wheat to water stress and salinity. Plant Physiol 131:1748–1755

Acevedo E, Ceccarelli S (1989) Role of physiologist-breeder in a breeding program for drought resistance conditions. In: Baker FWG. (ed) Drought resistance in cereals, CAB International, 117–139

Araus JL, Bort J, Ceccarelli S, Grando S (1997) Relationship between leaf posture and carbon isotope discrimination in field grown barley. Plant Physiol Biochem 35:533–541

Azhacuvel P, Vidya-Saraswathi D, Komatsuda T (2006) High-resolution linkage mapping for the non-brittle rachis locus btr1 in cultivated × wild barley (*Hordeum vulgare*). Plant Science 170: 1087–1094

Baum M, Grando S, Backes G, Jahoor A, Sabbagh A, Ceccarelli S (2003) QTLs for agronomic traits in the Mediterranean environment identified in recombinant inbred lines of the cross 'Arta' × *H. spontaneum* 41-1. Theor Appl Genet 107:1215–1225

Blum A (1988) Drought resistance. In: Plant breeding for stress environments. CRC Press, Boca Raton, FL, pp 43–76

Blum A (1996) Crop responses to drought and the interpretation of adaptation. Plant Growth Reg 20:135–148

Boyer JS (1982) Plant productivity and environment. Science 218:443–448

Bray EA (1993) Molecular responses to water deficit. Plant Physiol 103:1035–1040

Bruce WB, Edmeades GO, Barker TC (2002) Molecular and physiological approaches to maize improvement for drought tolerance. J Exp Bot 53:13–25

Brush SB (1999) Genes in the field: On-farm conservation of crop diversity. IPGRI/IDRC/Lewis Publishers, pp 51–76

Cattivelli L, Baldi P, Crosatti C, Di Fonzo N, Faccioli P, Grossi M, Mastrangelo AM, Pecchioni N, Stanca AM (2002) Chromosome regions and stress-related sequences involved in resistance to abiotic stress in Triticeae. Plant Mol Biol 48:649–665

Ceccarelli S (1994) Specific adaptation and breeding for marginal conditions. Euphytica 77:205–219

Ceccarelli S (1996a) Adaptation to low/high input cultivation. Euphytica 92:203–214

Ceccarelli S (1996b) Positive interpretation of genotype by environment interactions in relation to sustainability and biodiversity. In: Cooper M, Hammer GL (eds) Plant adaptation and crop improvement. CAB International, Wallingford, UK, ICRISAT, Andra Pradesh, India, IRRI, Manila, Philippines, pp 467–486

Ceccarelli S, Grando S (1996) Drought as a challenge for the plant breeder. Plant Growth Reg 20:149–155

Ceccarelli S, Acevedo E, Grando S (1991) Breeding for yield stability in unpredictable environments: single traits, interaction between traits, and architecture of genotypes. Euphytica 56:169–185

Ceccarelli S, Grando S, Baum M, Udupa SM (2004) Breeding for drought resistance in a changing climate. Crop science society of America and American society of agronomy, 677 S. Segoe Rd., Madison, WI 53711, USA. Challenges and Strategies for Dryland Agriculture. CSSA Special Publication no. 32

Cheong YH, Kim K-N, Pandey GK, Gupta R, Grant JJ, Luan S (2003) CBL1, a calcium sensor that differentially regulates salt, drought, and cold responses in Arabidopsis. Plant Cell 15:1833–1845

Choi DW, Koag MC, Close TJ (2000) Map locations of barley *Dhn* genes determined by gene-specific PCR. Theor Appl Genet 101:350–354

Choi DW, Rodriguez EM, Close TJ (2002) Barley *Cbf3* gene identification, expression pattern, and map location. Plant Physiol 129:1781–1787

Condon AG, Richards RA, Farquhar GD (1987) Carbon isotope discrimination is positively correlated with grain yield and dry matter production in field grown wheat. Crop Sci 27:996–1001

Cooper PJM, Gregory PJ, Tully D, Harris HC (1987) Improving water use efficiency of annual crops in the rainfed farming systems of West Asia and North Africa. Exp Agric 23:113–158

Diab AA, Teulat-Merah B, This D, Ozturk NZ, Benscher D, Sorrells ME (2004) Identification of drought-inducible genes and differentially expressed sequence tags in barley. Theor Appl Genet 109:1417–1425

Dubcovsky J, Lijavetzky D, Appendino L, Tranquilli G (1998) Comparative RFLP mapping of *Triticum monococcum* genes controlling vernalization requirement. Theor Appl Genet 97:968–975

Ellis RH, Hadley P, Roberts EH, Summerfield RJ (1990) Quantitative relationship between temperature and crop development and growth. In: Jackson M, Ford-Lloyd BV, Parry ML (Eds) Climatic change and plant genetic resources. Belhaven Press, London and New York, pp 85–115

Ellis RP, Forster BP, Robinson D, Handley LL, Gordon DC, Russell JR, Powell W (2000) Wild barley: a source of genes for crop improvement in the 21st century? J Exp Bot 51:9–17

Ellis RP, Forster BP, Gordon DC, Handley LL, Keith RP, Lawrence P, Meyer R, Powell W, Robinson D, Scrimgeour CM, Young G, Thomas WT (2002) Phenotype/genotype associations for yield and salt tolerance in a barley mapping population segregating for two dwarfing genes. J Exp Bot 2002:1163–1176

Eshed Y, Zamir D (1994) A genomic library of *Lycopersicon pennellii* in *Lycopersicon esculentum*-a tool for fine mapping of genes. Euphytica 79:175–179

Farquhar GD, Richards RA (1984) Isotopic composition of plant carbon correlates with water-use-efficiency of wheat genotypes. Aust J Plant Physiol 11:539–552.

Forster BP, Ellis RP, Moir J, Talame V, Sanguineti MC, Tuberosa R, This D, Teulat-Merah B, Ahmed I, Mariy SAEE, Bahri H, El Ouahabi M, Zoumarou-Wallis N, El-Fellah M, Ben Salem M (2004a) Genotype and phenotype associations with drought tolerance in barley tested in North Africa. Ann Appl Biol 144:157–168

Forster BP, Ellis RP, Thomas WTB, Newton AC, Tuberosa R, This D, El-Enein RA, Bahri MH, Ben Salem M (2004b) The development and application of molecular markers for abiotic stress tolerance in barley. J Exp Bot 51:19–27

Francia E, Rizza F, Cattivelli L, Stanca AM, Galiba G, Tóth B, Hayes PM, Skinner JS, Pecchioni N (2004) Two loci on chromosome 5H determine low-temperature tolerance in a 'Nure' (winter) × 'Tremois' (spring) barley map. Theor Appl Genet 108:670–680

Frary A, Doganlar S, Frampton A, Fulton T, Uhlig J, Yates H, Tanksley S (2000) Fine mapping of quantitative trait loci for improved fruit characteristics from *Lycopersicon chmielewskii* chromosome 1. Genome 46:235–243

Fulton TM, Nelson JC, Tanksley SD (1997) Introgression and DNA marker analysis of *Lycopersicon peruvianum*, a wild relative of the cultivated tomato, into *Lycopersicon esculentum*, followed through three successive backcross generations. Theor Appl Genet 95:895–902

Gottwald S, Börner A, Stein N, Sasaki T, Graner A (2004) The gibberellic-acid insensitive dwarfing gene *sdw3* of barley is located on chromosome 2HS in a region that shows high colinearity with rice chromosome 7L. Mol Genet Genomics 271:426–436

Grando S, von Bothmer R, Ceccarelli S (2001) Genetic diversity of barley: use of locally adapted germplasm to enhance yield and yield stability of barley in dry areas. In: Cooper HD, Spillane C, Hodgink T (eds) Broadening the genetic base of crop production. CABI/FAO/IPGRI, pp 351–372

Grando S, Baum M, Ceccarelli S, Goodchild A, Jaby El-Haramein F, Jahoor A, Backes G (2005) QTLs for straw quality characteristics identified in recombinant inbred lines of a *Hordeum vulgare/H. spontaneum* cross in a Mediterranean environment. Theor Appl Genet 110:688–695

Guo P, Baum M, Grando S, Valkoum J, Varshney RK, Graner A, Ceccarelli S (2005) Expression analysis of barely genes in response to drought stress during the reproductive growth stage using micrarray. Interdrought II, 24–28 June, Rome, Italy

Gupta PK, Roy JK, Prasad M (2001) Single nuleotide polymorphisms: a new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. Curr Sci 80:524–534

Hori K, Sato K, Nankaku N, Taked K (2005) QTL analysis in recombinant chromosome substitution lines and doubled haploid lines derived from a cross between *Hordeum vulgare* ssp *vulgare* and *Hordeum vulgare* ssp *spontaneum*. Mol Breed 16:295–311

Ingram J, Bartels D (1996) The molecular basis of dehydration tolerance in plants. Annu Rev Plant Physiol Plant Mol Biol 47:377–403

Ivandic V, Hackett CA, Zhang ZJ, Staub JE, Nevo E, Thomas WTB, Forster BP, (2000) Phenotypic responses of wild barley to experimentally imposed water stress. J Exp Bot 353:2021–2029

Ivandic V, Forster BP, Hackett CA, Nevo N, Keith R, Thomas TBW (2002) Analysis of simple sequence repeats (SSRs) in wild barley from the fertile crescent: associations with ecology, geography and flowering time. Plant Mol Biol 48:511–527

Ivandic V, Thomas WTB, Nevo E, Zhang Z, Forster BP (2003) Associations of simple sequence repeats with quantitative trait variation including biotic and abiotic stress tolerance in Hordeum spontaneum. Plant Breed 122:300–304

Karsai I, Hayes PM, Kling J, Matus IA, Meszaros K, Lang L, Bedo Z, Sato H (2004) Genetic variation in component traits of heading date in *Hordeum vulgare* subsp. *spontaneum* accessions characterized in controlled environments. Crop Sci 44:1622–1632

Kleinhofs A, Graner A (2001) An integrated map of the barley genome. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp 187–99

Laurie DA, Pratchett N, Bezant JH, Snape JW (1995) RFLP mapping of 5 major genes and 8 quantitative trait loci controlling flowering time in a winter – spring barley (*Hordeum vulgare* L.) cross. Genome 38:575–585

Li J, Huang XQ, Heinrichs F, Ganal MW, Röder MS (2005) Analysis of QTLs for yield, yield components, and malting quality in a BC_3-DH population of spring barley. Theor Appl Genet 110: 356–363

Li JZ, Huang XQ, Heinrichs F, Ganal MW, Roeder MS (2006) Analysis of QTLs for yield components, agronomic traits, and disease resistance in an advanced backcross population of spring barley. Genome 49:454–466

Liu X, Vance Baird WM (2004) Identification of a novel gene, HAABRC5, from *Helianthus annuus* (Asteraceae) that is upregulated in response to drought, salinity, and abscisic acid. Am J Botany 91:184–191

Long NR, Jeffries SP, Warner P, Karakousis A, Kretschmer JM, Hunt C, Lim P, Eckermann PJ, Barr A (2003) Mapping and QTL analysis of the barley population Mundah × Keel. Aust J Agric Res 54:1163–1171

Ludlow MM, Muchow RC (1990) A critical evaluation of traits for improving crop yields in water-limited environments. Adv Agron 43:107–153

Matus I, Corey A, Filchkin T, Hayes PM, Vales MI, Kling J, Riera-Lizarazu O, Sato K, Powell W, Waugh R (2003) Development and characterization of recombinant chromosome substitution lines (RCSLs) using *Hordeum vulgare* subsp. *spontaneum* as a source of donor alleles in a *Hordeum vulgare* subsp. *vulgare* background. Genome 46:1010–1023

Morgan JM (1983) Osmoregulation as a selection criterion for drought tolerance in wheat. Aust J Agric Res 34:607–614

Nakashima K, Yamaguchi-Shinozaki K (2006) Regulons involved in osmotic stress-responsive and cold stress-responsive gene expression in plants. Physiol Plantarum 126:62–71

Nevo E (1992) Origin, evolution, population genetics and resources for breeding of wild barley, Hordeum spontaneum, in the Fertile Crescent. In: Shewry PR (ed) Barley: genetics, biochemistry, molecular biology and biotechnology, CAB International, pp 19–43

Ozturk ZN, Talame V, Deyholos M, Michalowski CB, Galbraith DW, Gozukirmizi N, Tuberosa R, Bohnert HJ (2002) Monitoring large-scale changes in transcript abundance in drought- and salt-stressed barley. Plant Mol Biol 48:551–573

Pan A, Hayes PM, Chen F, Blake THH, Wright TKS, Karsai I, Bedö Z (1994) Genetic analysis of the components of winter hardiness in barley (*Hordeum vulgare* L.). Theor Appl Genet 89:900–910

Pellegrineschi A, Noguera LM, Skovmand B, Brito RM, Velazquez L, Salgado MM, Hernandez R, Warburton M, Hoisington D (2002) Identification of highly transformable wheat genotypes for mass production of fertile transgenic plants. Genome 45:421–30

Pellegrineschi A, Reynolds M, Pacheco M, Brito RM, Almeraya R, Yamaguchi-Shinozaki K, Hoisington D (2004) Stress-induced expression in wheat of the Arabidopsis thaliana DREB1A gene delays water stress symptoms under greenhouse conditions. Genome 47:493–500

Pillen K, Zacharias A, Léon J (2003) Advanced backcross QTL analysis in barley (*Hordeum vulgare* L.). Theor Appl Genet 107:340–352

Pillen K, Zacharias A, Léon J (2004) Comparative AB-QTL analysis in barley using a single exotic donor of *Hordeum vulgare* ssp. *spontaneum*. Theor Appl Genet 108:1591–1601

Reinheimer JL, Barr AR, Eglinton JK (2004) QTL mapping of chromosomal regions conferring reproductive frost tolerance in barley (*Hordeum vulgare* L.). Theor Appl Genet 109:1267–1274

Rostoks N, Mudie S, Cardle L, Russell J, Ramsay L, Booth A, Svensson Jt, Wanamaker SI, Walia H, Rodriguez EM, Hedley PE, Liu H, Morris J, Close TJ, Marshall DF, Waugh R (2005) Genome-wide

SNP discovery and linkage analysis in barley based on genes responsive to abiotic stress. Mol Gen Genomics 274:515–527

Schauer N, Semel Y, Roessner U, Gur A, Balbo I, Carrari F, Pleban T, Perez-Melis A, Bruedigam C, Kopka J, Willmitzer L, Zamir D, Fernie AR (2006) Comprehensive metabolic profiling and phenotypeing of interspecific introgression lines for tomato improvement. Nat Biotechnol 24: 447–454

Seki M, Narusaka M, Abe H, Kasuga M, Yamaguchi-Shinozaki K, Carninici P, Hayashizaki Y, Shinozaki K, (2001) Monitoring the expression pattern of 1300 Arabidopsis genes under drought and cold stress by using a full-length cDNA microarray. Plant Cell 13:61–72

Seki M, Narusaka M, Ishida J, Nanjo T, Fujita M, Oono Y, Kamiya A, Nakajima M, Enju A, Sakurai T et al (2002) Monitoring the expression profiles of 7,000 Arabidopsis genes under drought, cold and high-salinity stresses using a full-length cDNA microarray. Plant J 31:279–292

Shinozaki K, Yamaguchi-Shinozaki K, Seki M (2003) Regulatory network of gene expression in the drought and cold stress responses. Curr Opin Plant Biol 6:410–417

Singh M, Malhotra RS, Ceccarelli S, Sarker A, Grando S, Erskine W (2003) Spatial variability models to improve dryland field trials. Exp Agric 39:1–10.

Skinner J, Szucs P, von Zitzewitz J, Marquez-Cedillo L, Filichkin T, Stockinger EJ, Thomashow MF, Chen THH, Hayes PM (2006) Mapping of barley homologs to genes that regulate low temperature tolerance in *Arabidopsis.* Theor Appl Genet 112:832–842

Snape JW, Sarma R, Quarrie SA, Fish L, Galiba G, Sutka J (2001) Mapping genes for flowering time and frost tolerance in cereals using precise genetic stocks. Euphytica 120:309–315

Talamé V, Ozturk N, Bohnert HJ, Moris LN, Charles PN (2003) Microarray analysis of transcript abundance in barley under conditions of water deficit. Plant Animal Genome Conf XI, January 11–15, San Diego, CA, USA, p 15

Talamé V, Sanguineti MC, Chiapparino E, Bahri H, Ben Salem M, Forster BP, Ellis RP, Rhouma S, Zoumarou W, Waugh R, Tuberosa R (2004) Identification of *Hordeum spontaneum* QTL alleles improving field performance of barley grown under rainfed conditions. Ann Appl Biol 144: 309–319

Tanksley SD, Nelson JC (1996) Advanced backcross QTL analysis: a method of the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. Theor Appl Genet 92:191–203

Tanksley SD, Grandillo S, Fulton TM, Zamir D, Eshed Y, Petiard V, Lopez J, Beck-Brunn T (1996) Advanced backcross QTL analysis in a cross between an elite processing line of tomato and its wild relative *L. pimpinellifolium*. Theor Appl Genet 92:213–224

Teulat B, This D, Khairallah M, Borries C, Ragot C, Sourdille P, Leroy P, Monneveux P, Charrier A (1998) Several QTLs involved in osmotic adjustment trait variation in barley (*Hordeum vulgare* L.). Theor Appl Genet 96:688–698

Teulat B, Borries C, This D (2001a) New QTLs identified for plant water status, water-soluble carbohydrates, osmotic adjustment in a barley population grown in a growth-chamber under two water regimes. Theor Appl Genet 103:161–170

Teulat B, Merah O, Souyris I, This D (2001b) QTLs for agronomic traits from Mediterranean barley progeny grown in several environments. Theor Appl Genet 103:774–787

Teulat B, Merah O, Sirault X, Borries C, Waugh R, This D (2002) QTLs for grain carbon isotope discrimination in field-grown barley. Theor Appl Genet 106:118–126

Teulat B, Zoumarou-Wallis N, Rotter B, Ben Salem M, Bahri H, This D (2003) QTL for relative water content in field-grown barley and their stability across Mediterranean environments. Theor Appl Genet 108:181–188

Tondelli A, Francia E, Barabaschi D, Aprile A, Skinner JS, Stockinger EJ, Stanca AM, Pecchioni N (2006) Mapping regulatory genes as candidates for cold and drought stress tolerance in barley. Theor Appl Genet 112:445–454

Turner NC (2002) Optimizing water use. In: Nösberger L, Geiger HH, Struick PC (eds) Crop science: progress and prospects. CAB International, Wallingford, UK pp 119–135

Ueda A, Kathiresan A, Inada M, Narita Y, Nakamura T, Shi W, Takabe T, Bennett J (2004) Osmotic stress in barley regulates expression of a different set of genes than salt stress does. J Exp Bot 55: 2213–2218

Umezawa T, Fujita M, Fujita Y, Yamaguchi-Shinozaki K, Shinozaki K (2006) Engineering drought tolerance in plants: discovering and tailoring genes to unlock the future. Curr Opin Biotech 17:113–122

Van Oosterom EJ, Ceccarelli S, Peacock JM (1993) Yield response of barley to rainfall and temperature in Mediterranean environments. J Agric Sci 121:307–313

Vom Brocke K, Presterl T, Christinck A, Weltzien E, Geiger HH (2002) Farmers' seed management practices open up new base populations for pearl millet breeding in a semi-arid zone of India. Plant Breed 121:36–42

von Korff M (2005) Detection of QTL for agronomic traits and disease resistance in two advanced backcross populations derived from a wild barley accession (*Hordeum vulgare* ssp. *spontaneum*). Shaker Verlag Aachen. ISBN:3-8322-4293-7. Dissertation Universität Bonn

von Korff M, Wang H, Léon J, Pillen K (2004) Development of candidate introgression lines using an exotic barley accession (*H. vulgare* ssp. *spontaneum*) as donor. Theor Appl Genet 109:1736–1745

von Korff M, Wang H, Léon J, Pillen K (2005) Detection of resistance genes against powdery mildew, leaf rust and scald introgressed from wild barley (*H. vulgare* ssp. *spontaneum*). Theor Appl Genet 111:583–590

von Korff M, Wang H, Léon J, Pillen K (2006) AB-QTL analysis in spring barley: II. Detection of favourable exotic alleles for agronomic traits introgressed from wild barley (*H. vulgare* ssp. *spontaneum*). Theor Appl Genet 112:1221–1231

Wilson ID, Barker GL, Edwards KJ (2003) Genotype to phenotype: a technological challenge. Ann Appl Biol 142:33–39

Xiao J, Grandillo S, Ahn SN, McCouch SR, Tanksley SD, Li J, Yuan L (1996) Genes from wild rice improve yield. Nature 384:223–24

Xiao J, Li J, Grandillo S, Nag SN, Yuan L, Tanksley SD, McCouch SR (1998) Identification of trait-improving quantitative trait loci alleles from a wild rice relative, *Oryza rufipogon*. Genetics 150:899–909

Xu D, Duan X, Wang B, Hong B, Ho T, Wu R (1996) Expression of a late embryogenesis abundant protein gene, hva1, from barley confers tolerance to water deficit and salt stress in transgenic rice. Plant Physiol 110:249–257

Yun SJ, Gyenis L, Bossolini E, Hayes PM, Matus I, Smith KP, Steffenson BJ, Tuberosa R, Muehlbauer GJ (2006) Validation of quantitative trait loci for multiple disease resistance in barley using advanced backcross lines developed with a wild barley. Crop Sci 46:1179–1186

Zhang JZ, Creelman RC, Zhu JX (2004) From laboratory to field using information from Arabidopsis to engineer salt, cold, and drought tolerance in crop plants. Plant Physiol 135:615–621

Zhang F, Chen G, Huang Q, Orion O, Krugman T, Fahima T, Korol AB, Nevo E, Gutterman Y (2005) Genetic basis of barley caryopsis dormancy and seedling desiccation tolerance at the germination stage. Theor Appl Genet 110:445–453

CHAPTER 4

MOLECULAR MARKERS FOR GENE PYRAMIDING AND DISEASE RESISTANCE BREEDING IN BARLEY

WOLFGANG FRIEDT[1,*] AND FRANK ORDON[2]

[1]*Institute of Crop Science and Plant Breeding I, Justus-Liebig-University, Heinrich-Buff-Ring 26-32, 35392 Giessen, Germany*

[2]*Institute of Epidemiology and Resistance Resources, Federal Centre for Breeding Research on Cultivated Plants, Erwin-Baur-Str. 27, 06484 Quedlinburg, Germany*

Abstract: Barley (*Hordeum vulgare* L.) is one of the oldest crop plants and among the most important cereals worldwide which is cultivated from the polar circle to the tropics. Barley crops can be infected and severely damaged by many fungal, viral and bacterial pathogens as well as insect pests. Therefore, breeding for disease and pest resistance is of special importance in barley in order to prevent or reduce yield losses. Furthermore, genetic resistance allows the reduction of agrochemical applications and thereby greatly contributes to environment and consumer protection. Resistance breeding has already been very successful in the past and provided many resistant or tolerant barley varieties highly adapted to adverse growing conditions. The genetic basis of such resistance depends on the respective pathogen or pest: Many cases of monogenic (major gene) resistance have been described but oligo- or polygenic types of resistance are also widespread and appreciated by breeders and growers due to their superior durability. Today, molecular markers are available for many major resistance genes and resistance QTL (quantitative trait loci) against a wide range of pathogens in barley. Such markers are the basis for an efficient marker assisted selection (MAS) in scientific breeding research and commercial barley breeding. Furthermore, marker assisted backcrossing procedures allow an enhanced incorporation of resistance genes derived from non-adapted germplasm (e.g. *H. bulbosum*). Beyond this, pyramiding of resistance genes by marker-assisted combination breeding may lead to longer lasting resistance and enable the further use of resistance genes already overcome by single strains or isolates of the respective pathogen. This chapter reviews the present state of the art concerning markers available for major resistance genes and resistance QTL in barley and exemplifies marker assisted backcrossing and pyramiding strategies in the pathosystem barley – barley yellow mosaic virus disease caused by BaYMV and BaMMV.

*Author for correspondence: wolfgang.friedt@agrar.uni-giessen.de

R.K. Varshney and R. Tuberosa (eds.), Genomics Assisted Crop Improvement: Vol. 2: Genomics Applications in Crops, 81–101.

1. INTRODUCTION

Next to wheat, maize, and rice, barley (*Hordeum vulgare* ssp. *vulgare*) which is grown world-wide on an area of about 57 million hectares (FAO 2005) is number four of the world's cereal crops being very important for livestock feeding and human alimentation. Like in other crops, severe yield losses occur in barley every year due to fungal and viral diseases and as a result of insect pests. Therefore, breeding for disease and pest resistance has to be regarded as a major goal of barley breeding reducing yield losses and contributing to a sustainable environment and consumer friendly barley production due to reduced pesticide applications.

As a result of intensive breeding, modern barley cultivars combine pathogen resistances with high grain yield and superior seed quality. In general, combining of resistances or the introgression of new resistance genes from unadapted germplasms or related species, respectively, is achieved by sexual recombination, i.e. crossing of parental lines followed by phenotypic selection in the segregating offspring. In this case the success of breeding entirely depends on extensive field or glasshouse tests for resistance to the respective pathogens. But, as barley is damaged by many pathogens which often show a rapid adaptation to its hosts resistance genes, breeding for resistance is a very complex task and the identification of desired recombinants expressing resistance to most diseases by phenotypic selection only has nearly reached the limits of manageability. But, today plant biotechnology methods allow to improve the efficiency of selection and to accelerate the whole breeding process: for example, anther or microspore culture starting from F_1 donor plants immediately leads to homozygous doubled haploid plants facilitating easier and more accurate selection than in conventional segregating F_2 populations (Friedt & Ordon 2004, Werner et al. 2006). In addition, especially the development of molecular marker systems and the isolation of resistance genes transferring selection to some extent from the phenotypic to the genotypic level offer new opportunities for a more efficient breeding for resistance.

Since barley is grown from the polar circle to the tropics it is hit by a wide range of different fungal and viral diseases. Consequently, many corresponding resistance genes and quantitative trait loci (QTL) have been tagged already. In the following chapter an overview is given on such major resistance genes and QTL for resistance and possibilities to use respective markers in different marker based selection schemes are elucidated.

2. MOLECULAR MARKER TAGGED RESISTANCE GENES AND RESISTANCE QTL IN BARLEY

The development of molecular markers starting from restriction fragment length polymorphisms (RFLPs, Botstein et al. 1980), via the application of PCR-based procedures (Saiki et al. 1985, 1988), has dramatically increased the possibilities of an efficient application of marker based selection procedures in plant breeding. Meanwhile, closely linked PCR-based markers have been developed for many major resistance genes against important pathogens of barley (Table 1).

Table 1. List of mapped major resistance genes in barley based on Graner (1996) and Graner et al. (2000a), modified and extended

Resistance gene	Chromsomal location	Reference(s)
Powdery mildew (Blumeria graminis)		
mlo	4HL	Hinze et al. 1991, Büschges et al. 1997
Mlg	4HL	Görg et al. 1993, Kurth et al. 2001
Mla	1HS	Schüller et al. 1992, Schwarz et al. 1999, 2002, Wei et al. 1999, Zhou et al. 2001, Haltermann et al. 2001, 2004, 2006,
mlt	7HS	Schoenfeld et al. 1996
Mlj	5HL	Schoenfeld et al. 1996
Mlf	7HL	Schoenfeld et al. 1996
MlLa	2HL	Hilbers et al. 1992, Giese et al. 1993, Mohler & Jahoor 1996
Mlh	6H	Hilbers, cit Graner 1996
MlHb	2HS	Pickering et al. 1995, Graner et al. 1996b
Ml(TR)	5H	Falak et al. 1999
Mli	1HL	Jahoor, cit Graner 1996
Puccinia graminis		
Rpg1	7HS	Horvath et al. 1995, Penner et al. 1995, Han et al. 1999, Brueggemann et al. 2002
rpg4	5HL	Borovkova et al. 1995, Kilian et al. 1997
Puccinia hordei		
Rph2	5HS	Borovkova et al. 1997
Rph3	7HS	Park et al. 2003
Rph4	1H	Collins et al. 2001, Park et al. 2003
Rph5	3H	Mammadov et al. 2003, 2005
Rph7	3HS	Graner et al. 2000b, Brunner et al. 2000, 2003, Scherer et al. 2005
Rph9	5HL	Borovkova et al. 1998
Rph12	5HL	Borovkova et al. 1998, Park et al. 2003
Rph15	2HS	Weerasena et al. 2004
Rph16	2HS	Ivandic et al. 1998, Perovic et al. 2004
Rph19	7HL	Park & Karakousis 2002
Puccinia striiformis		
rpsGZ	4H	Yan & Chen 2006
Rhynchosporium secalis		
Rh (Rrs1)	3HL	Graner & Tekauz 1996, Penner et al. 1996, Reitan et al. 2002, Genger et al. 2003
Rhy	3HL	Barua et al. 1993

(*Continued*)

Table 1. (*Continued*)

Resistance gene	Chromsomal location	Reference(s)
Rh2	7HS	Schweizer et al. 1995, Schmidt et al. 2001
Rrs4	4HL	Patil et al. 2003
Rrs13	6HS	Abbott et al. 1995
Rrs14	1H	Garvin et al. 2000
Rrs15	7HL	Genger et al. 2005
Pyrenophora teres		
Pt.a	3HL	Graner et al. 1996
Pt.d	2HS	Graner & Tekauz, unpubl., cit. Graner et al. 2000
Rpt4	7H	Williams et al. 2004
Cochliobolus sativus		
Rcs5	7HS	Steffenson et al. 1996
Typhula incarnata		
Ti	1HS	Graner et al. 1996
Pyrenophora graminea		
Rdg1	2HL	Thomsen et al. 1997
Rdg2	7HS	Tacconi et al. 2001, Bulgarelli et al. 2004
Heterodera avenae		
Ha2	2HS	Kretschmer et al. 1997
Ha4	5HL	Barr et al. 1998
Barley stripe mosaic virus (BSMV)		
Rsm	7HS	Edwards & Stephenson 1996
Barley yellow dwarf virus (BYDV)		
Ryd2	3HL	Collins et al. 1996, Paltridge et al. 1998, Ford et al. 1998
Ryd3	6H	Niks et al. 2004
Barley yellow mosaic virus (BaYMV), Barley mild mosaic virus (BaMMV)		
rym1	4HL	Okada et al. 2004
rym3	5HS	Saeki et al. 1999, Werner et al. 2003a
rym4	3HL	Graner & Bauer 1993, Ordon et al. 1995, Weyen et al. 1996, Stein et al. 2005, Kanyuka et al. 2005
rym5	3HL	Graner et al. 1999a, Pellio et al. 2005, Stein et al. 2005, Kanyuka et al. 2005
rym7	1HS	Graner et al. 1999b
rym8	4HL	Bauer et al. 1997
rym9	4HL	Bauer et al. 1997, Werner et al. 2000
rym10	3HL	Graner et al. 1995,
rym11	4HL	Bauer et al. 1997, Nissan-Azzous et al. 2005
rym12	4HL	Graner et al. 1996a
rym13	4HL	Werner et al. 2003b
$Rym14^{Hb}$	6HS	Ruge et al. 2003
rym15	6H	Le Gouis et al. 2004
$Rym16^{Hb}$	2HL	Ruge-Wehling et al. 2006

Besides major resistance genes many QTL contributing to resistance against important pathogens have been detected. In contrast to resistance encoded by major genes it is consensus and to some extent supported by evidence that quantitative resistance is more durable *per se*. In the past, studies on quantitative, oligo- or polygenically inherited resistance have mainly focused on epidemiological studies and biometrics, but today efficient tools such as molecular marker techniques and sophisticated software packages are available, facilitating the dissection of quantitative resistance into individual Mendelian loci and the determination of their effect(s) and interactions. Regarding QTL-mapping, the development of locus-specific PCR-based markers, especially SSRs (Ramsay et al. 2000, Maccaulay et al. 2001) and the AFLP-technique (Voss et al. 1995) facilitating an accelerated construction of genetic maps in comparison to RFLPs (Graner et al. 1991), has enhanced QTL detection. Consequently, the number of disease resistance QTL has risen considerably and may rise in the future due to the development of more efficient finger print techniques in barley e.g. DArTs (Wenzl et al. 2004).

With respect to *Blumeria graminis* the causal agent of powdery mildew, first QTL have been mapped in barley in the beginning of the 1990s on chromosome 7H and 5H (Heun 1992). Since that time many additional QTL have been detected all over the barley genome (Saghai Maroof et al. 1994, Backes et al. 1995, Thomas et al. 1995, Backes et al. 2003), some of them mapping in the vicinity of major resistance genes e.g. on chromosome 1H (Backes et al. 1996, Emebri et al. 2005). Using Advanced Backcross QTL analysis (AB-QTL), QTL for resistance to powdery mildew, leaf rust (*Puccinia hordei*) and scald (*Rhynchosporium secalis*) have been detected in *H. spontaneum* (v. Korff et al. 2005).

For *P. hordei* three QTL conferring leaf rust resistance at the seedling stage (explaining 55% of the phenotypic variance), and 5 QTL for resistance in the adult plant stage (explaining 60% of the phenotypic variance) with two QTL in common have been identified (Qi et al. 1998, 2000, van Berloo et al. 2002). In additional studies it turned out that only one QTL had a substantial effect in both developmental stages, and it was shown that some of the QTL exhibit isolate-specific reactions. In addition, one QTL for resistance to leaf rust has been identified in the position of the *Rph16* gene (Kicherer et al. 2000). In a very recent approach, a QTL for resistance to *P. hordei* has also been detected using association mapping (Kraakman et al. 2006).

With regard to yellow or stripe rust (*P. striiformis* f. sp. *hordei*) first QTL had been detected on chromosomes 5HL and 4HL (Chen et al. 1994). In later studies these QTL were confirmed and additional QTL have been detected (Toojinda et al. 1998, 2000, Castro et al. 2002, Vales et al. 2005) and combined in order to improve resistance ("pyramided", see Castro et al. 2002, 2003a, 2003b, 2003c).

Regarding resistance to *Pyrenophora teres,* the causal agent of barley net blotch, 7 QTL responsible for resistance in the adult stage accounting for 67% of the phenotypic variation, and 3 QTL for resistance in the seedling stage explaining 47% of the phenotypic variance were detected by Steffenson et al. (1996). In more recent studies, QTL for resistance against this pathogen have been located on

chromosome 6H (Manninen et al. 2000, Cakir et al. 2003, Emebri et al. 2005). Ma et al. (2004) detected two QTL on chromosomes 6HS and 2HS explaining about 70% of the phenotypic variance. Analysing adult plant resistance to the spot form of net blotch (*Pyrenophora teres* f. *maculata*), QTL on chromosome 7H, 4H and 5H were identified (Williams et al. 2003). Furthermore, it has been shown that resistance to spot blotch *Cochliobolus sativus* (anamorph *Bipolaris sorokiniana*) at the adult plant stage is governed by two QTL explaining 70% of the phenotypic variance (Steffenson et al. 1996). Differences in QTL localisation for resistance to spot blotch were detected using different DH populations (Bilgic et al. 2005).

For resistance to scald (*Rhynchosporium secalis*) QTL were mapped on chromosomes 2HL (Backes et al. 1995), 3H, and 7H (Bjørnstad et al. 2004) as well as on 3H, 4H and 6H (Jensen et al. 2002). With respect to *Pyrenophora graminea,* the causal agent of barley leaf stripe, QTL were located on chromosomes 7HL, 2HS and 4HS (Pecchioni et al. 1996, 1999, 2000). Meanwhile, additional isolate and non isolate-specific loci were identified (Arru et al. 2002, 2003).

Among the fungal diseases of barley, Fusarium head blight (FHB, caused by *Fusarium* sp.) has recently gained evident importance. In this respect, three distinct regions contributing to resistance have been detected on chromosome 2H (Mesfin et al. 2003). QTL influencing mycotoxin (DON) content have also been detected on chromosome 2HL (Dahleen et al. 2004) and an additional major FHB QTL on chromosome 6H. Some of these QTL have been validated in different crosses and environments (Canci et al. 2004). In recent studies QTL for resistance to FHB have again been located on chromosome 2H, but additional QTL have been detected on the short arm of chromosome 5H (Hori et al. 2005) and on 4H (Horsley et al. 2006).

Besides these pathogens of major importance, QTL for resistance to bacterial leaf streak caused by *Xanthomonas campestris* pv. *hordei* (Attari et al. 1998) and to the blast disease caused by *Pyricularia oryzae* (Sato et al. 2001) have been detected. In addition to genetic loci for fungal and bacterial diseases, QTL for tolerance against *Barley yellow dwarf virus* (BYDV) have been identified (Toojinda et al. 2000, Scheurer et al. 2001). For BYDV-PAV and BYDV-MAV seven QTL were mapped and substantial QTL x environment interaction was observed (Toojinda et al. 2000). Scheurer et al. (2001) used a German isolate of BYDV-PAV and detected two QTL on chromosomes 2HL and 3HL in the region of the *Ryd2* gene together explaining 46% of the phenotypic variance. An additional QTL accounting for more than 70% of the phenotypic variance for tolerance to the German BYDV-PAV isolate has recently been mapped on chromosome 6H (Niks et al. 2004) and has been named *Ryd3* due to the large portion of phenotypic variance explained. QTL for resistance to the vectors of BYDV, i.e. several aphid species, have been detected on barley chromosomes 1H and 7H (Moharramipour et al. 1997).

It is interesting to note that some QTL map in chromosomal regions of barley where previously major resistance genes had been identified. For example, this has been shown for powdery mildew (Backes et al. 1996, 2003), yellow rust, leaf rust (Thomas et al. 1995, Kicherer 2000), and net blotch (Richter et al. 1998) resistances.

3. MARKER BASED STRATEGIES IN BREEDING FOR RESISTANCE

As described above, closely linked markers are available for many major resistance genes of barley as well as for numerous QTL. Besides this, some major resistance genes like *mlo* (Büschges et al. 1997), *Mla* (Haltermann et al 2001, Zhou et al. 2001), *Rpg1* (Brueggemann et al. 2002, Nirmala et al. 2006), or *rym4* (Stein et al. 2005, Kanyuka et al. 2005) have been isolated already, facilitating the development of allele specific functional genetic markers. These markers facilitating selection on the allele level as well as closely linked markers can be used in marker based selection procedures on the single plant level, i.e. in early generations, independently from the natural occurrence of respective pathogens in the field and from labor and time consuming greenhouse tests. Based on the known position of respective genes and QTL it can be calculated whether these are combinable at all and which population size is needed to combine these genes. In applied barley breeding e.g. markers for BaMMV/BaYMV resistance genes have been widely applied (Schiemann & Backes 2000) or for BYDV-resistance (Ovesna et al. 2000). Besides the application in mere marker based selection procedures, respective markers can be used to enhance backcrossing procedures and to pyramid resistance genes.

3.1. Marker Assisted Backcrossing

An incorporation of recessive resistance genes derived from exotic or un-adapted germplasm normally requires long-lasting backcrossing procedures to combine these resistances with other agronomic traits, in particular superior yield. Incorporating recessive resistance genes by phenotypic selection takes twice as long as incorporating dominant genes, since a selfing generation is required after each backcross (BC) step for the phenotypic identification of homozygous recessive resistant plants to be used in the next BC cycle. However, by using co-dominant markers, like SSRs, heterozygous carriers of the resistance encoding allele to be used for the next BC step can be detected directly in F_1, thereby saving one year for each BC cycle. The same holds true for dominant markers generating an additional fragment linked to the resistance encoding allele (cf. Ordon et al. 1999). Differences between conventional selection and a MAS procedure are exemplified in Figure 1 for the BaMMV resistance gene *rym15*. The time saving of this procedure becomes even more evident in the case that barley mild mosaic virus (BaMMV) and barley yellow mosaic virus (BaYMV) resistances are not encoded by the same but by different genes, like in the case of the Japanese variety ‘Chikurin Ibaraki 1’, carrying the BaMMV resistance *rym15* on chromosome 6H (Le Gouis et al. 2004) and a separate BaYMV resistance gene on chromosome 5H (Werner et al. 2003a). Since BaYMV is not efficiently transmissible by mechanical means, an additional selfing generation is needed to obtain a sufficient quantity of seeds for field testing, while in the MAS approach the track of both genes can be followed simultaneously. The use of markers leads

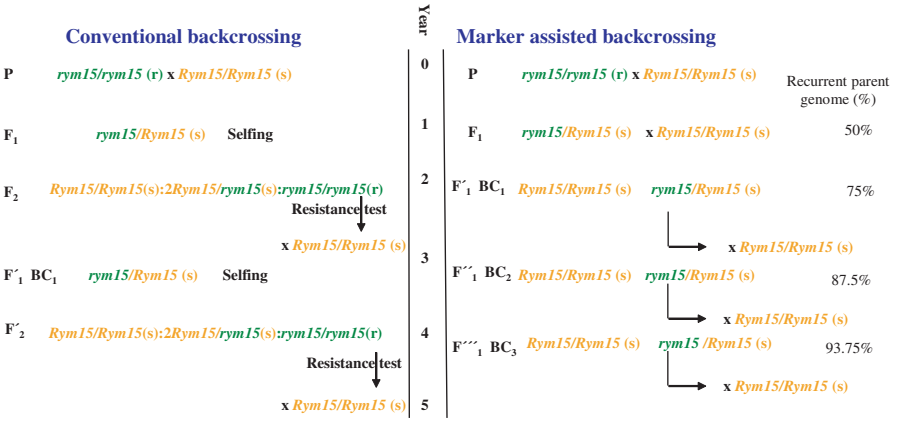


Figure 1. Comparison of conventional and marker-assisted backcrossing programmes for the incorporation of *rym15* (from Ordon et al. 1999, mod.) *(See Plate 1)*

to a drastic enhancement of BC procedures which may be further improved if the genomic portion of the recurrent parent is determined additionally. Results in different species have shown that there may be a significant deviation from the 75% genomic portion of the recurrent parent expected in the BC_1 generation (Powell et al. 1996, Uptmoor et al. 2006), thus demonstrating the usefulness of genotype-based selection.

3.2. Pyramiding Resistance Genes

Monogenic (qualitative) resistance, i.e. resistance conferred by a single major gene (Leach et al. 2001), is assumed to be non-durable in most cases, with exceptions like *mlo* resistance, due to the generally high mutation rate in plant pathogens (Takken & Joosten 2000, Kiyosawa 1982, Kloppers & Pretorius 1997). This can lead to the selection of new pathogen strains, able to overcome the effect(s) of an individual resistance gene. Additionally, the probability of disease resistance breakdown is increased by the large-scale and long-term cultivation of varieties carrying single genes enabling the pathogen to overcome the resistance (Singh et al. 2001). Thus, on the one hand screening for new sources of resistance is needed (Huang et al. 1997), followed by the detection, tagging and incorporation in adapted cultivars. On the other hand, strategies have been developed for extending the useful life-time of single resistance genes in cultivars in combination with creating cultivars with broader resistance-spectra, i.e. the accumulation of host resistance genes in a single line or cultivar, a procedure called pyramiding (Nelson 1978). Besides the extension of resistance durability the strategy of gene pyramiding is also applied for increasing the level and spectrum of disease resistance (Huang at al. 1997, Kinane & Jones 2000), leading to broad-spectrum resistance. This is of special interest, because resistance genes often differ in their specificity against the

various pathogen strains so that their combination can result in a broader spectrum of resistance than conferred by each gene alone.

For pyramiding, markers facilitating the simultaneous selection of several resistance genes and avoiding the need of repeated inoculation with different isolates followed by phenotyping are especially useful. By phenotypic selection alone it appears very demanding, difficult or even impossible to reliably combine and detect multiple resistance genes in one genotype, due to dominance and epistasis effects and/or simply masking of one gene's effect by another or others.

Attempts for pyramiding resistance genes have meanwhile been carried out in barley for several pathosystems. For example, pyramiding of different scald resistance genes resulted in improved resistances to the fungal pathogen *R. secalis* in comparison to lines carrying a single resistance gene: While Brown et al. (1996) performed the pair-wise combination of genes for resistance to scald by the use of linked isozyme markers, Raman et al. (1999) developed AFLP markers for this purpose. Pyramiding of genes *rym3* and *rym5* was achieved in Japan by using isozyme markers and phenotypic selection procedures (Saeki et al. 1999). Besides pyramiding major genes, trials were also carried out to combine positive QTL-alleles, e.g. Castro et al. (2000, 2003a, 2003b) created pyramids of resistance QTL for stripe rust of barley (*Puccinia striiformis* f.sp. *hordei*) on chromosomes 4H, 1H, and 5H with the objective of determining the relationship between QTL number and level of resistance. Pyramiding was carried out by using microsatellite markers, and a reduction of disease severity has been observed in genotypes with positive alleles at two or more loci. Similar results have recently been obtained by Richardson et al. (2006).

Pyramiding of major resistance genes may be of special importance in the future especially aiming at an extension of the utility of major resistance genes. For example, many resistance genes against the barley yellow mosaic virus complex have been described (Table 1). However, out of the recessive genes derived from the primary *H. vulgare* gene-pool only *rym11* is still effective against all strains known in Europe (Kanyuka et al. 2004, Habekuß et al. 2005). *Rym4* is not effective against BaYMV-2, *rym9* is ineffective against BaYMV and BaYMV-2, and *rym5* is not effective against BaMMV-SIL (Kanyuka et al. 2004) and the new strain of BaMMV (BaMMV-TASL) recently identified in Germany (Habekuß et al. 2005).

Studies on combining resistance genes against the barley yellow mosaic virus complex have focused on genes for which molecular markers were identified since the 1990s, i.e, *rym4, rym5, rym9* and *rym11* (Bauer et al. 1997; Graner et al. 1999a). While *rym4* and *rym5* map to the same marker interval in the telomeric region of chromosome 3HL (Graner et al. 1999a, Pellio et al. 2005) and represent alleles of the same gene, i.e. the eukaryotic translation initiation factor 4E (*Hv-eIF4E*, Stein et al. 2005), *rym9* is located in the telomeric and *rym11* in the centromeric region of chromosome 4HL (Bauer et al. 1997; Nissan-Azzouz et al. 2005). For pyramiding such recessive genes doubled haploid (DH) populations are better suited than segregating F_2 populations, since homozygous recessive

genotypes are much more frequent among DHs. In addition, dominant markers are as informative as co-dominant ones in such populations. Therefore, DH lines from crosses comprising the following gene combinations were generated*: rym4 x rym9, rym4 x rym11, rym5 x rym11, rym5 x rym9, rym9 x rym11*. Starting from these crosses two different strategies have been followed involving one or two cycles of DH production (Werner et al. 2005). For the first, F_1 derived DH populations were screened with appropriate closely linked PCR-based markers in order to identify plants carrying both genes in a homozygous recessive state, theoretically present at a frequency of 25%. These genotypes were inter-crossed followed by a DH line production step and re-analyses by the molecular markers to identify genotypes homozygous recessive for two or three resistance genes, respectively (Figure 2).

For the second strategy, F_1 plants having one resistance gene in common, e.g. [(*rym4* x *rym9*), (*rym4* x *rym11*)], were inter-crossed and at least 100 F'_1 kernels were produced for each cross (Figure 3, Werner et al. 2005). Genotypes homozygous recessive at the common resistance locus were identified using respective SSR markers. The selected genotypes were then screened for heterozygosity at the other resistance loci and DHs were produced from the resultant selections (Figure 4, Werner et al. 2006). In theory 6.25 plants of the 100 F'_1 plants should carry the expected genotype. Since less than the expected number of plants were identified in four of the crosses, genotypes were also selected that were heterozygous at all three resistance loci to enrich the pool. Among the DH offspring of these plants 20 out of 107 carried *rym4*, *rym9* and *rym11* and 27 out of 187 tested carried *rym5*, *rym9* and *rym11* in homozygous recessive states. In addition, DH-lines carrying all possible two-gene-combinations were achieved simultaneously (for details cf. Werner et al. 2005).

Since DH line production itself is time-consuming and costly the second strategy, including only one DH step, is considered more efficient than the first strategy with two DH steps. Although there are no cost differences concerning molecular analyses it should be noted, that in the one step strategy, co-dominant markers are a prerequisite because the first selection step takes place in heterozygous genotypes. For all other steps of both strategies dominant markers are as informative as co-dominant ones since heterozygous genotypes are absent in DH populations. Due to the appearance of the resistance breaking BaMMV-strain (Kanyuka et al. 2004, Habekuß et al. 2005) this strategy has gained actual importance. For example, combining the gene *rym5*, which at this time represents the sole basis of European barley cultivars resistant to BaMMV, BaYMV and BaYMV-2, with *rym9* effective against BaMMV and BaMMV-SIL (Kanyuka et al. 2004) should result in resistance against all yellow mosaic inducing viruses currently known in Europe.

In practical barley breeding pyramiding using molecular markers has to be repeated in each crossing cycle since the combined resistance genes will segregate in the progeny. However, this holds true only until respective major genes will be fixed within the actual breeding material, e.g. in Europe.

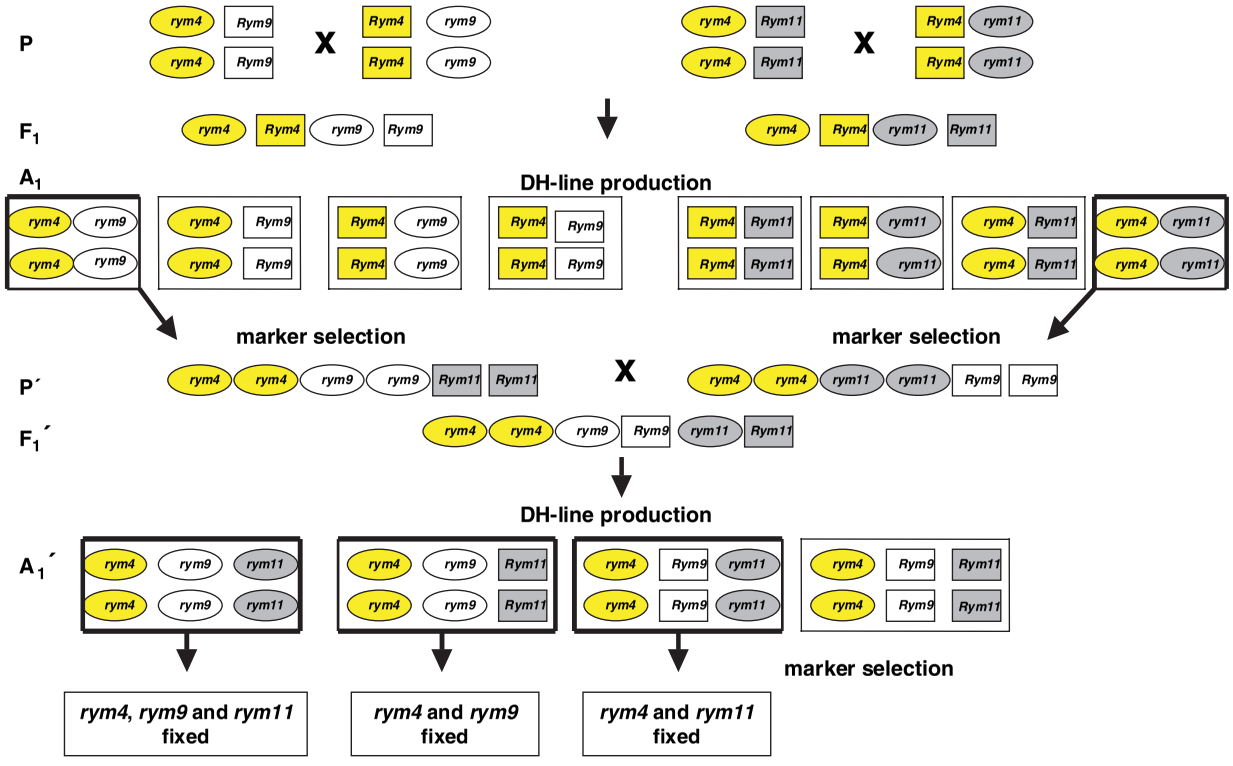


Figure 2. Scheme of pyramiding resistance genes *rym4*, *rym9* and *rym11* by two haploidy steps (Werner et al. 2005) *(See Plate 2)*

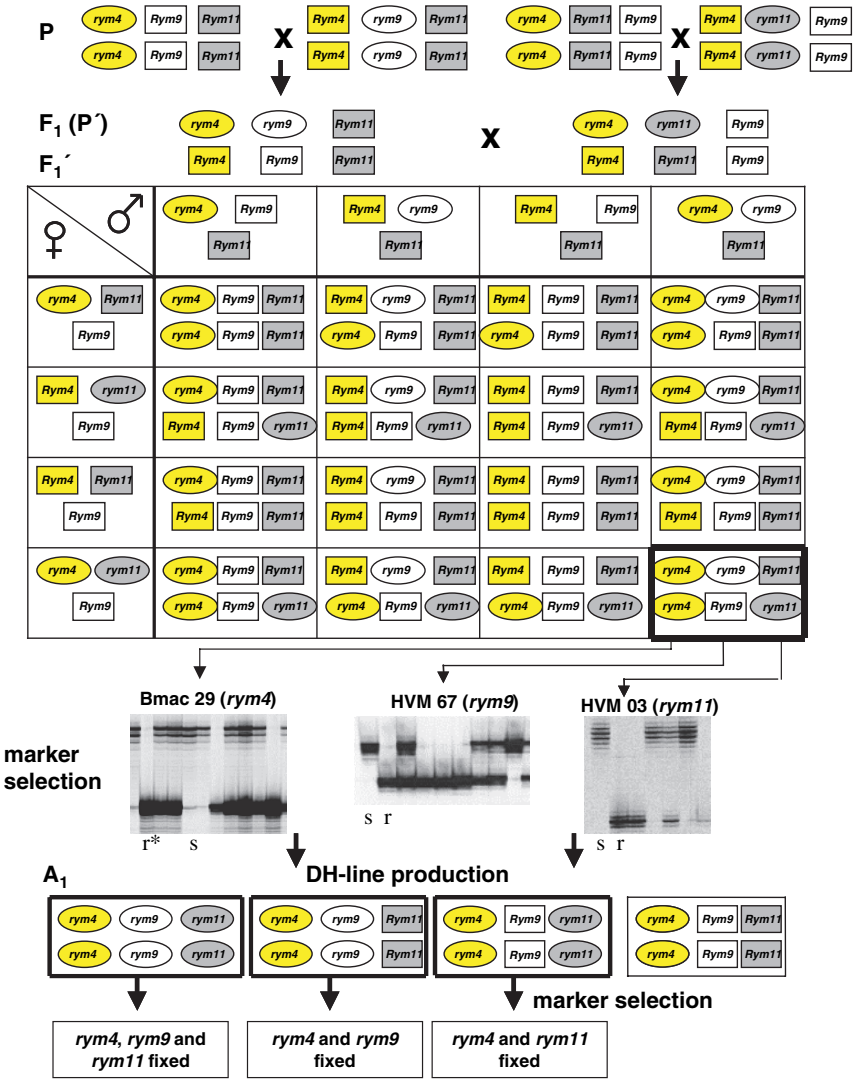


Figure 3. Scheme of pyramiding resistance genes *rym4*, *rym9* and *rym11* by one haploidy step (Werner et al. 2005) *(See Plate 3)*

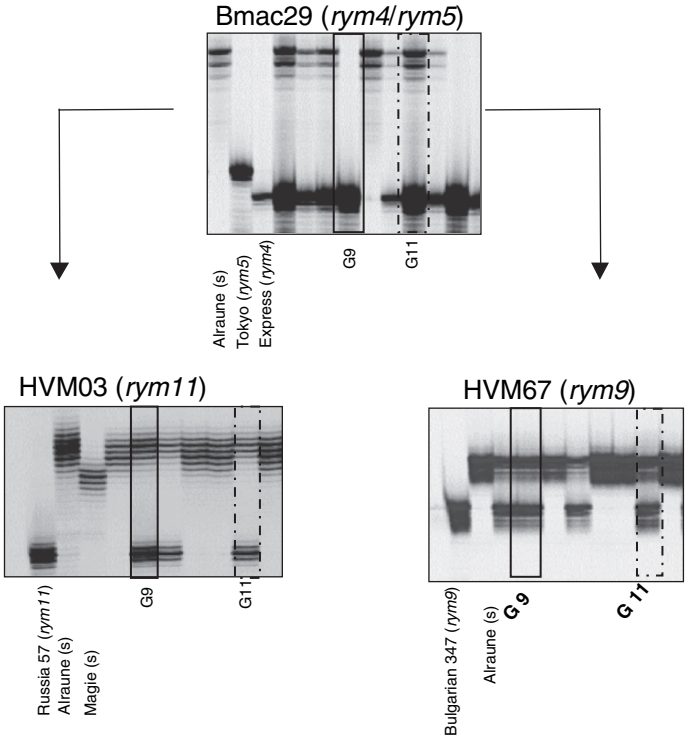


Figure 4. Scheme for selecting F_1 genotypes homozygous at the resistance locus in common (*rym4*) and heterozygous at the others (*rym11/rym9*, G9) and of genotypes heterozygous at all three loci (G11) as the starting point for pyramiding three and two resistance genes, respectively, by one DH step (Werner et al. 2006)

4. CONCLUSIONS AND FUTURE PERSPECTIVES

The results presented elucidate that molecular markers are available for many major resistance genes and QTL in barley facilitating – besides mere marker based selection procedures which are already applied in practical barley breeding for resistance to many pathogens e.g. BaMMV/BaYMV (Schiemann & Backes 2000) or BYDV (Ovesna et al. 2000) – the implementation of new breeding strategies like marker assisted backcrossing procedures or pyramiding of resistance genes. In the last few years the number of isolated resistance genes has risen considerably. However, the availability of expression profiling techniques and knowledge on synteny between the sequenced rice genome and barley will lead to an enhanced isolation of resistance genes in this species and a deeper understanding of respective pathosystems in the future. The isolation of genes involved in resistance and/or tolerance will transfer breeding for resistance in barley to the next level facilitating the identification of novel alleles and their directed use in molecular breeding strategies in order to enhance resistance.

REFERENCES

Abbott DC, Lagudah ES, Brown AHD (1995) Identification of RFLPs flanking a scald resistance gene in barley. J Hered 86:152–154

Arru L, Niks RE, Lindhout P, Vale G, Francia E, Pecchioni N (2002) Genomic regions determining resistance to leaf stripe (*Pyrenophora graminea*) in barley. Genome 45:460–466

Arru L, Francia E, Pecchioni N (2003) Isolate-specific QTLs of resistance to leaf stripe (*Pyrenophora graminea*) in the 'Steptoe' × 'Morex' spring barley cross. Theor Appl Genet 106:668–675

Attari El H, Hayes PM, Rebai A, Barrault G, Dechamp-Guillaume G, Sarrafi A (1998) Potential of doubled-haploid lines and localization of quantitative trait loci (QTL) for partial resistance to bacterial leaf streak (*Xanthomonas campestris* pv. *hordei*) in barley. Theor Appl Genet 96:95–100

Backes G, Graner A, Foroughi-Wehr B, Fischbeck G, Wenzel G, Jahoor A (1995) 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 90:294–302

Backes G, Schwarz G, Wenzel G, Jahoor A (1996) Comparison between QTL analysis on powdery mildew resistance in barley based on detached primary leaves and on field data. Plant Breed 115:419–421

Backes G, Madsen LH, Jaiser H, Stougaard J, Herz M, Mohler V, Jahoor A (2003) 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 106:353–362

Barr AR, Chalmers KJ, Karakousis A, Kretschmer JM, Manning S, Lange RCM, Lewis L, Jeffries SP, Langridge P (1998) RFLP mapping of a new cereal cyst nematode resistance locus in barley. Plant Breed 117:185–187

Barua UM, Chalmers KJ, Hackett CA, Thomas WTB, Powell W, Waugh R (1993) Identification of RAPD markers linked to a *Rhynchosporium secalis* resistance locus in barley using isogenic lines and bulked segregant analysis. Heredity 71:177–184

Bauer E, Weyen J, Schiemann A, Graner A, Ordon F (1997) Molecular mapping of novel resistance genes against barley mild mosaic virus (BaMMV). Theor Appl Genet 95:1263–1269

Bilgic H, Steffenson B, Hayes P (2005) Comprehensive genetic analyses reveal differential expression of spot blotch resistance in four populations of barley. Theor Appl Genet 111:1238–1250

Bjørnstad A, Grønnerød S, Key JM, Tekauz A, Crossa J, Martens H (2004) Resistance to barley scald (*Rhynchosporium secalis*) in the Ethiopian donor lines 'Steudelli' and 'Jet', analyzed by partial least squares regression and interval mapping. Hereditas (Lund) 141:166–179

Borovkova IG, Steffenson BJ, Jin Y, Rassmussen JB, Kilian A, Kleinhofs A, Rossnagel BG, Kao KN (1995) Identification of molecular markers linked to the stem rust resistance gene *rpg4* in barley. Phytopathology 85:181–185

Borovkova IG, Jin Y, Steffenson BJ, Kilian A, Blake TK, Kleinhofs A (1997) Identification and mapping of a leaf rust resistance gene in barley line Q21861. Genome 40:236–241

Borovkova IG, Jin Y, Steffenson BJ (1998) Chromosomal location and genetic relationship of leaf rust resistance genes *Rph9* and *Rph12* in barley. Phytopathology 88:76–80

Botstein D, White RL, Skolnik M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314–331

Brown AHD, Garvin DF, Burdon JJ, Abott DC, Read BJ (1996) The effect of combining scald resistance genes on disease levels, yield and quality traits in barley. Theor Appl Genet 93:361–366

Brueggeman R, Rostocks N, Kudrna D, Killian A, Han F, Chen J, Druka A, Steffenson B, Kleinhofs A (2002) The barley stem rust-resistance gene *Rpg*1 is a novel disease-resistance gene with homology to receptor kinases. Proc Natl Acad Sci USA 99:9328–9333

Brunner S, Keller B, Feuillet C (2000) Molecular mapping of the *Rph7* leaf rust resistance gene in barley (*Hordeum vulgare* L.). Theor Appl Genet 101:783–788

Brunner S, Keller B, Feuillet C (2003) A large rearrangement involving genes and low-copy DNA interrupts the microcollinearity between rice and barley at he *Rph*7 locus. Genetics 164:673–683

Bulgarelli D, Collins NC, Tacconi G, Dellaglio E, Brueggeman R, Kleinhofs A, Stanca AM, Vale G (2004) High-resolution genetic mapping of the leaf stripe resistance gene *Rdg*2a in barley. Theor Appl Genet 108:1401–1408

Büschges R, Hollrichter K, Panstruga R, Simons G, Wolter M, Frijiters A, van Daelen R, van der Lee T, Diergaarde P, Groendijk J, Töpsch S, Vos P, Salamini F, Schulze Lefert P (1997) The barley *Mlo* gene: a novel control element of plant pathogen resistance. Cell 88:695–705

Cakir M, Gupta S, Platz GJ, Ablett GA, Loughman R, Emebiri LC, Poulsen D, Li CD, Lance RCM, Galwey NW, Jones MGK, Appels R (2003) Mapping and validation of the genes for resistance to *Pyrenophora teres* f. *teres* in barley (*Hordeum vulgare* L.). Aust J Agric Res 54:1369–1377

Canci PC, Nduulu LM, Muehlbauer GJ, Dill-Macky R, Rasmusson DC, Smith KP (2004) Validation of quantitative trait loci for Fusarium head blight and kernel discoloration in barley. Mol Breed 14:91–104

Castro A, Corey A, Filichkin T, Hayes P, Sandoval-Islas JS, Vivar H (2000) Stripe rust resistance QTL pyramids in barley. 8th International Barley Genetics Symposium, Adelaide, Australia. Vol II, Contrib Papers, pp 86–88

Castro AJ, Chen XM, Hayes PM, Knapp SJ, Line RF, Toojinda T, Vivar H (2002) Coincident QTL which determine seedling and adult plant resistance to stripe rust in barley. Crop Sci 42:1701–1708

Castro AJ, Capettini F, Corey AE, Filichkina T, Hayes PM, Kleinhofs A, Kudrna D, Richardson K, Sandoval-Islas S, Rossi C, Vivar H (2003a) Mapping and pyramiding of qualitative and quantitative resistance to stripe rust in barley. Theor Appl Genet 107:922–930

Castro AJ, Chen XM, Corey A, Filichkina T, Hayes PM, Mundt C, Richardson K, Sandoval-Islas S, Vivar H (2003b) Pyramiding and validation of quantitative trait locus (QTL) alleles determining resistance to barley stripe rust: effects on adult plant resistance. Crop Sci 43:2234–2239

Castro AJ, Chen XM, Hayes PM, Johnston M (2003c) Pyramiding quantitative trait locus (QTL) alleles determining resistance to barley stripe rust: effects on resistance at the seedling stage. Crop Sci 43:651–659

Chen FQ, Prehn D, Hayes PM, Mulrooney D, Corey A, Vivar H (1994) Mapping genes for resistance to barley stripe rust (*Puccinia striiformis* f. sp. *hordei*). Theor Appl Genet 88:215–219

Collins NC, Paltridge NG, Ford CM, Symons RH (1996) The *Yd2* gene for barley yellow dwarf virus resistance maps close to the centromere on the long arm of barley chromosome 3. Theor Appl Genet 92:858–864

Collins N, Park R, Spielmeyer W, Ellis J, Pryor AJ (2001) Resistance gene analogs in barley and their relationship to rust resistance genes. Genome 44:375–381

Dahleen LS, Agrama HA, Horsley RD, Steffenson BJ, Schwarz PB, Mesfin A, Franckowiak JD (2004) Identification of QTLs associated with Fusarium head blight resistance in Zhedar 2 barley. Theor Appl Genet 108:95–104

Edwards MC, Steffenson BJ (1996) Genetics and mapping of barley stripe mosaic virus resistance in barley. Phytopathology 86:184–187

Emebiri LC, Platz G, Moody DB (2005) Disease resistance genes in a doubled haploid population of two-rowed barley segregating for malting quality attributes. Aust J of Agric Res 56:49–56

Falak I, Falk DE, Tinker NA, Mather DE (1999) Resistance to powdery mildew in a doubled haploid barley population and its association with marker loci. Euphytica 107:185–192

Ford CM, Paltridge NG, Rathjen JP, Moritz RL, Simpson RJ, Symons RH (1998) Rapid and informative assays for *Yd*2, the barley yellow dwarf virus resistance gene, based on the nucleotide sequence of a closely linked gene. Mol Breed 4:23–31

Friedt W, Ordon F (2004) Breeding for virus resistance of barley: amalgamation of classical and biotechnological approaches. Proceedings of the 9th International Barley Genetics Symposium, Brno, 329–337

Garvin DF, Brown AHD, Raman H, Read BJ (2000) Genetic mapping of the barley *Rrs*14 scald resistance gene with RFLP, isozyme and seed storage protein markers. Plant Breed 119:193–196

Genger RK, Brown AHD, Knogge W, Nesbitt K, Burdon JJ (2003) Development of SCAR markers linked to a scald resistance gene derived from wild barley. Euphytica 134:149–159

Genger RK, Nesbitt K, Brown AHD, Abbott DC, Burdon JJ (2005) A novel barley scald resistance gene: genetic mapping of the *Rrs15* scald resistance gene derived from wild barley, *Hordeum vulgare* ssp *spontaneum*. Plant Breed 124:137–141

Giese H, Holm-Jensen AG, Jensen HP, Jensen J (1993) Localization of the *Laevigatum* powdery mildew resistance gene to barley chromosome 2 by the use of RFLP markers. Theor Appl Genet 85:697–900

Görg R, Hollricher K, Schulze-Lefert P (1993) Functional analysis and RFLP-mediated mapping of the *Mlg* resistance locus in barley. Plant J 3:857–866

Graner A (1996) Molecular mapping of genes conferring disease resistance: the present state and future aspects. In: Scoles G, Rossnagel B (eds) Proceeding of the V International Oat Conference & VII International Barley Genetics Symposium, Inv Papers, Saskatoon, Canada, University Extension Press, Saskatoon, Sakatchewan, pp 157–166

Graner A, Bauer E (1993) RFLP mapping of the *ym4* virus resistance gene in barley. Theor Appl Genet 86:689–693

Graner A, Tekauz A (1996) RFLP mapping in barley of a dominant gene conferring resistance to scald (*Rhynchosporium secalis*). Theor Appl Genet 93:421–425

Graner A, Jahor A, Schondelmaier J, Siedler H, Pillen K, Fischbeck G, Wenzel G, Herrman RG (1991) Construction of an RFLP map in barley. Theor Appl Genet 83:250–256

Graner A, Bauer E, Kellermann A, Proeseler G, Wenzel G, Ordon F (1995) RFLP analysis of resistance to the barley yellow mosaic virus complex. Agronomie 15:475–479

Graner A, Bauer E, Chojecki J, Tekauz A, Kellermann A, Proeseler G, Michel M, Valkov V, Wenzel G, Ordon F (1996a) Molecular mapping of disease resistance in barley. In: Scoles G, Rossnagel B (eds) Proceeding of the International Oat Conference & VII International Barley Genetics Symposium, Poster Sessions Vol 1. Saskatoon, Canada, University Extension Press, Saskatoon, Sakatchewan, pp 253–255

Graner A, Foroughi-Wehr B, Tekauz A (1996b) RFLP mapping of a gene in barley conferring resistance to net blotch (*Pyrenophora teres*). Euphytica 91:229–234

Graner A, Streng S, Kellermann A, Schiemann A, Bauer E, Waugh R, Pellio B, Ordon F (1999a) Molecular mapping and genetic fine-structure of the *rym5* locus encoding resistance to different strains of the barley yellow mosaic virus complex. Theor Appl Genet 98:285–290

Graner A, Streng S, Kellermann A, Proeseler G, Schiemann A, Peterka H, Ordon F (1999b) Molecular mapping of genes conferring resistance to soil-borne viruses in barley – an approach to promote understanding of host–pathogen interactions. J Plant Dis and Prot 106:405–410

Graner A, Michalek W, Streng S (2000a) Molecular mapping of genes conferring resistance to viral and fungal pathogens. 8th International Barley Genetic Symposium Adelaide, Australia, Inv Papers, Vol I; 45–52

Graner A, Streng S, Drescher A, Jin Y, Borovkova T, Steffenson BJ (2000b) Molecular mapping of the leaf rust resistance gene *Rph*7 in barley. Plant Breed 119:389–392

Habekuss A, Kühne T, Rabenstein F, Krämer I, Ehrig F, Ruge-Wehling B, Huth W, Ordon F (2005) Detection of an *rym*5 resistance breaking virus strain in Germany. 6th Symposium International Working Group on Plant Viruses with Fungal Vectors, Bologna, Italy, p 60

Haltermann DA, Wise RP (2004) A single-amino acid substitution in the sixth leucine-rich repeat of barley *MLA6* and *MLA13* alleviates dependence on *RAR1* for disease resistance signalling. Plant J 38:215–226

Haltermann DA, Wise RP (2006) Upstream open reading frames of the barley *Mla13* powdery mildew resistance gene function co-operatively to down-regulate translation. Mol Plant Biol 7:167–176

Haltermann D, Zhou F, Wie F, Wise RP, Schulze-Lefert P (2001) The *Mla6* coiled-coil, NBS-LRR protein confers AvrMla6-dependent resistance specificity to *Blumeria graminis f. sp. hordei* in barley and wheat. Plant J 25:335–348

Han F, Kilian A, Chen JP, Kudrna D, Steffenson B, Yamamoto K, Matsumoto T, Sasaki T, Kleinhofs A (1999) Sequence analysis of a rice BAC covering the syntenous barley *Rpg*1 region. Genome 42:1071–1076

Heun M (1992) Mapping quantitative powdery mildew resistance of barley using a restriction fragment length polymorphism map. Genome 35:1019–1025

Hilbers S, Fischbeck G, Jahoor A (1992) Localization of the *Laevigatum* resistance gene *MlLa* against powdery mildew in the barley genome by the use of RFLP markers. Plant Breed 109: 335–338

Hinze K, Thompson RD, Ritter E, Salamini F, Schulze-Lefert P (1991) Restriction fragment length polymorphism-mediated targeting of the *mlo* resistance locus in barley (*Hordeum vulgare*). Proc Natl Acad Sci USA 88:3691–3695

Hori K, Kobayashi T, Sato K, Takeda K (2005) QTL analysis of Fusarium head blight resistance using a high-density linkage map in barley. Theor Appl Genet 111:1661–1672

Horsley RD, Schmieder D, Maier C, Kudrna D, Urrea CA, Steffenson BJ, Schwarz PB, Franckowiak JD, Green MJ, Zhang B, Kleinhofs A (2006) Identification of QTLs associated with Fusarium head blight resistance in barley accession Ciho 4196. Crop Sci 46:145–156

Horvarth DP, Dahleen LS, Stebbing JA, Penner G (1995) A co-dominant PCR-based marker for assisted selection of durable stem rust resistance in barley. Crop Sci 35:1445–1450

Huang B, Angeles ER, Domingo J, Magpantay G, Singh S, Zhang G, Kumaravadivel N, Benett J, Khush GS (1997) Pyramiding of bacterial blight resistance genes in rice: marker-assisted selection using RFLP and PCR. Theor Appl Genet 95:313–320

Ivandic V, Walther U, Graner A (1998) Molecular mapping of a new gene in wild barley conferring complete resistance to leaf rust (*Puccinia hordei* Otth). Theor Appl Genet 97: 1235–1239

Jensen J, Backes G, Skinnes H, Giese H (2002) Quantitative trait loci for scald resistance in barley localized by a non-interval mapping procedure. Plant Breed 121:124–128

Kanyuka K, McGrann G, Alhudaib K, Hariri D, Adams MJ (2004) Biological and sequence analysis of a novel European isolate of Barley mild mosaic virus that overcomes the barley *rym*5 gene. Arch Virol 149:1469–1480

Kanyuka K, Druka A, Caldwell DG, Tymon A, McCullam N, Adams MJ (2005) Evidence that the recessive bymovirus resistance locus *rym4* in barley corresponds to the eucaryotic translation initiation factor 4E gene. Mol Plant Pathol 6:449–458

Kicherer S, Backes G, Walther U, Jahoor A (2000) Localising QTLs for leaf rust resistance and agronomic traits in barley (*Hordeum vulgare* L.). Theor Appl Genet 100:881–888

Kilian A, Chen J, Han F, Steffenson B, Kleinhofs A (1997) Towards map-based cloning of the barley stem rust resistance genes *Rpg*1 and *rpg*4 using rice as an intergenomic cloning vehicle. Plant Mol Biol 35:187–195

Kinane JT, Jones PW (2000) Components of partial resistance to powdery mildew in wheat mutants. Eur J Plant Pathol 106:607–616

Kiyosawa S (1982) Genetics and epidemiological modelling of breakdown of plant disease resistance. Ann Rev Phytopathol 20:93–117

Kloppers FJ, Pretorius ZA (1997) Effects of combinations amongst genes *Lr13*, *Lr34* and *Lr37* on components of resistance in wheat to leaf rust. Plant Pathol 46:737–750

Kraakman ATW, Martinez F, Mussiraliev B, Eeuwijk FA, Niks RE (2006) Linkage disequilibrium mapping of morphological, resistance, and other agronomically relevant traits in modern spring barley cultivars. Mol Breed 17:41–58

Kretschmer JM, Chalmers KJ, Manning S, Karakousis A, Barr AR, Islam AKMR, Logue SJ, Choe YW, Barker SJ, Lance RCM, Langridge P (1997) RFLP mapping of the *Ha*2 cereal cyst nematode resistance gene in barley. Theor Appl Genet 94:1060–1064

Leach JE, Cruz CMV, Bai J, Leung H (2001) Pathogen fitness penalty as a predictor of durability of disease resistance genes. Ann Rev Phytopathol 39:187–224

Le Gouis J, Devaux P, Werner K, Hariri D, Bahrman N, Beghin D, Ordon F (2004) *Rym15* from the Japanese cultivar 'Chikurin Ibaraki 1' is a new Barley mild mosaic virus (BaMMV) resistance gene mapped on chromosome 6H. Theor Appl Genet 108:1521–1525

Ma ZQ, Lapitan NLV, Steffenson B (2004) QTL mapping of net blotch resistance genes in a doubled-haploid population of six-rowed barley. Euphytica 137:291–296

Macaulay M, Ramsay L, Powell W, Waugh R (2001) A representative, highly informative 'genotyping set' of barley SSRs. Theor Appl Genet 102:801–809

Mammadov JA, Zwonitzer JC, Biyashev RM, Griffey CA, Jin Y, Steffenson BJ, Maroof MAS (2003) Molecular mapping of leaf rust resistance gene *Rph*5 in barley. Crop Sci 43:388–393

Mammadov JA, Steffenson BJ, Maroof MAS (2005) High-resolution mapping of the barley leaf rust resistance gene *Rph5* using barley expressed sequence tags (ESTs) and synteny with rice. Theor Appl Genet 111:1651–1660

Manninen O, Kalendar R, Robinson J, Schulman AH (2000) Application of BARE-1 retrotransposon markers to the mapping of a major resistance gene for net blotch in barley. Mol Gen Genet 264:325–334

Mesfin A, Smith KP, Dill-Macky R, Evans CK, Waugh R, Gustus CD, Muehlbauer GJ (2003) Quantitative trait loci for fusarium head blight resistance in barley detected in a two-rowed by six-rowed population. Crop Sci 43:307–318

Moharramipour S, Tsumuki H, Sato K, Yoshida H (1997) Mapping resistance to cereal aphids in barley. Theor Appl Genet 94:592–596

Mohler V, Jahoor A (1996) Allele-specific amplification of polymorphic sites for the detection of powdery mildew resistance loci in cereals. Theor Appl Genet 93:1078–1082

Nelson RR (1978) Genetics of horizontal resistance to plant diseases. Annu Rev Phytopathol 16:359–378

Niks RE, Habekuß A, Bekele B, Ordon F (2004) A novel major gene on chromosome 6H for resistance to barley against the barley yellow dwarf virus. Theor Appl Genet 109:1536–1543

Nirmala J, Brueggemann R, Maier C, Clay C, Rostocks N, Kannangara CG, von Wettstein D, Steffenson BJ, Kleinhofs A (2006) Subcellular localization and functions of the barley stem rust resistance receptor-like serine/threonin-specific protein kinase *Rpg*1. Proc Natl Acad Sci 103:7518–7523

Nissan-Azzous F, Graner A, Friedt W, Ordon F (2005) Fine-mapping of the BaMMV, BaYMV-1 and BaYMV-2 resistance of barley (*Hordeum vulgare*) accession PI1963. Theor Appl Genet 110:212–218

Okada Y, Kanatani R, Arai S, Ito K (2004) Interaction between Barley yellow mosaic disease resistance gene *rym1* and *rym5*, in the response to BaYMV strains. Breed Sci 54:319–325

Ordon F, Bauer E, Friedt W, Graner A (1995) Marker-based selection for the *ym4* BaMMV-resistance gene in barley using RAPDs. Agronomie 15:481–485

Ordon F, Schiemann A, Pellio B, Dauck V, Bauer E, Streng S, Friedt W, Graner A (1999) Application of molecular markers in breeding for resistance to the barley yellow mosaic virus complex. J Plant Dis and Prot 106:256–264

Ovesna J, Vacke J, Kucera L, Chrpova J, Novakova I, Jahoor A, Sip V (2000) Genetic analysis of resistance in barley to barley yellow dwarf virus. Plant Breed 119:481–486

Paltridge NG, Collins NC, Bendahmane A, Symons RH (1998) Development of YLM, a codominant PCR marker closely linked to the *Yd2* gene for resistance to barley yellow dwarf disease. Theor Appl Genet 96:1170–1177

Park RF, Karakousis A (2002) Characterization and mapping of gene *Rph*19 conferring resistance to *Puccinia hordei* in the cultivar 'Reka 1' and several Australian barleys. Plant Breed 121:232–236

Park RF, Poulsen D, Barr AR, Cakir M, Moody DB, Raman H, Read BJ (2003) Mapping genes for resistance to *Puccinia hordei* in barley. Aust J Agric Res 54:1323–1333

Patil V, Bjørnstad A, Mackey J (2003) Molecular mapping of a new gene *Rrs*4 (CI11549) for resistance to barley scald (*Rhynchosporium secalis*). Mol Breed 12:169–183

Pecchioni N, Faccioli P, Toubia Rahme H, Valè G, Terzi V (1996) Quantitative resistance to barley leaf stripe (*Pyrenophora graminea*) is dominated by one major locus. Theor Appl Genet 93:97–101

Pecchioni N, Vale G, Toubia-Rahme H, Faccioli P, Terzi V, Delogu G (1999) Barley–*Pyrenophora graminea* interaction: QTL analysis and gene mapping. Plant Breed 118:29–35

Pecchioni N, Tacconi G, Arru L, Bellini L, Vale G (2000) The resistance of barley to leaf stripe caused by *Pyrenophora graminea*. Czech J Genet Plant Breed 36:88–91

Pellio B, Streng S, Bauer E, Stein N, Perovic D, Schiemann A, Friedt W, Ordon F, Graner A (2005) High-resolution mapping of the Rym4/Rym5 locus conferring resistance to the barley yellow mosaic virus complex (BaMMV, BaYMV, BaYMV-2) in barley (*Hordeum vulgare* ssp. *vulgare* L.) Theor Appl Genet 110:283–293

Penner GA, Stebbing JA, Legge B (1995) Conversion of an RFLP marker for the barley stem rust resistance gene *Rpg*1 to a specific PCR-amplifiable polymorphism. Mol Breed 1:349–354

Penner GA, Tekauz A, Reimer E, Scoles GJ, Rossnagel BG, Eckstein PE, Legge WG, Burnett PA, Ferguson T, Helm JF (1996) The genetic basis of scald resistance in western Canadian barley cultivars. Euphytica 92:367–374

Perovic D, Stein N, Zhang N, Drescher H, Prasad M, Kota R, Kopahnke D, Graner A (2004) An integrated approach for comparative mapping in rice and barley with special reference to the *Rph16* resistance locus. J Funct Integr Genomics 4:74–83

Pickering RA, Hill AM, Michel M, Timmermann-Vaughan GM (1995) The transfer of a powdery mildew resistance gene from *Hordeum bulbosum* L. to barley *H. vulgare* L chromosome 2 (2L). Theor Appl Genet 91:1288–1292

Powell W, Baird E, Booth A, Lawrence M, MacAulay M, Bonar N, Young G, Thomas WTB, McNicol JW, Waugh R (1996) Single locus and multi-locus molecular assays for barley breeding research. In: Scoles G, Rossnagel B (eds) Proceeding of the V Internation Oat Conference & VII International Barley Genetics Symposium Inv. Papers, Saskatoon, Canada, University Extension Press, Saskatoon, Sakatchewan, pp 174–181

Qi X, Niks RE, Stam P, Lindhout P (1998) Identification of QTLs for partial resistance to leaf rust (*Puccinia hordei*) in barley. Theor Appl Genet 96:1205–1215

Qi X, Fufa F, Sijtsma D, Niks RE, Lindhout P, Stam P (2000) The evidence for abundance of QTLs for partial resistance to *Puccinia hordei* on the barley genome. Mol Breed 6:1–9

Raman H, Read BJ, Brown AHD, Abbott DC (1999) Molecular markers and pyramiding of multiple genes for resistance to scald in barley. Proceedings of the 9th Australian Barley Technical Symposium, 1999

Ramsay L, Macaulay M, degli Ivanissevich S, MacLean K, Cardle L, Fuller J, Edwards KJ, Tuvesson S, Morgante M, Massari A, Maestri E, Marmiroli N, Sjakste T, Ganal M, Powell W, Waugh R (2000) A simple sequence repeat-based linkage map of barley. Genetics 156:1997–2005

Reitan L, Gronnerod S, Ristad TP, Salamati S, Skinnes H, Waugh R, Bjørnstad A (2002) Characterization of resistance genes against scald (*Rhynchosporium secalis* (Oudem.) J.J. Davis) in barley (*Hordeum vulgare* L.) lines from central Norway, by means of genetic markers and pathotype tests. Euphytica 123:31–39

Richardson KL, Vales MI, Kling JG, Mundt CC, Hayes PM (2006) Pyramiding and dissecting disease resistance QTL to barley stripe rust. Theor Appl Genet 113:485–495

Richter K, Schondelmaier J, Jung C (1998) Mapping of quantitative loci affecting *Drechslera teres* resistance in barley with molecular markers. Theor Appl Genet 97:1225–1234

Ruge B, Linz A, Proeseler G, Pickering G, Greif P, Wehling P (2003) Mapping of *Rym*14^{HB}, a gene introgressed from *Hordeum bulbosum* and conferring resistance to BaMMV and BaYMV in barley. Theor Appl Genet 107:965–971

Ruge B, Linz A, Habekuß A, Flath K, Wehling P (2004). Introgression and mapping of novel resistance genes from the secondary genepool of barley, *Hordeum bulbosum*. Proceedings of the 9th International Barley Genetics Symposium, Brno, Czech Republic, pp 729–736

Ruge-Wehling B, Linz A, Habekuß A, Wehling P (2006) Mapping of Rym16^{HB}, the second soil-borne virus-resistance gene introgressed from *Hordeum bulbosum*. Theor Appl Genet 113:867–873

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

Saghai Maroof MA, Zhang Q, Biyashev RM (1994) Molecular marker analyses of powdery mildew resistance in barley. Theor Appl Genet 88:733–740

Saiki RK, Scharf SJ, Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N (1985) Enzymatic amplification of ß-globulin genomic sequences and restriction site analysis for diagnosis of sickle cell anima. Science 230:1350–1354

Saiki RK, Gelfand S, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA (1988) Primer directed enzymatic amplification of DNA with thermostable DNA polymerase. Science 239:487–491

Sato K, Inukai T, Hayes PM (2001) QTL analysis of resistance to the rice blast pathogen in barley (*Hordeum vulgare*). Theor Appl Genet 102:916–920

Scherer B, Isidore E, Klein P, Kim JS, Bellec A, Chalhoub B, Keller B, Feuillet C (2005) Large intraspecific haplotype variability at the *Rph*7 locus results from rapid and recent divergence in the barley genome. Plant Cell 17:361–374

Scheurer KS, Friedt W, Huth W, Waugh R, Ordon F (2001) QTL analysis of tolerance to a German strain of BYDV-PAV in barley (*Hordeum vulgare* L.). Theor Appl Genet 103:1074–1083

Schiemann A, Backes G (2000) The use of molecular markers in practical barley breeding. In: Proceeding of the 8th International Barley Genetics Symposium Vol 3. Adelaide Australia pp 42–44

Schmidt D, Röder MS, Dargatz H, Wolf N, Schweizer GF, Tekauz A, Ganal MW (2001) Construction of a YAC library from barley cultivar Franka and identification of YAC-derived markers linked to the *Rh*2 gene conferring resistance to scald (*Rhynchosporium secalis*). Genome 44:1031–1040

Schönfeld M, Ragni A, Fischbeck G, Jahoor A (1996) RFLP mapping of three new loci for resistance genes to powdery mildew (*Erysiphe graminis* f. sp. *hordei*) in barley. Theor Appl Genet 93:48–56

Schüller C, Backes G, Fischbeck G, Jahoor A (1992) RFLP markers to identify the alleles on the *Mla* locus conferring powdery mildew resistance in barley. Theor Appl Genet 84:330–338

Schwarz G, Michalek W, Mohler V, Wenzel G, Jahoor A (1999) Chromosome landing at the *Ml*a locus in barley (*Hordeum vulgare* L.) by means of high-resolution mapping with AFLP-markers. Theor Appl Genet 98:521–530

Schwarz G, Michalek W, Jahoor A, Mohler V (2002) Direct selection of expressed sequences on a YAC clone revealed proline-rich-like genes and BARE-1 sequences physically linked to the complex *Mla* powdery mildew resistance locus of barley (*Hordeum vulgare* L.). Plant Sci 163:307–311

Schweizer GF, Baumer M, Daniel G, Rugel H, Röder MS (1995) RFLP markers linked to scald (*Rhynchosporium secalis*) resistance gene *Rh2* in barley. Theor Appl Genet 90:920–924

Singh S, Sidhu JS, Huang N, Vikal Y, Li Z, Brar DS, Dhaliwal HS, Khush GS (2001) Pyramiding three bacterial blight resistance genes (*xa5*, *xa13* and *Xa21*) using marker-assisted selection into indica rice cultivar PR106. Theor Appl Genet 102:1011–1015

Steffenson BJ, Hayes PM, Kleinhofs A (1996) Genetics of seedling and adult plant resistance to net blotch (*Pyrenophora teres* f. *teres*) an spot blotch (*Cochliobolus sativus*) in barley. Theor Appl Genet 92:552–558

Stein N, Perovic D, Kumlehn J, Pellio B, Stracke S, Streng S, Ordon F, Graner A (2005). The eukaryotic translation initiation factor 4E confers multialleic recessive bymovirus resistance in *Hordeum vulgare* (L.). Plant J 42:912–922

Tacconi G, Cattivelli L, Faccini N, Pecchioni N, Stanca AM, Valé G (2001) Identification and mapping of a new leaf stripe resistance gene in barley (*Hordeum vulgare* L.). Theor Appl Genet 102:1286–1291

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

Thomas WTB, Powell W, Waugh R, Chalmers KJ, Barua UM, Jack P, Lea V, Forster BP, Swanston JS, Ellis RP, Hanson PR, Lance RCM (1995) Detection of quantitative trait loci for agronomic, yield, grain and disease characters in spring barley (*Hordeum vulgare* L.) Theor Appl Genet 91:1037–1047

Thomsen SB, Jensen HP, Jensen J, Skou JP, Jørgensen JH (1997) Localization of a resistance gene and identification of sources of resistance to barley leaf stripe. Plant Breed 116:455–459

Toojinda T, Baird E, Booth A, Broers L, Hayes P, Powell W, Thomas W, Vivar H, Young G (1998) Introgression of quantitative trait loci (QTL) determining stripe rust resistance in barley: an example of marker-assisted line development. Theor Appl Genet 96:123–131

Toojinda T, Broers LH, Chen XM, Hayes PM, Kleinhofs A, Korte J, Kudrna D, Leung H, Line RF, Powell W, Ramsay L, Vivar H, Waugh R (2000) Mapping quantitative and qualitative disease resistance genes in a doubled haploid population of barley (*Hordeum vulgare*). Theor Appl Genet 101:580–589

Uptmoor R, Wenzel W, Ayisi K, Donaldson G, Gehringer A, Friedt W, Ordon F (2006) Variation of the genomic proportion of the recurrent parent in BC1 and its relation to yield performance in sorghum (*Sorghum bicolor*) breeding for low-input conditions. Plant Breed 125:532–534.

Vales MI, Schon CC, Capettini F, Chen XM, Corey AE, Mather DE, Mundt CC, Richardson KL, Sandoval-Islas JS, Utz HF, Hayes PM (2005) Effect of population size on the estimation of QTL: a test using resistance to barley stripe rust. Theor Appl Genet 111:1260–1270

Van Berloo R, Aalbers H, Werkman A, Niks RE (2002) Resistance QTL confirmed through development of QTL–NILs for barley leaf rust resistance. Mol Breed 8:187–195

Von Korff M, Wang H, Leon J, Pillen K (2005) AB-QTL analysis in spring barley. I. Detection of resistance genes against powdery mildew, leaf rust and scald introgressed from wild barley. Theor Appl Genet 111:583–590

Voss P, Hogers R, Bleeker M, Reijans M, Van de Le T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23:4407–4414

Weerasena JS, Steffenson BJ, Falk AB (2004) Conversion of an amplified fragment length polymorphism marker into a co-dominant marker in the mapping of the *Rph*15 gene conferring resistance to barley leaf rust, *Puccinia hordei* Otth. Theor Appl Genet 108:712–719

Wei F, Gobelman-Werner K, Morroll SM, Kurth J, Mao L, Wing R, Leister D, Schulze-Lefert P, Wise RP (1999) 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 153:1929–1948

Wenzl P, Carling J, Kudrna D, Jaccoud D, Huttner E, Kleinhofs A, Killian A (2004) Diversity Arrays Technology (DarT) for whole-genome profiling of barley. Proc Natl Acad Sci 101:9915–9920

Werner K, Friedt W, Ordon F (2000) Strategies for pyramiding resistance genes against the barley yellow mosaic virus complex based on molecular markers and DH-lines. Proceeding of the 8th International Barley Genetics Symposium, Adelaide, Australia. Vol II, Contrib Papers: 200–202

Werner K, Friedt W, Laubach E, Waugh R, Ordon F (2003a) Dissection of resistance to soil-borne yellow mosaic inducing viruses of barley (BaMMV, BaYMV, BaYMV-2) in a complex breeders cross by SSRs and simultaneous mapping of BaYMV/BaYMV-2 resistance of 'Chikurin Ibaraki 1'. Theor Appl Genet 106:1425–1432

Werner K, Rönnicke S, Le Gouis J, Friedt W, Ordon F (2003b) Mapping of a new BaMMV-resistance gene derived from the variety 'Taihoku A'. J Plant Dis and Prot 110:304–311

Werner K, Friedt W, Ordon F (2005) Strategies for pyramiding resistance genes against the barley yellow mosaic virus complex (BaMMV, BaYMV, BaYMV-2). Mol Breed 16:45–55

Werner K, Friedt W, Ordon F (2006) Localisation and combination of resistance genes against soil-borne viruses of barley (BaMMV, BaYMV) using doubled haploids and molecular markers. Euphytica DOI: 10.1007/s10681-006-9206-4

Weyen J, Bauer E, Graner A, Friedt W, Ordon F (1996) RAPD mapping of the distal portion of chromosome 3 of barley, including BAMMV/BaYMV resistance gene *ym4*. Plant Breed 115:285–287

Williams KJ, Platz GJ, Barr AR, Cheong J, Willsmore K, Cakir M, Wallwork H (2003) A comparison of the genetics of seedling and adult plant resistance to the spot form of net blotch (*Pyrenophora teres* f. *maculata*). Aust J Agric Res 54:1387–1394

Williams KJ, Lichon A, Gianquitto P, Kretschmer JM, Karakousis A, Manning S, Langridge P, Wallwork H (2004) Identification and mapping of a gene conferring resistance to the spot form of net blotch (*Pyrenophora teres* f *maculata*) in barley. Theor Appl Genet 99:323–327

Yan GP, Chen XM (2006) Molecular mapping of a recessive gene for resistance to stripe rust in barley. Theor Appl Genet 113:529–537

Zhong SB, Effertz RJ, Jin Y, Franckowiak JD, Steffenson BJ (2003) Molecular mapping of the leaf rust resistance gene *Rph*6 in barley and its linkage relationships with *Rph*5 and *Rph*7. Phytopathology 93:604–609

Zhou FS, Kurth JC, Wei FS, Elliot C, Vale G, Yahiaoui N, Keller B, Somerville S, Wise S, Schulze-Lefert P (2001) Cell-autonomous expression of barley *Mla*1 confers race-specific resistance to the powdery mildew fungus via a *Rar*1-independent signalling pathway. Plant Cell 13:337–350

CHAPTER 5

CLONING GENES AND QTLS FOR DISEASE RESISTANCE IN CEREALS

BEAT KELLER*, STÉPHANE BIERI, ELIGIO BOSSOLINI AND NABILA YAHIAOUI

Institute of Plant Biology, University of Zürich, Zollikerstrasse 107, 8008 Zürich, Switzerland

Abstract: A number of resistance genes against biotrophic pathogens recently have been cloned from wheat and barley. These include the barley stem rust resistance gene Rpg1, the leaf rust resistance genes Lr10 and Lr21 in wheat and several alleles from the highly diverse powdery mildew resistance loci Mla in barley and Pm3 in wheat. In addition, the durable and recessive mlo gene also conferring powdery mildew resistance as well as the viral resistance genes rym4 / rym5 were isolated from barley. There are many advanced projects in a number of research groups aimed at the isolation of additional resistance genes, including some quantitative trait loci with major effects on resistance against biotrophic and necrotrophic pathogens. The availability of these genes for transgenic approaches as well as the development of highly diagnostic markers to test for the presence of the gene in plants will allow new breeding strategies. Resistance breeding, possibly more than breeding for any other major trait, will benefit enormously and rapidly from this new molecular information: a rapid diagnosis of resistance genes as well as a rational combination of qualitative and quantitative resistance factors based on molecular knowledge will become feasible in the next decade.

1. INTRODUCTION

Plant pathogens are causing losses both by reducing the quantity of the harvested crop as well the quality of the product. It is estimated that in wheat, barley and rice around 10% of the total crop is destroyed by plant diseases. Yield losses reach 30% considering also pests and weeds (Oerke and Dehne 1997). Quality problems of the harvested product can be due to shrivelled seed which are frequently found as a consequence of the infection by leaf pathogens such as mildews, rusts,

*Author for correspondence: bkeller@botinst.uzh.ch

R.K. Varshney and R. Tuberosa (eds.), Genomics Assisted Crop Improvement: Vol. 2: Genomics Applications in Crops, 103–127.

septoria, stagonospora blotch as well as many others. There also are direct quality problems caused by pathogens. In cereals, *Fusarium* species are the major culprit for mycotoxin contamination from the harvested grain, causing economic losses and in the worst case human and animal health problems. Consequently, the improvement of resistance against the *Fusarium* pathogens has been one of the main goals in classical and molecular pathology during the last decade.

Disease resistant plants are essential for sustainable agriculture. Although fungicides are available against many of the important cereal diseases, they are not affordable for farmers in many parts of the world, are sometimes difficult to apply because of weather conditions and can cause environmental problems. Historically, the applied goal of the genetic and later the molecular analysis of naturally occurring disease resistance has been to support and to improve resistance breeding in crop plants. However, the study of the classical and molecular genetics of disease resistance in plants also has been a scientifically rewarding area of basic research, revealing elementary mechanisms in the interactions of plants with pathogens at different levels. These studies have also resulted in the discovery of surprising similarities between pathogen defense in plants and defense mechanisms involved in innate immunity of animals and humans (Nuernberger et al. 2004).

There are recent reviews on the topic of cereal resistance genes which describe in some more detail several aspects which can only briefly be discussed in this chapter (Ayliffe and Lagudah 2004; Keller et al. 2005). Here, we will focus on the cloning of resistance genes and QTL in cereal species in the narrower sense (wheat, barley, rye and oat), but occasionally data and discussion on cereals in the wider sense (including the two crops maize and rice) will be included.

2. SINGLE GENE RESISTANCE VS. QUANTITATIVE RESISTANCE: TWO GENETICALLY DIFFERENT TYPES OF RESISTANCE AND THEIR CONSEQUENCES FOR BREEDING STRATEGIES

The studies on the inheritance of resistance revealed that there are two clearly distinct genetic mechanisms for disease resistance: monogenic resistance is based on single genes or at least single genetic loci whereas quantitative resistance depends on two or more genes. Genes involved in quantitative resistance have additive genetic effects, with none of them individually conferring sufficient resistance to be of agronomic interest. In most cases (for some exceptions see below), it was found that single resistance genes conferred complete resistance but were only active against certain races of the pathogen, i.e. they showed a genetic interaction with genes from the pathogen. If such a gene conferred a complete resistance against all known pathogen races, large-scale cultivation frequently resulted in the appearance of new, virulent pathogen races possibly created by mechanisms of recombination

or mutation. Alternatively, these virulent races might have been present at very low frequencies in the pathogen populations already before introduction of a particular resistance gene. Thus, single gene based, race-specific resistance is often rapidly overcome by new pathogen races, thereby limiting its use as a single defense line in a crop species. A possible use of resistance genes in a more sustainable way in breeding is based on combination of genes (McDonald and Linde 2002). This is achieved by combining ("pyramiding") several resistance genes in the same cultivar, or by using line or cultivar mixtures. There, different *R* genes are present in different plant lines and the combined planting of the lines results in an overall improvement of resistance without putting an enormous selection pressure on the pathogen.

Single gene, race-specific resistance is based on a direct or indirect recognition event between the product of the plant resistance gene and the effector protein encoded by the avirulence gene of the pathogen. This interaction is studied intensively both in model plants as well as in crop species and in some cases is now quite well understood (see e.g Schulze-Lefert and Bieri 2005 and references therein). At the genetic level, resistance genes are often clustered at some loci in the genome or exist as different alleles, each allele conferring resistance towards a specific set of pathogen races carrying the corresponding avirulence allele. In contrast to this resistance based on major single genes, quantitative resistance rarely shows a genetic interaction with the pathogen and thus is effective against all races of a particular pathogen species. It slows down the disease development by increasing latency period, lowering sporulation frequency or other parameters related to the development of the epidemic. Quantitative resistance shows a relevant environmental dependence which makes breeding for it cost- and labor intensive in large-scale breeding programmes, particularly if a number of quantitative traits have to be incorporated into the same elite genetic material.

Although plant pathogen interactions are sometimes simplified by stating that quantitative resistance is durable, whereas this is not the case for single-gene based resistance, one should keep in mind that this is not always the case. There are several examples of single genes which have been effective for many years: these include the *Sr2* stem rust resistance gene in wheat, the *mlo* powdery mildew resistance gene in barley and the *N* gene against tobacco mosaic virus. In addition, there are indications that certain quantitative resistances can also erode and become less effective, albeit not in the rapid form as observed for race-specific resistance genes in the boom-and-bust cycles (McDonald and Linde 2002).

Resistance breeding in crop plants depends on both major *R* gene as well as quantitative resistance. Monogenic resistance genes are easy to work with but are frequently not durable as described above due to rapid selection of pathogen races which overcome the specific resistance. Consequently, quantitative resistance is an attractive alternative although it is difficult to handle in a breeding program if there are no diagnostic markers available for selection of the relevant chromosomal regions. The application of molecular markers has allowed the

genetics of quantitative resistance to be determined and quantitative trait loci involved in resistance have been identified and mapped. The focus of current research is on cloning a large number of quantitative as well as major resistance genes from different cereal species and the understanding of their function and evolution.

3. MAP-BASED CLONING OF CEREAL DISEASE RESISTANCE GENES

Currently, the only grass genome which is completely sequenced is the rice genome. It was proposed early on as the model genome for all grass species based on large-scale synteny between rice and other species. The possible use and the pitfalls of using rice as a genomic model for gene cloning in other grasses have been described in two recent reviews and shall not be repeated here (Bennetzen and Ma 2003; Xu et al. 2005). In summary, rice provides an excellent model to saturate genetic regions of interest with molecular markers. In addition, it can provide candidate genes for further study, but it has to be kept in mind that resistance genes are poorly conserved at orthologous positions in grass genomes (Leister et al. 1998; Guyot et al. 2004). The rice genome can not replace barley and wheat genomic tools as, at least for resistance genes, a physical contig spanning the locus has to be derived from the species, or even better, from the cultivar of interest. An alternative "bridging" genome could possibly be the genome of *Brachypodium distachyon*, a species which is more closely related to wheat than rice and will be sequenced in the near future.

Once the first resistance genes had been cloned from model organisms, it became clear that many of them encoded proteins with conserved domains such as nucleotide binding site (NBS) and leucine rich repeat (LRR) domains. Many studies subsequently described the isolation of resistance gene analogs (RGAs, genes encoding proteins with domain structures very similar to those encoded by functional resistance genes) using PCR amplification based on primers developed from conserved regions. A large number of cereal RGAs were subsequently isolated and characterized from several species (see e.g. Leister et al. 1998). It is disappointing that until now none of these RGAs was proven to be an active resistance gene by functional analysis in wheat and barley. Amazingly enough and in a sort of contradiction to the unsuccessful RGA approaches, the recently cloned cereal resistance genes have mostly turned out to encode NBS-LRR type of proteins. The failure of the RGA strategy for cloning of resistance genes is possibly due to the large number of RGAs present in cereal genomes. In rice, around 600 RGAs were found, and this class of genes accounts for approximately 1% or more of the rice genes (Bai et al. 2002; Monosi et al. 2004). A similar number can be expected for the barley genome and each of the three genomes of hexaploid wheat. Thus, the sheer number of RGAs makes the identification of a specific gene difficult. Consequently, all the known wheat and barley resistance genes have been isolated by map-based cloning and not by homology with the rice genome or the RGA approach.

Map-based cloning in cereal genomes is challenging because of the large genome sizes (5×10^9 bp in barley, 1.6×10^{10} bp in bread wheat) and the high content of repetitive DNA making chromosome walking tedious. The steps towards cloning of a gene include (1) the establishment of a large population segregating for the trait of interest, (2) high resolution mapping of the gene using molecular markers and (3) establishment of a physical contig of BAC clones containing flanking markers of the gene. Both for marker development as well as physical mapping, the availability of genomic resources is crucial. In wheat and barley, high density maps and a large number of markers, mainly highly polymorphic microsatellite markers as well as EST sequences are available (reviewed in Keller et al. 2005). In an analysis for wheat published by McIntosh et al. (2003), there were 5537 RFLP loci, 1620 microsatellites and 2049 protein loci and genes controlling phenotypic traits. This makes the wheat genome one of the most densely mapped plant genomes, except for those with a completely known genome sequence. More information is available on the Graingene web site http://wheat.pw.usda.gov/GG2/index.shtml).

There is a growing list of wheat and barley BAC libraries available for physical mapping. For barley, there are BAC libraries from cultivars Morex and Cepada Capa, with several others currently being made (Yu et al. 2000; Isidore et al. 2005a). Two libraries from hexaploid wheat cultivar Chinese Spring are available as spotted filter sets for hybridization screening (Allouis et al. 2003, Shen et al. 2005). In addition, pooled PCR screenable BAC libraries have been constructed from hexaploid wheat cv. Glenlea (Nilmalgoda et al. 2003), as well as from single chromosomes (Safar et al. 2004) and from pooled chromosomes (Janda et al. 2004) of cv. Chinese Spring. The BAC library of hexaploid Chinese Spring as well as the chromosome 3B specific BAC library have recently become available for general use (see http://cnrgv.toulouse.inra.fr/ENG/collection.html). Genomic studies in the complex hexaploid wheat genome can be simplified by specific analysis of one of the three individual wheat genomes. These can be accessed through the diploid A, B, and D-genome donors. Therefore, a number of BAC libraries have been constructed from diploid ancestors of wheat, e.g. for the A genomes of *T. monococcum* (Lijavetzky et al. 1999) and *T. urartu* (Akhunov et al. 2005), for the wild grass which is most similar to the B genome (*Aegilops speltoides*, Akhunov et al. 2005) and for the D genome (*Aegilops tauschii*, Moullet et al. 1999, Luo et al. 2003, Akhunov et al. 2005), as well as from the tetraploid AB genomes (Cenci et al. 2003). Given the observed large differences in haplotypes of different varieties in barley (Scherrer et al. 2005) and wheat (Isidore et al. 2005b), it will be very helpful for the cloning of a specific gene if several or many BAC libraries are available. Then, the BAC library from the cultivar which has the most closely related haplotype can be used. Map-based cloning has resulted in the isolation of a number of fungal disease resistance genes in wheat and barley. The isolated genes are listed in Table 1. In addition, the table also lists the cloned maize and rice disease resistance genes against maize rust and rice blast/bacterial blight. Below, we discuss in more detail the cloning of the wheat and barley genes.

Table 1. Cloned disease resistance genes in wheat, barley, rice and maize

Species	Gene	Disease	Type of gene	Reference
Wheat	*Lr10*	Leaf rust	NBS-LRR	Feuillet et al. 2003
Wheat	*Lr21*	Leaf rust	NBS-LRR	Huang et al. 2003
Wheat	*Pm3a-g*	Alleles against specific races of powdery mildew	NBS-LRR	Yahiaoui et al. 2004 Srichumpa et al. 2005 Yahiaoui et al. 2006
Barley	*mlo*	Powdery mildew	7 transmembrane domain protein	Büschges et al. 1997
Barley	*Mla* alleles *1,6,7,10,12,13*	Alleles against specific races of powdery mildew	NBS-LRR	Zhou et al. 2001 Halterman et al. 2001 Halterman et al. 2003 Shen et al. 2003 Halterman and Wise 2004
Barley	*Rpg1*	Stem rust	Tandem kinase	Brueggemann et al. 2002
Maize	*Rp1-D*	Leaf rust	NBS-LRR	Collins et al. 1999
Rice	*Xa1*	Leaf blight	NBS-LRR	Yoshimura et al. 1998
Rice	*xa5*	Leaf blight	TFIIAγ	Iyer and McCouch 2004 Jiang et al. 2006
Rice	*Xa-21*	Leaf blight	RLK	Song et al. 1995
Rice	*Xa-26*	Leaf blight	RLK	Sun et al. 2004
Rice	*Pib*	Rice blast	NBS-LRR	Wang et al. 1999
Rice	*Pi-ta*	Rice blast	LRD	Bryan et al. 2000
Rice	*Pi-9*	Rice blast	NBS-LRR	Qu et al. 2006

3.1. Powdery Mildew Resistance Genes in Cereals: The Well Studied Cases of the Barley *Mla* and Wheat *Pm3* Allelic Series of Resistance Genes

Barley powdery mildew (*Blumeria graminis* f. sp. *hordei*) is a biotrophic fungal pathogen of economical importance in most barley growing areas. Breeding for resistance against this disease already has a long history, such that a high number of resistance genes (denominated *Ml* genes) have been identified in the cultivated and wild gene pools of barley (Jorgensen 1993; Weibull et al. 2003). The *Mla* locus (first described in the cultivar "Algerian" = *Mla1*) was shown to encode a large number of alleles specifying race specific resistances that are distinguishable based on the differential response of barley lines towards a set of different mildew isolates. A total of 32 *Mla* alleles have been defined genetically (*Mla1* to *Mla34*; *Mla4* was shown to be an independent locus and renamed *Mlk1*; *Mla15* is identical with *Mla7*) (Jorgensen 1993; Weibull et al. 2003), and many more can be expected in as yet uncharacterized germplasm. The *Mla* locus is thus among the most variable disease resistance loci found in plants.

Initially, a physical map spanning the *Mla* locus was established using the cv. Morex BAC library, which was the only barley BAC library available at that time (Wei et al. 1999). The BAC contig encompassing 261 kb represented 0.25 cM in an F_2 population of 1800 individuals used for high resolution mapping. The contig was fully sequenced and annotated (Wei et al. 2002). The cv. Morex sequence revealed the presence of 8 NBS-LRR resistance gene homologs (RGHs) that were grouped in three families based on sequence similarity. A function in disease resistance could not be assigned to any of these candidate genes as Morex does not express a known powdery mildew resistance gene at the *Mla* locus. Thus, a functional *Mla* gene had to be isolated from another barley accession. In a second step, this was done concomitantly, yet independently, for both *Mla1* and *Mla6* using different approaches.

To isolate *Mla1*, a cv. Algerian derived *Mla1* near-isogenic line was γ-ray mutagenized to recover *Mla1* mutants. Two mutants showed an altered RFLP banding pattern when hybridized with selected probes derived from the Morex contig. One of these probes was a RGH and was subsequently used to isolate the *Mla1* sequence from a cosmid library of a *Mla1* expressing barley line. The *Mla1* gene identity was confirmed through genetic complementation in a single cell transient expression assay (see below). *Mla1* encodes a coiled-coil nucleotide-binding-site leucine-rich-repeat (CC-NBS-LRR) type of R-protein of 958 amino acids with 11 imperfect LRRs (Zhou et al. 2001). To isolate *Mla6*, the corresponding LRR regions of different Morex RGH family members were amplified from a *Mla6* expressing cv. Franger derived near-isogenic line. These LRR fragments were then hybridized to a cDNA library developed from the same accession. Only probes from one RGH family (the same as for *Mla1*) identified clones in the library. These cDNAs corresponded to one potentially full-length CC-NBS-LRR protein of 956 amino acids. A genomic clone from the same gene was isolated through a PCR based screen of a cosmid library constructed from the same accession. The identity of *Mla6* was confirmed similarly as *Mla1* (Halterman et al. 2001). The *Mla1* and *Mla6* gene function was later corroborated in transgenic barley. An 8kb genomic fragment of either one of the genes was able to confer complete, race specific resistance that was indistinguishable from the respective donor lines (Bieri et al. 2004).

Subsequently, further *Mla* alleles were isolated more easily: *Mla12* was isolated from a cosmid library (Shen et al. 2003) and *Mla13* from a cDNA library (Halterman et al. 2003) using the previously isolated *Mla* genes as probes. *Mla7* and *Mla10* were amplified by a RT-PCR strategy with primers designed from sequences conserved in the previously isolated genes (Halterman and Wise 2004). Sequence comparison of *Mla* coding regions revealed that they are closely related (>93% overall identity) and represent gene variants (probably alleles) of a common progenitor. The N-terminal part of the proteins is much more conserved (>98% identical) than the C-terminal part (>82%). Accordingly, the region determining the difference in recognition specificity between Mla1 and Mla6 could be mapped in a domain swap experiment to the C-terminal part of the protein encompassing the non-LRR C-terminus plus

three to nine LRRs (Mla6 and Mla1, respectively; Shen et al. 2003). The same region also has a number of hypervariable residues that likely are under diversifying selection.

In bread wheat, the *Pm3* locus encodes one of the earliest identified allelic series for wheat disease resistance. Ten *Pm3* alleles (*Pm3i* to *j*) have been genetically identified in wheat lines or cultivars originating from four continents (Zeller and Hsam 1998). Each of them confers race-specific resistance to wheat powdery mildew (*Blumeria graminis* f. sp. *tritici*). The first of these alleles identified at the molecular level was the *Pm3b* allele which was cloned using a positional cloning strategy (Yahiaoui et al. 2004). Two wheat species with lower ploidy levels, the diploid wheat *T. monococcum* and the tetraploid *T. turgidum* cv. *durum*, provided models for the genome of hexaploid wheat and allowed the establishment of a physical contig spanning the *Pm3* locus. Although the haplotypes at the *Pm3* locus differed dramatically between the three levels of ploidy, a large resistance gene-like family was consistently found at the *Pm3* locus and a candidate gene cosegregating with *Pm3b* was identified. A single mutant showing no obvious large deletion at the *Pm3* locus showed a single base-pair deletion in the coding region of the candidate gene, and no mRNA could be detected from the candidate gene by RT-PCR in this mutant. In the transient assay described below for powdery mildew resistance genes, the identity of the candidate gene as *Pm3b* was demonstrated by its race-specific interaction with the powdery mildew pathogen (Yahiaoui et al. 2004).

For the isolation of the remaining *Pm3* genes, haplotype analysis of ten lines carrying different *Pm3* alleles was performed. In all these lines a conserved genomic region delimited by markers co-segregating with *Pm3b* was identified, including a structurally very well conserved, *Pm3b*-like gene (Srichumpa et al. 2005). Based on this haplotype conservation, six additional *Pm3* resistance alleles (*Pm3a*, *c*, *d*, *e f*, and *g*) were isolated using a PCR-based strategy (Srichumpa et al. 2005; Yahiaoui et al. 2006). The *Pm3* resistance alleles encode a coiled-coil nucleotide binding site leucine rich repeat type of protein. Haplotype conservation in *Pm3* lines and high sequence conservation (more than 97% sequence identity) observed between the *Pm3* resistance genes indicated that all *Pm3* specificities form a true allelic series. Highly specific functional markers derived from the allelic sequences now allow for the detection of individual *Pm3* alleles with complete accuracy in wheat breeding programs (Tommasini et al., 2006). In addition to the functional alleles, a susceptible variant of the *Pm3* genes was identified. The comparison of the sequence of the *Pm3* resistance alleles with the susceptible allele P*m3CS* provided important information on the evolutionary processes that generated the resistance alleles. One group of four *Pm3* resistance alleles showed few, clearly delimited, polymorphic sequence blocks of ancient origin, possibly derived from gene conversion. A second group of three alleles differed from *Pm3CS* by only two to five mutations, all non-synonymous and all in the LRR-encoding region. Transient transformation assays confirmed that these few differences in the LRR domain of *Pm3* proteins are responsible for the specificity of resistance against different powdery mildew isolates. Moreover, the very high sequence conservation

between these alleles and the susceptible *Pm3CS* allele together with the absence of synonymous mutations between all these genes indicated a recent evolution of the *Pm3* resistance genes probably after wheat domestication 10,000 years ago (Yahiaoui et al. 2006).

Interestingly, both the *Mla* and *Pm3* genes are located in the distal region of chromosome group 1 (1HS and 1AS, respectively) and both loci are characterized by a large number of functionally different alleles. However, the two loci are not orthologous as the *Mla* probes derived from barley sequences map about 5 cM from the wheat *Pm3b* gene in a wheat genetic map (Zhou et al. 2001 and our unpublished results). In addition, the two genes are not similar at the sequence level (except for the overall domain structure of CC, NBS, LRR domains). We conclude that *Mla* and *Pm3* represent a case of convergent evolution where quite different genes were recruited to confer resistance against very similar pathogens. It is relevant to study at the molecular level the direct or indirect mechanisms of MLA and PM3 protein interaction with fungal avirulence gene products and the subsequent signal transduction. It is also interesting to note that two different genetic loci have evolved into major resistance genes in the two closely related crop species wheat and barley. This strongly suggests that for most resistance loci it will be necessary to isolate them from the crop species where they are active as there is very low functional similarity between orthologous genes even in closely related species. Thus, the isolation and comparison of the *Mla* and *Pm3* genes has resulted in exciting new findings and demonstrates that, in addition to putative practical applications, it is well worth isolating resistance genes from crop plants instead of limiting this type of work to model species.

3.2. The *mlo* Powdery Mildew Resistance in Barley

Classical monogenic resistance is race-specific and genetically dominant or co-dominant. The *Mla* and *Pm3* genes described above are typical examples of such resistance genes. In barley, there is a second and highly unusual recessive resistance against powdery mildew based on the absence of the *mlo* gene. Absence of *Mlo* confers a broad spectrum resistance which is active against most powdery mildew races in barley and has been durable for a long time. The *mlo* haplotype introgressed in many European spring barley lines is derived from a natural gene silencing event found in Ethiopian landraces. The complex molecular basis of this mutation has recently been identified and is described in Piffanelli et al. (2004). Resistance based on the *mlo* gene was obtained in all powdery mildew susceptible barley lines which were mutagenized and screened for resistance. Therefore, it seems that loss of the *Mlo* gene is sufficient to confer this broad spectrum resistance which is based on the formation of cell wall appositions. The *mlo* gene was actually the first gene cloned from a large cereal genome by map-based cloning (Büschges et al. 1997). This was based on a segregating F_2 population with more than 2000 individuals and the identification of closely linked AFLP markers. In this early time of gene cloning in barley there was no large insert library available yet and the authors made a YAC

library from cultivar Ingrid which has an *Mlo* genotype. The *Mlo* gene could be mapped to a short physical sequence of approximately 30 kb, revealing only one good candidate gene for *Mlo*. PCR-based sequencing of the *Mlo* gene candidate in several independently generated *mlo* mutants from different genetic background revealed mutations in the coding region in all of them. This mutant analysis as well as experiments detecting intragenic recombination events demonstrated the identity of the cloned gene as *Mlo*. The MLO protein is encoded by 12 exons and contains 533 amino acids, with a number of putative transmembrane domains. The analysis of MLO function in barley as well as in Arabidopsis is currently a very active area in research and ideally this type of resistance could also be developed in other cereal species, particularly wheat.

3.3. Rust Resistance Genes from Wheat and Barley

There are three rust diseases in wheat and barley: leaf rust, stem rust and stripe rust. Although there are more than 200 resistance genes described against these diseases in the two crops (see e.g. McIntosh et al. 1995), until now only the three leaf rust resistance genes *Lr1*, *Lr10* and *Lr21* have been cloned from wheat. In addition, one stem rust resistance gene, *Rpg1*, has been isolated from barley.

Similar to wheat genes *Pm3* and *Lr10* as well as the barley *Mla* gene, the wheat *Lr21* gene is located in the distal region of the short arm of a group 1 chromosome, in this case 1DS. It was isolated by map-based cloning, using a shuttle mapping strategy between the diploid species *Aegilops tauschii*, the D-genome donor of bread wheat, and the hexaploid bread wheat genome (Huang et al. 2003). *Lr21* had originally been introgressed into bread wheat from an *Ae. tauschii* line collected in Iran. High-resolution mapping was done in the hexaploid recipient wheat, whereas the large insert clones (in this case cosmid clones) were generated from the diploid *Ae. tauschii* donor line. Although one recombinant, which later turned out to be intragenic, complicated the identification of a candidate gene considerably, Huang et al. (2003) could identify a promising candidate gene which was then confirmed to be *Lr21* based on stably transformed transgenic wheat lines. The *Lr21* gene encodes a 1080 amino acid protein of the NBS-LRR type. In the amino terminal domain it has a highly unusual 151 amino acid domain which is not found in other NBS-LRR proteins. The LR21 protein has only 13 imperfect leucine-rich repeats, which is relatively short compared to other NBS-LRR type of resistance genes, but very similar in size compared to other LRR domains in plant proteins (De Lorenzo et al. 2001; Baumberger et al. 2001).

The *Lr10* leaf rust resistance gene is located on chromosome 1AS and was described in the gene pool of hexaploid wheat but not in any diploid relative. A high resolution genetic map was established based on 3120 F2 plants. Closely linked markers were used to identify BAC clones from the region of interest. At the time of this work, there were only BAC libraries available from the A^m genome from *T. monococcum* and the D-genome from *Ae. tauschii*. As the *Lr10* gene is located on chromosome 1AS, the library from *T. monococcum* DV92 was used to establish

a physical contig around the gene (Stein et al. 2000). Two chromosomal walking steps were necessary to establish a contig including flanking markers. These steps were difficult because of the highly repetitive nature of the wheat genome, and the development of specific probes for the isolation of overlapping clones was time-consuming. Sequencing of the relevant part of the contig revealed the presence of two NBS-LRR candidate genes between the flanking markers (Wicker et al. 2001). It later turned out that the choice of the diploid wheat for establishing a contig had been a lucky one, as the line *T. monococcum* DV92 has a haplotype at the *Lr10* locus which is very similar to the one in *Lr10* containing hexaploid wheat line (Feuillet et al. 2003). It was also found later that most hexaploid wheat lines have a haplotype with a large deletion at the *Lr10* locus and the discovery of the gene would have been impossible in a BAC contig from such a line (Isidore et al. 2005b).

Using EMS mutagenesis, three independent mutations were identified in one of the candidate genes, demonstrating that the affected gene is *Lr10* (Feuillet et al. 2003). Transformation with the two candidate genes only resulted in resistant and fertile plants when both candidate genes were used, a fact that is still not completely understood (Feuillet et al. 2003). The *Lr10* gene encodes a CC-NBS-LRR type of protein with 919 amino acids and 14 imperfect leucine-rich repeats. Amazingly, diversifying selection was shown to act on the N-terminus of the gene, whereas the LRR domain was more conserved. This contrasts with the finding in other resistance genes (McDowell et al. 1998; Meyers et al. 1998). *Lr10* is clearly different in sequence compared to other cloned cereal resistance genes. In particular, it is more closely related to the Arabidopsis *Rpm1* bacterial resistance gene than to the second wheat leaf rust resistance gene cloned, *Lr21* (Feuillet et al. 2003). *Lr10* is similar to *Rpm1* in two more aspects: first, both genes are present only as one copy in the genome which contrasts with the observed gene clusters at many other resistance loci in cereals (e.g. *Mla*, *Pm3*, *Lr21*). Second, for both genes there is a characteristic presence or absence polymorphism in the gene pool (Isidore et al. 2005b): a wheat cultivar either has an *Lr10* gene or a very closely related gene or it has a large deletion, very similar to the *Rpm1* locus. The molecular relevance of these similarities for resistance gene function of these two genes remains to be determined. Very recently, a third leaf rust resistance, *Lr1*, has been cloned but this work has not yet been published (S. Cloutier, B. McCallum, M. Jordan, C. Feuillet, B. Keller, unpublished data).

A single rust resistance gene has been isolated from barley: the *Rpg1* stem rust resistance gene located on chromosome 7HS which confers a highly valuable and durable (since 1942) resistance in many North America cultivars. The gene orginates from an old Swiss landrace from the Canton Lucerne and was cloned by Brueggeman et al. (2002). The story of the *Rpg1* cloning is one of several good case studies revealing the limits of the use of the rice genome for cloning of disease resistance genes (summarized by Ayliffe and Lagudah 2004). For *Rpg1* cloning, a high resolution map of 8,518 gametes was used as well as a BAC contig of 330 kb established by chromosome walking from the cultivar Morex which

contains *Rpg1*. High resolution mapping delimited the interval containing *Rpg1* to 110 kb. Two candidate genes with sequence similarity to receptor-kinase like genes were identified and one was putatively assigned as being *Rpg1* by comparing allelic sequences from susceptible and resistance cultivars. The RPG1 protein is encoded by 14 exons and contains 837 amino acids (Brueggeman et al. 2002). For a resistance protein it has a highly unusual protein structure with two tandem kinase domains but lacking any LRRs. It remains to be seen if it is a membrane-bound protein or localized intracellularly. The proposed *Rpg1* gene was transformed into the susceptible barley cultivar "Golden Promise", making it resistant to stem rust and confirming that the candidate gene was *Rpg1* (Horvath et al. 2003). Interestingly, and similarly to *Lr10* (Feuillet et al. 2003), transgenic plants showed an increased resistance compared to the line carrying the endogenous gene. While this can be attributed to overexpression by a heterologous promoter in the case of *Lr10*, *Rpg1* was under control of its own promoter and the reason for increased resistance in transgenic plants remains unclear.

3.4. The *Rp1* Common Rust Resistance Locus in Maize

The *rp1* locus conferring resistance to maize common rust (*Puccinia sorghi*) is a complex locus on the distal end of the short arm of maize chromosome 10. Genetic studies showed that different *Rp1* specificities could recombine, demonstrating the complexity of this locus (Saxena and Hooker 1968). Recombination events in the *rp1* complex have been associated with the generation of variants with novel phenotypes. Some of these variants have non-specific reactions to maize rust and display lesion mimic phenotypes (Hu et al. 1996), some confer a reduced level of resistance, others show novel race specificities (Richter et al. 1995). The *Rp1-D* specificity was isolated using a transposon tagging strategy combined with a candidate gene approach (Collins et al. 1999). In a first step, resistance gene analogs were amplified from the maize genome and one of them, PIC20 hybridized with a small gene family at the *rp1* locus. For the final identification of *Rp1-D*, independent mutants using *Mutator* and *Ds* transposons were obtained for the *Rp1-D* gene. In one family of *Rp1-D* mutants generated using the *Mutator* element, a *Hin*dIII restriction fragment of 5 kb hybridized to the *Mutator* probe and cosegregated with the mutant allele. In the *Ds* mutants of *Rp1-D* , the PIC20 probe was used to detect presence or absence of insertions at the *Rp1-D* locus. PIC20 detected insertions of the *Ds* transposon in a *Rp1D*-mutant and excision in three *Ds* revertants. The mutated sequences were then isolated from λ genomic DNA libraries and were found to correspond to the same gene. The *Rp1-D* gene encodes a NBS-LRR type of protein of 1292 amino acid (Collins et al. 1999). In contrast to a majority of NBS LRR proteins, RP1-D does not show a TIR or a leucine zipper motif at the N terminus. The *Rp1-D* haplotype was found to contain 9 paralogous genes. Only one is functional and is located at the most distal end of the cluster (Smith et al. 2004). Genes at the *rp1* locus often mispair at meiosis which should result in the creation of new haplotypes or new recombinant genes. Recently, recombination events have been identified at the *rp1*

locus which resulted in the generation of a novel specificity (Smith and Hulbert 2005). Based on haplotype analysis of the recombinants, the novel resistance is believed to be the result of a combination of existing *Rp1* genes by intergenic recombination rather than the result of the creation of a new gene.

3.5. Viral Resistance Genes *rym4 / rym5* in Barley

In winter barley, viral infection by the two bymoviruses *Barley Yellow Mosaic* (BaYMV) and *Barley Mild Mosaic Virus* (BaMMV) can cause considerable yield loss. This group of viruses is transmitted by a soil-borne fungus, *Polymyxa graminis*, which is almost impossible to fight with fungicides. In the barley gene pool, there are recessive resistance genes against BaYMV and BaMMV such as the *rym4 / rym5* complex on chromosome 3HL. Recently, these genes were shown to be allelic and they were isolated by a map-based cloning strategy (Stein et al. 2005). First, a contig of 650 kb was established at the *Rym4/Rym5* locus. It did not span the gene despite considerable efforts to extend the contig to include flanking markers on both sides of the resistance gene. The region turned out to have (i) a suppressed recombination and (ii) to be extremely gene poor. It was completely sequenced and represents the largest contiguous sequence in barley and wheat that has been analysed todate (Wicker et al. 2005). A gene encoding a typical eukaryotic translation initiation facter 4E was located on the contig and represented a very promising candidate for the resistance gene. When the gene for *Hv-eIF4E* from a susceptible barley line was transformed into a virus resistant cultivar, the transformants turned out to be susceptible, confirming that the cloned gene was the susceptibility factor. Comparative sequencing of *Hv-eIF4E* in 56 barley accessions revealed a low degree of polymorphism in the gene pool. Both for the *rym4* and *rym5* alleles two characteristic and diagnostic polymorphisms were detected, suggesting an independent origin of the two resistance alleles. The amino acid changes caused by the DNA polymorphisms were all located on the surface of the protein model for eIF4E in close proximity of the cap-binding domain (Stein et al. 2005). The viral RNAs encode a "virus genome-linked protein" (VPg) with binding affinity for eIF4E but the exact molecular mechanism of this interaction and its consequences resulting in susceptibility or resistance remain elusive. Interference with viral translation or viral movement are both possible explanations. The characterization of amino acids putatively involved in resistance or susceptibility should greatly facilitate the determination of the molecular basis leading to the particular outcome of the virus-host interaction. The results described by Stein et al. (2005) clearly suggest that each of the *eIF4E* genes in cereal genomes represents a good candidate for having allelic variants conferring viral resistance.

3.6. Ongoing Single Gene Isolation Projects in Wheat, Barley and Rye

In addition to the genes which have already been cloned from the complex wheat and barley genomes, there is an increasing number of projects targeting other

relevant disease resistance loci in wheat, barley and rye. This work is profiting from the improved genomic resources described earlier, as well as from an increased interest in genetic resistance of cereals. An exhaustive description of all the ongoing projects would be beyond the scope of this chapter, but a few of the cloning projects which are well advanced will be described in a bit more detail and should give a representation of the type of genes that will possibly be cloned in the near future.

The short arm of rye chromosome 1RS is homoeologous to the group 1S chromosome arms in barley and wheat and carries a number of economically important resistance genes against stem, leaf and stripe rust as well as powdery mildew. The 1RS chromosome translocation is present in many bread wheat cultivars and, therefore, these resistance genes are present in a broad set of elite wheat varieties. Mago et al. (2005) have recently described their progress in the isolation of the *Sr31*, *Lr26* and *Yr9* genes on chromosome 1RS. They used two wheat lines each with a different rye translocation (one of which did not carry the resistance genes) for high resolution mapping. This elegant approach circumvented the problem of the lack of recombination within the rye segment in a normal cross between two wheat lines, of which only one carries the translocation. A first result from recombination and deletion mapping experiments was the observation that the three resistance activities are caused by three physically distinct, but closely linked genes. Although a homolog of the barley *Mla* gene mapped very closely to the resistance gene cluster, a deletion mutant lacking all three resistance activities retained all *Mla* markers. This suggests that *Sr31*, *Lr26* and *Yr9* are not related to *Mla*. Nevertheless, the high density map around the genes as well as the physical knowledge of the barley *Mla* region should simplify the cloning of these genes in the near future.

The isolation of an additional stem rust resistance gene in wheat, *Sr2*, is well advanced. *Sr2* confers partial and recessive resistance against stem rust, being analyzed genetically as a QTL (see below). The gene has confered durable resistance over many decades and thus represents a very important target for cloning. Because *Sr2* can be scored with certainty under suitable environmental conditions, it can be mapped at high resolution similar to single genes. *Sr2* is completely linked, or pleiotropic, with pseudo-black chaff, a phenotypic marker which causes dark pigmentation on the stem internodes and glumes. *Sr2* maps to chromosome 3BS – a region where QTL for *Fusarium* resistance and Stagonospora resistance also are found as well as the leaf rust resistance gene *Rph7* in barley (see below). This region is the focus of an international effort to assemble and map large physical contigs for sequencing. The cloning of *Sr2* will be greatly supported by this effort. Using mapping information from the rice genome, Kota et al. (2006) have recently mapped *Sr2* at high resolution with the closest flanking markers spanning an interval of 0.4 cM, and the same authors started to build a physical contig from this locus (Spielmeyer and Lagudah, personal communication).

In barley, there are also many advanced map-based cloning projects. Only three of them will be mentioned. In contrast to wheat, no leaf rust resistance gene has yet been isolated from barley. However, the two genes *Rph5* and *Rph7* have been

located to very small genetic intervals in different regions of chromosome 3 HS (Mammadov et al. 2005; Scherrer et al. 2005). The *Rph5* gene has been located in a small 0.3 cM genetic interval in the barley map and the syntenic rice genomic region was established (Mammadov et al. 2005). The *Rph7* gene was located in a genetic interval of 0.13 cM and a physical BAC contig containing the flanking markers was established from a barley cultivar containing the *Rph7* gene (Scherrer et al. 2005). The complete BAC sequence was determined but no obvious candidate gene could be identified. Thus, *Rph7* possibly represents a new type of resistance gene. Finally, the advance in the isolation of the barley *Rdg2a* gene against the hemi-biotrophic seed-borne pathogen *Pyrenophora graminea* (leaf stripe) should be mentioned. *Rdg2a* was located in a 0.14 cM interval between two resistance gene analogs and might therefore be a member of a larger RGA cluster at this locus (Bulgarelli et al. 2004).

There is also progress in the characterization of genes involved in resistance against necrotrophic cereal pathogens. These pathogens are often characterized by production of toxins which result in necrosis of the host tissue. Genetically, sensitivity is conferred by a dominant gene, suggesting that a specific receptor/transporter is responsible for the observed sensitivity. Thus, the recessive allele is desired in breeding material. The wheat *Tsn1* gene belongs to this class of genes and confers sensitivity to the proteinaceous, host-selective toxin Ptr ToxA produced by the tan spot fungal pathogen (*Pyrenophora tritici repentis*). The group of J. Faris has recently mapped this gene at high resolution in an interval of 0.8 cM and started to establish BAC contigs from this region (Lu et al. 2006; Lu and Faris 2006). The same group has also made significant progress in the isolation of the *Snn1* sensitivity gene for a toxin produced by the *Stagonospora nodorum* glume blotch pathogen. This gene has been mapped in a population of 2,050 gametes to a genetic interval of 0.9 cM. Using the current genetic resolution, a BAC contig of 250 kb cosegregates with the *Snn1* gene (J. Faris, personal communication).

4. CLONING OF RESISTANCE QTL FROM CEREALS

Breeding for quantitative, horizontal resistance is a promising approach to achieve durable resistance. Quantitatively acting defence genes do not confer a high degree of resistance to pathogens, but they slow down disease development. The molecular isolation of QTLs from cereal genomes is not a prerequisite for breeding, but an effective breeding program would profit enormously from diagnostic and very closely linked molecular markers for QTLs. Positional cloning requires intermediate steps, like the generation of molecular markers which are genetically tightly linked to the target locus. These markers can then be used in traditional breeding programmes (non-transgenic) to introgress the QTL into different genetic backgrounds. In addition, QTL cloning and functional analysis will be essential to learn more about the molecular mechanisms that result in partial resistance. As described above, gene isolation in cereal genomes is difficult, and QTL isolation is further complicated by the small phenotypic effect of a specific locus. Dissection of the phenotypic

variance, validation of single loci in recombinant, near isogenic backgrounds and testing in replicated trials are essential steps to achieve a precise map location of a specific QTL. Near isogenic lines for a specific QTL represent ideal biological material to also characterize the molecular or physiological mechanisms that lead to enhanced resistance. Recently, an excellent discussion on QTL cloning in plants was published and more details on the technology for QTL isolation are found there (Salvi and Tuberosa 2005).

The development of collections of near isogenic genotypes sharing the same QTLs for horizontal resistance and differing for a number of major resistance genes could result in highly valuable breeding material and be an efficient combination of the two different genetic resources. Thus, the high resolution mapping and cloning of resistance QTLs will form the basis of new strategies for resistance breeding which could not systematically be followed in the past because of a lack of tools to detect the different loci involved in resistance. Currently, despite the many studies focused on mapping resistance QTL, only a few of them are at a stage where cloning of the underlying gene is within reach.

A short summary on some advanced cloning projects for disease resistance QTL is given in Table 2. The table only lists examples of QTLs for which one or several labs are intensively working on high-resolution mapping and cloning. Typically, these large projects are made in some form of national or international collaborations. In wheat, a QTL for Fusarium head blight (FHB) resistance, *Qfhs.ndsu-3BS* from the resistant cv. Sumai 3 has been located in a very small region on chromosome 3B (Liu and Anderson 2003a,b) and the isolation of this gene can be expected in the near future. FHB (scab) is a very serious problem for wheat production. The disease is caused by several species of the genus *Fusarium,* and the International Maize and Wheat Improvement Center (CIMMYT) has described FHB to be a major limiting factor to wheat production in many parts of the world (Dubin et al. 1997). The fungus attacks the spike mostly at the flowering stage, and with favourable weather conditions, infection can result in serious yield losses. Crop quality is often compromised by harmful mycotoxins that are produced by the fungus in diseased grain. Fine mapping of *Qfhs.ndsu-3BS* gene is ongoing using the rice genome as a model to derive new markers for the wheat 3BS region. QTL cloning and

Table 2. Advanced projects for cloning disease resistance QTL and durable resistance genes in wheat

Target gene	Chromosomal location	Disease	QTL/single gene	Reference
Qfhs.ndsu-3BS	3BS	Fusarium Head blight	QTL	Liu and Anderson 2003b
Sr2	3BS	Stem rust	QTL/single gene	Kota et al. 2006
Lr34	7DS	Leaf rust	QTL	Schnurbusch et al. 2004b; Spielmeyer et al. 2005
Qsnb.sfr-3BS	3BS	Stagonospora nodorum	QTL	Schnurbusch et al. 2004a

characterization will shed light on the molecular mechanism underlying resistance to this major disease (Liu and Anderson, 2003a,b). A similar approach has been undertaken to map and isolate the horizontal slow rusting resistance loci *Lr34/Yr18* (Schnurbusch et al. 2004b, Spielmeyer et al. 2005), *Lr46/Yr29* (Rosewarne et al. 2006), and *Sr2* (Kota et al. 2006) as well as a major QTL for Stagonospora nodorum glume blotch resistance which is also located on chromosome 3BS (Schnurbusch et al. 2004a). The first two loci located on chromosomes 7DS and 1BL, respectively, are conferring quantitative and durable resistance to both leaf and stripe rust. *Lr34/Yr18* was also described to be linked to resistance against powdery mildew and *Barley Yellow Dwarf Virus*.

5. PROBLEMS FOR GENE IDENTIFICATION AND VALIDATION: GENOMIC RESOURCES ARE MOSTLY NOT FROM THE SPECIFIC CULTIVAR CARRYING THE RESISTANCE GENE OR QTL OF INTEREST

Once an *R* gene or a resistance QTL has been located in a physical interval, there is still the problem of the identification of the resistance gene from a number of candidate genes. As described above, there is only a very limited number of cultivars or lines for which genomic BAC libraries are available in cereals, mostly because of the genome size of these crops which make the development of such resources costly. Thus, there are hundreds of specific resistance genes identified in wheat and barley, as well as dozens of resistance QTL, all identified in different cultivars of the species. In contrast, there are only one or very few BAC libraries available for each cereal species, making it unlikely that there is a specific library of the line carrying the resistance gene of interest. The major question then is the following: does the genomic region of the cultivar from which the BAC library was made accurately corresponds to the region in the cultivar in which genetic mapping was performed? Or in other words, does the BAC cultivar contain an allelic form of the gene giving resistance? Are the two haplotypes similar enough to each other? This is not a trivial point as there are a number of cases where the haplotypes in susceptible cultivars simply do not contain the region of interest because of deletion events. This was found e.g. for the *Rpm1* Arabidopsis gene against *P. syringae* (Grant et al. 1998) and the *Lr10* leaf rust resistance gene in wheat (Feuillet et al., 2003). In the latter case, most modern hexaploid varieties carry the deletion haplotype, including the two cultivars from which BAC libraries are available: Renan and Chinese Spring. Thus, *Lr10* was only cloned by the fortunate situation that there was a BAC library available from diploid wheat *T. monococcum* cv. DV92 which turned out to have the same haplotype at the *Lr10* locus as the hexaploid wheat lines carrying the gene (Wicker et al. 2001, Feuillet et al. 2003, Isidore et al. 2005b). In other situations, particularly in the case of resistance gene clusters, the haplotypes at these loci differ dramatically between different cultivars. One of the best studied large and complex resistance loci, the *Dm3* locus is not from cereals but from lettuce. Nevertheless, it is probably a good example for the

enormous within-species or within-genus variability of resistance loci. The *Dm3* downy mildew resistance locus shows a great variability in the gene cluster of NBS-LRR genes, with cultivars having from 9 to 39 genes (Kuang et al. 2004). The actual line carrying the active *Dm3* gene has a cluster of 32 copies of the gene. A similar situation is found at the *Rp1* resistance locus which also shows a dramatic variability of the number of copies in the gene cluster (Collins et al. 1999; Smith et al. 2004). In such cases, in the absence of a BAC library from a cultivar with the gene, it is difficult to identify the actual resistance gene. Several strategies have been developed to solve this problem: cosmid and cDNA libraries from the lines carrying functional *Mla* resistance alleles were used for the isolation these genes (Halterman et al. 2001; Zhou et al. 2001). In a second strategy, a non-gridded BAC library of the cultivar of interest can be made. This was successfully done in the case of the *Rph7* leaf rust resistance gene in barley present in the cultivar Cepada Capa, although the nature of the resistance gene is still not clear (Isidore et al. 2005b, Scherrer et al. 2005). Non-gridded libraries can be screened using pooling strategies. Although this approach is not quite as efficient as the use of a fully arrayed or pooled BAC library, such libraries are very valuable and cost effective if only a single gene needs to be isolated from a specific line.

It is not yet clear if resistance gene loci are particularly variable among cultivars of a specific species or if variability, and thus the problems of cultivar specific haplotypes described above, is a general problem of gene and QTL isolation in crop species. Given recent evidence in maize (Fu and Dooner 2002, Brunner et al. 2005, Morgante et al. 2005) and barley (Scherrer et al. 2005), there is considerable variability at many genetic loci among cultivars of the same species. This might be mostly due to gene fragments moved by helitron transposons (Morgante et al. 2005) and therefore have little phenotypic consequences. However, the detailed aspects of intra-species variability has to be determined in much more detail to learn if haplotype variability will represent a common problem for cereal gene and QTL isolation. It might also be worth to consider the construction of BAC libraries for several wild accessions from a particular species. This might result in a good representation of the haplotypes originally present in the gene pool and provide the necessary basis for molecular work on most haplotypes present in modern cultivars.

6. VALIDATION OF CANDIDATE RESISTANCE GENES

Map-based cloning can delineate a resistance gene or resistance QTL to a small region of the genome. However, in most cases it will not result in only one candidate gene but in a number of genes which might be the gene of interest. Validation remains a thorny and often underestimated part of map-based cloning projects which today, with more and better tools available than just a few years ago, can result in relatively rapid progress towards the identification of the candidate genomic region.

In the case of race-specific, major resistance genes which are frequently encoded by NBS-LRR type of genes, the identification of a single NBS-LRR gene in the genetic interval carrying the resistance gene makes it likely that this is actually

the gene of interest. Nevertheless, such a simple situation is rather the exception than the rule as NBS-LRR genes frequently occur in gene clusters, making the identification of the actual *R* gene problematic (see e.g. Yahiaoui et al. 2004) and in many cases, more sophisticated approaches are needed. Once one or more candidate genes have been identified for a specific resistance gene or QTL, there are a number of possible ways to confirm or exclude its identity as a resistance factor. Ideally, functional null mutants are available. Sequence analysis of candidate genes in these mutants can rapidly reveal whether the candidate gene is the resistance gene. The barley *mlo* powdery mildew resistance gene was quickly confirmed this way, as well as the barley *Mla1* and *Mla12* alleles (Shen et al. 2003; Zhou et al. 2001), the *Lr10* leaf rust resistance gene in wheat (Feuillet et al. 2003) and the wheat powdery mildew resistance gene *Pm3b* (Yahiaoui et al. 2004). Knock out mutants are much more easily identified for major resistance genes than for resistance QTL. Nevertheless, for major resistance QTL such as the leaf rust resistance locus *Lr34*, mutants have been identified and will be highly useful once candidate genes are identified (R. Singh, CIMMYT, personal communication).

Powdery mildew is a highly relevant pathogen in several cereal crops. Its interaction with wheat or barley occurs exclusively in epidermal cells which can be transiently transformed by particle bombardment (Douchkov et al. 2005; Schweizer et al. 1999; Shirasu et al. 1999). Using marker genes such as genes encoding beta-glucuronidase or green fluorescent protein, the transformed cells can be identified. By cobombardment with a candidate resistance gene and subsequent infection with a specific race of the powdery mildew pathogen, transformed cells can be examined for compatible or incompatible interaction with the pathogen. This allows for a test of whether a candidate gene confers resistance and has been successfully used for the *mlo* gene (Shirasu et al.1999) and the *Mla* alleles (Halterman et al. 2001; Shen et al. 2003; Zhou et al. 2001) in barley as well as for the *Pm3* allelic series in wheat (Srichumpa et al. 2005; Yahiaoui et al. 2006; Yahiaoui et al. 2004). Such transient assays are of course extremely valuable but are currently restricted to young seedlings and the epidermal tissue. Many pathogens, including the rusts, interact mainly with mesophyll cells and transient expression is not feasible for functional analysis of candidate resistance genes against these pathogens.

Stable transformation in cereals such as wheat and barley as well as maize is now routine but still tedious and time consuming before results can be obtained in the T_1 or T_2 generation. In barley, stable Agrobacterium-mediated transformation is mostly restricted to the cultivar Golden Promise, whereas in wheat, particle bombardment is still the most widely used method. In wheat, several sublines of cultivar Bobhwite, the Canadian cultivar Fielder as well as a number of other cultivars are transformed in different projects. The restricted availability of lines which can be transformed efficiently is a potential problem if a specific gene should be tested. For example, both cultivar Fielder as well as cultivar Bobwhite S26, which can be easily transformed, already contain several leaf rust resistance genes, among them the *Lr10* gene (Feuillet et al. 2003). Thus, there is still a need to have a broader set of genotypes that can be transformed. This will increase the chance

of finding a line which is highly susceptible against the specific disease of interest, allowing reliable phenotyping of transgenic plants containing candidate genes.

Recently, virus-induced gene silencing (VIGS) based on the *Barley Stripe Mosaic Virus* (BSMV) has been developed as a new tool for functional gene analysis in barley and wheat (Holzberg et al. 2002, Scofield et al. 2005). In this system, short fragments (up to 1200 nucleotides) are cloned into the γ-genome of the tripartite viral genome. After infection of the plants at the young seedling stage (barley or wheat), the target gene is effectively silenced. In the case of a candidate gene actually being the resistance gene, after infection of seedling genotypes carrying the resistance gene of interest with the relevant pathogen race, we expect silencing of the resistance gene and susceptibility. The power of VIGS in cereals was recently shown for the wheat *Lr21* leaf rust resistance gene which was efficiently silenced by VIGS and seedlings became susceptible (Scofield et al. 2005). This confirmed the results previously obtained by stable transformation (Huang et al. 2003). Thus, candidate genes whose function can be determined at the seedling stage can be efficiently silenced and resistance genes can be identified. There are two possible problems with this system which have to be considered in its application: first, in the barley gene pool there is widespread resistance against BSMV, preventing its use in these lines. Therefore, for each barley line to be used for VIGS, it has to be established first that it is susceptible to the virus. In wheat cultivars, resistance to BSMV is rare or absent, making VIGS more straightforward (S. Bieri and B. Keller, unpublished data). The second problem is related to the fact that *R* genes are frequently organized as gene clusters of closely related genes at a specific locus. Several tandemly oriented and highly similar copies of NBS-LRR genes represent a considerable problem for VIGS identification of the correct gene. As VIGS is silencing all genes which are >80% homologous (Holzberg et al. 2002), all members of an *R* gene family at a specific locus will be silenced and VIGS will only be useful for confirming that the gene of interest is a member of the gene family, but will not help to identify the particular family member which is the resistance gene. Possibly, the use of more specific 3' untranslated regions of individual gene members of the family can solve this problem.

7. OUTLOOK

The improvement of genomic resources for application in cereals promises to make map-based cloning of resistance genes and resistance QTL an increasingly important topic of research. After many decades of classical resistance breeding there is an enormous pool of lines carrying well defined resistance traits against most of the important diseases in cereals. This is probably one of the broadest and best characterized genetic resources in higher plants, but it still waits to be explored for a better molecular understanding of plant disease resistance. Given the many ongoing map-based cloning projects for resistance genes and QTLs, we will see a rapid progress in the isolation of genetic factors contributing to different types of resistances in the next years. Map-based cloning and the identification

of candidate genes will possibly also profit from association mapping which has a high potential for genetic resolution in suitable populations. It remains to be determined how easy it will be to assemble good, diverse populations derived from the wheat breeding pool for association mapping. However, the first successful association mapping study in wheat has recently been published (Breseghello and Sorrells 2006). Cloning of resistance genes in cereals will not only contribute to better molecular understanding of disease resistance but also, as the work is done in crop plants, to improved resistance breeding. Two main approaches will support breeding: first, it will become easier to combine resistance factors by marker-assisted selection and second, transgenic approaches which are based on a better molecular understanding of the fundamental processes underlying disease resistance will become feasible.

REFERENCES

Akhunov E, Akhunova A, Dvorak J (2005) BAC libraries of *Triticum urartu*, *Aegilops speltoides* and *Ae. tauschii*, the diploid ancestors of polyploid wheat. Theor Appl Genet 111:1617–1622

Allouis S, Moore G, Bellec A, Sharp R, Faivre Rampant P, Mortimer L, Pateyron S, Foote TN, Griffiths S, Caboche M, Chalhoub B (2003) Construction and characterisation of a hexaploid wheat (*Triticum aestivum* L.) BAC library from the reference germplasm 'Chinese Spring'. Cereal Res Commun 31:331–338

Ayliffe MA, Lagudah ES (2004) Molecular genetics of disease resistance in cereals. Ann Bot 94:765–773

Bai JF, Pennill LA, Ning JC, Lee SW, Ramalingam J, Webb CA, Zhao BY, Sun Q, Nelson JC, Leach JE, Hulbert SH (2002) Diversity in nucleotide binding site-leucine-rich repeat genes in cereals. Genome Res 12:1871–1884

Baumberger N, Ringli C, Keller B (2001) The chimeric leucine-rich repeat/extensin cell wall protein LRX1 is required for root hair morphogenesis in *Arabidopsis thaliana*. Genes Dev 15:1128–1139

Bennetzen JL, Ma J (2003) The genetic colinearity of rice and other cereals on the basis of genomic sequence analysis. Curr Opin Plant Biol 6:128–133

Bieri S, Mauch S, Shen Q-H, Peart J, Devoto A, Casais C, Ceron F, Schulze S, Steinbiss H-H, Shirasu K, Schulze-Lefert P (2004) RAR1 positively controls steady state levels of barley MLA resistance proteins and enables sufficient MLA6 accumulation for effective resistance. Plant Cell 16:3480–3495

Breseghello F, Sorrells ME (2006) Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. Genetics 172:1165–1177

Brueggeman R, Rostoks N, Kudrna D, Kilian A, Han F, Chen J, Druka A, Steffenson B, Kleinhofs A (2002) The barley stem rust-resistance gene *Rpg1* is a novel disease-resistance gene with homology to receptor kinases. Proc Natl Acad Sci USA 99:9328–9333

Brunner S, Fengler K, Morgante M, Tingey S, Rafalski A (2005) Evolution of DNA sequence non-homologies among maize inbreds. Plant Cell 17:343–360

Bryan GT, Wu KS, Farrall L, Jia Y, Hershey P et al (2000) A single amino acid difference distinguishes resistant and susceptible alleles of the rice blast resistance gene *Pi-ta*. Plant Cell 12:2033–2046

Bulgarelli D, Collins NC, Tacconi G, Dellaglio E, Brueggeman R, Kleinhofs A, Stanca AM, Vale G (2004) High genetic-resolution mapping of the leaf stripe resistance gene *Rdg2a* in barley. Theor Appl Genet 108:1401–1408

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 (1997) The barley *Mlo* gene: a novel control element of plant pathogen resistance. Cell 88:695–705

Cenci A, Chantret N, Kong X, Gu Y, Anderson OD, Fahima T, Distelfeld A, Dubcovsky J (2003) Construction and characterization of a half million clone BAC library of durum wheat (*Triticum turgidum* ssp. *durum*). Theor Appl Genet 107:931–939

Collins N, Drake J, Ayliffe M, Sun Q, Ellis J, Hulbert S, Pryor T (1999) Molecular characterization of the maize *Rp1-D* rust resistance haplotype and its mutants. Plant Cell 11:1365–1376

De Lorenzo G, D'Ovidio R, Cervone F (2001) The role of polygalacturonase-inhibiting proteins (PGIPS) in defense against pathogenic fungi. Ann Rev Phytopathol 39:313–335

Douchkov D, Nowara D, Zierold U, Schweizer, P (2005) A high-throughput gene-silencing system for the functional assessment of defense-related genes in barley epidermal cells. Mol Plant Microbe Interact 18:755–761

Dubin HJ, Gilchrist L, Reeves J, McNab A, (eds) (1997) Fusarium head scab: global status and prospects. CIMMYT, Mexico, DF, Mexico. 130 p

Feuillet C, Travella S, Stein N, Albar L, Nublat A, Keller B (2003) Map-based isolation of the leaf rust disease resistance gene *Lr10* from the hexaploid wheat (*Triticum aestivum* L.) genome. Proc Natl Acad Sci USA 100:15253–15258

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

Grant MR, McDowell JM, Sharpe AG, de Torres Zabala M, Lydiate DJ, Dangl JL (1998) Independent deletions of a pathogen-resistance gene in *Brassica* and *Arabidopsis.* Proc Natl Acad Sci USA 95:15843–15848

Guyot R, Yahiaoui N, Feuillet C, Keller B (2004) *In silico* comparative analysis reveals a mosaic conservation of genes within a novel colinear region in wheat chromosome 1AS and rice chromosome 5S. Funct Integr Genomics 4:47–58

Halterman DA, Wise RP (2004) 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 J 38:215–226

Halterman D, Zhou FS, Wei FS, Wise RP, Schulze-Lefert P (2001) The MLA6 coiled-coil, NBS-LRR protein confers *AvrMla6*-dependent resistance specificity to *Blumeria graminis* f. sp. *hordei* in barley and wheat. Plant J 25:335–348

Halterman DA, Wei FS, Wise RP (2003) Powdery mildew-induced *Mla* mRNAs are alternatively spliced and contain multiple upstream open reading frames. Plant Physiol 131:558–567

Holzberg S, Brosio P, Gross C, Pogue GP (2002) Barley stripe mosaic virus-induced gene silencing in a monocot plant. Plant J 30:315–327

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

Hu GS, Richter TE, Hulbert SH, Pryor T (1996) Disease lesion mimicry caused by mutations in the rust resistance gene *rp1*. Plant Cell 8:1367–1376

Huang L, Brooks SA, Li W, Fellers JP, Trick HN, Gill BS (2003) Map-based cloning of leaf rust resistance gene *Lr21* from the large and polyploid genome of bread wheat. Genetics 164:655–664

Isidore E, Scherrer B, Bellec A, Budin K, Faivre-Rampant P, Waugh R, Keller B, Caboche M, Feuillet C, Chalhoub B (2005a) Direct targeting and isolation of genes of interest using an improved pooled BAC libraries cloning and screening strategy. Funct Integr Genomics 5:97–103

Isidore E, Scherrer B, Chalhoub B, Feuillet C, Keller B (2005b) Ancient haplotypes resulting from extensive molecular rearrangements in the wheat A genome have been maintained in species of three different ploidy levels. Genome Res 15:526–536

Iyer AS, McCouch SR (2004) The rice bacterial blight resistance gene *xa5* encodes a novel form of disease resistance. Mol Plant Microbe Interact 17:1348–1354

Janda J, Bartos J, Safar J, Kubalakova M, Valarik M, Cihalikova J, Simkova H, Caboche M, Sourdille P, Bernard M, Chalhoub B, Dolezel J (2004) Construction of a subgenomic BAC library specific for chromosomes 1D, 4D and 6D of hexaploid wheat. Theor Appl Genet 109:1337–1345

Jiang GH, Xia ZH, Zhou YL, Wan J, Li DY, Chen RS, Zhai WX, Zhu LH (2006) Testifying the rice bacterial blight resistance gene *xa5* by genetic complementation and further analyzing *xa5 (Xa5)* in comparison with its homology TFIIA gamma 1. Mol Gen Genet 275:354–366

Jorgensen JH (1993) Coordinator's reports: disease and pest resistance genes. Barley Genet Newsl 22:10–133

Keller B, Feuillet C, Yahiaoui N (2005) Map-based isolation of disease resistance genes from bread wheat: cloning in a supersize genome. Genet Res Camb 85:93–100

Kota R, Spielmeyer W, McIntosh RA, Lagudah ES (2006) Fine genetic mapping fails to dissociate durable stem rust resistance gene *Sr2* from pseudo-black chaff in common wheat (*Triticum aestivum* L.). Theor Appl Genet 112:492–499

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

Leister D, Kurth J, Laurie DA, Yano M, Sasaki T, Devos K, Graner A, Schulze-Lefert P (1998) Rapid reorganization of resistance gene homologues in cereal genomes. Proc Natl Acad Sci USA 95:370–375

Lijavetzky D, Muzzi G, Wicker T, Keller B, Wing R, Dubcovsky J (1999) Construction and characterization of a bacterial artificial chromosome (BAC) library for the A genome of wheat. Genome 42:1176–1182

Liu S, Anderson JA (2003a) Marker assisted evaluation of *Fusarium* head blight resistant wheat germplasm. Crop Sci 43:760–766

Liu S, Anderson JA (2003b) Targeted molecular mapping of a major wheat QTL for Fusarium head blight resistance using wheat ESTs and synteny with rice. Genome 46:817–823

Lu H-J, Faris JD (2006) Macro- and microcolinearity between the genomic region of wheat chromosome 5B containing the *Tsn1* gene and the rice genome. Funct Integr Genomics 6:90–103

Lu H-J, Fellers JP, Friesen TL, Meinhardt SW, Faris JD (2006) Genomic analysis and marker development for the *Tsn1* locus in wheat using bin-mapped ESTs and flanking BAC contigs. Theor Appl Genet 112:1132–1142

Luo MC, Thomas C, You FM, Hsiao J, Ouyang S, Buell CR, Malandro M, McGuire PE, Anderson OD, Dvorak J (2003) High-throughput fingerprinting of bacterial artificial chromosomes using the snapshot labeling kit and sizing of restriction fragments by capillary electrophoresis. Genomics 82:378–389

Mago R, Miah H, Lawrence GJ, Wellings CR, Spielmeyer W, Bariana HS, McIntosh RA, Pryor AJ, Ellis JG (2005) High-resolution mapping and mutation analysis separate the rust resistance genes *Sr31*, *Lr26* and *Yr9* on the short arm of rye chromosome 1. Theor Appl Genet 112:41–50

Mammadov JA, Steffenson BJ, Saghai Maroof MA (2005) High-resolution mapping of the barley leaf rust resistance gene *Rph5* using barley expressed sequence tags (ESTs) and synteny with rice. Theor Appl Genet 111:1651–1660

McDonald BA, Linde C (2002) Pathogen population genetics, evolutionary potential, and durable resistance. Ann Rev Phytopathol 40:349–379

McDowell JM, Dhandaydham M, Long TA, Aarts MGM, Goff S, Holub EB, Dangl JL (1998) Intragenic recombination and diversifying selection contribute to the evolution of downy mildew resistance at the RPP8 locus of *Arabidopsis*. Plant Cell 10:1861–1874

McIntosh RA, Wellings CR, Parks RF (1995). (In: Cloud-Guest A, Jeans K) (eds) Wheat rusts: an atlas of resistance genes. Kluwer, Dordrecht, pp 1–200

McIntosh RA, Yamzaki Y, Devos KM, Dubcovsky J, Rogers WJ (2003) Catalogue of gene symbols for wheat. In: Pognat NE, Romano M, Pogna E, Galterio G (eds) Proceedings of the 10th International Wheat Genetics Sysmposium, Istituto Sperimentale per la Cerealicoltura Paestum, Italy pp 1–34

Meyers BC, Shen KA, Rohani P, Gaut BS, Michelmore RW (1998) Receptor-like genes in the major resistance locus of lettuce are subject to divergent selection. Plant Cell 10:1833–1846

Monosi B, Wisser RJ, Pennill L, Hulbert SH (2004) Full-genome analysis of resistance gene homologues in rice. Theor Appl Genet 109:1434–1447

Morgante M, Brunner S, Pea G, Fengler K, Zuccolo A, Rafalski A (2005) Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize. Nat Genet 37:997–1002

Moullet O, Zhang HB, Lagudah ES (1999) Construction and characterisation of a large DNA insert library from the D genome of wheat. Theor Appl Genet 99:305–313

Nilmalgoda SD, Cloutier S, Walichnowski AZ (2003) Construction and characterization of a bacterial articial chromosome (BAC) library of hexaploid wheat (*Triticum aestivum* L.) and validation of genome coverage using locus-specific primers. Genome 46:870–878

Nürnberger T, Brunner F, Kemmerling B, Piater L (2004) Innate immunity in plants and animals: striking similarities and obvious differences. Immunol Rev 198:249–266

Oerke E-C, Dehne H-W (1997) Global crop production and the efficacy of crop protection – current situation and future trends. Europ J Plant Pathol 103:203–215

Piffanelli P, Ramsay L, Waugh R, Benabdelmouna A, D'Hont A, Hollricher K, Jorgensen JH, Schulze-Lefert P, Panstruga R (2004) A barley cultivation-associated polymorphism conveys resistance to powdery mildew. Nature 430:887–891

Qu SH, Liu GF, Zhou B, Bellizzi M, Zeng LR, Dai LY, Han B, Wang GL (2006) The broad-spectrum blast resistance gene *Pi9* encodes a nucleotide-binding site-leucine-rich repeat protein and is a member of a multigene family in rice. Genetics 172:1901–1914

Richter TE, Pryor TJ, Bennetzen JL, Hulbert SH (1995) New rust resistance specificities associated with recombination in the *Rp1* complex in maize. Genetics 141:373–381

Rosewarne GM, Singh RP, Huerta-Espino J, William HM, Bouchet S, Cloutier S, McFadden H, Lagudah ES (2006) Leaf tip necrosis, molecular markers and β1-proteasome subunits associated with the slow rusting resistance genes *Lr46/Yr29*. Theor Appl Genet 112:500–508

Safar J, Bartos J, Janda J, Bellec A, Kubalakova M, Valarik M, Pateyron S, Weiserova J, Tuskova R, Cihalikova J, Vrana J, Simkova H, Faivre-Rampant P, Sourdille P, Caboche M, Bernard M, Dolezel J, Chalhoub B (2004) Dissecting large and complex genomes: flow sorting and BAC cloning of individual chromosomes from bread wheat. Plant J 39:960–968

Salvi S, Tuberosa R (2005) To clone or not clone plant QTLs: present and future challenges. Trends Plant Sci 10:297–304

Saxena KMS, Hooker AL (1968) On structure of a gene for disease resistance in maize. Proc Natl Acad Sci USA 61:1300–1305

Scherrer B, Isidore E, Klein P, Kim J, Bellec A, Chalhoub B, Keller B, Feuillet C (2005) Large-intra-specific haplotype variability at the *Rph7* locus results from rapid and recent divergence in the barley genome. Plant Cell 17:361–374

Schnurbusch T, Paillard S, Schori A, Messmer M, Schachermayr G, Winzeler M, Keller B (2004a) Dissection of quantitative and durable leaf rust resistance in Swiss winter wheat reveals a major resistance QTL in the *Lr34* chromosomal region. Theor Appl Genet 108:477–484

Schnurbusch T, Bossolini E, Messmer M, Keller B (2004b) Tagging and validation of a major quantitative trait locus for leaf rust resistance and leaf tip necrosis in winter wheat cultivar Forno. Phytopathology 94:1036–1041

Schulze-Lefert P, Bieri S (2005) Recognition at a distance. Science 308:506–508

Schweizer P, Pokorny J, Abderhalden O, Dudler R (1999) A transient assay system for the functional assessment of defense-related genes in wheat. Mol Plant Microbe Interact 12:647–654

Scofield SR, Huang L, Brandt AS, Gill BS (2005) Development of a virus-induced gene-silencing system for hexaploid wheat and its use in functional analysis of the *Lr21*-mediated leaf rust resistance pathway. Plant Physiol 138:2165–2173

Shen QH, Zhou F, Bieri S, Haizel T, Shirasu K, Schulze-Lefert P (2003) Recognition specificity and RAR1/SGT1 dependence in barley *Mla* disease resistance genes to the powdery mildew fungus. Plant Cell 15:732–744

Shen B, Wang DM, McIntyre CL, Liu CJ (2005) A 'Chinese Spring' wheat (*Triticum aestivum* L.) bacterial artificial chromosome library and its use in the isolation of SSR markers for targeted genome regions. Theor Appl Genet 111:1489–1494

Shirasu K, Nielsen K, Piffanelli P, Oliver R, Schulze-Lefert P (1999) Cell-autonomous complementation of *mlo* resistance using a biolistic transient expression system. Plant J 17:293–299

Smith SM, Hulbert SH (2005) Recombination events generating a novel *Rp1* race specificity. Mol Plant Microbe Interact 18:220–228

Smith SM, Pryor AJ, Hulbert SH (2004) Allelic and haplotypic diversity at the *Rp1* rust resistance locus of maize. Genetics 167:1939–1947

Song WY, Wang GL, Chen LL, Kim HS, Pi LY, Holsten T, Gardner J, Wang B, Zhai WX, Zhu LH, Fauquet C, Ronald P (1995) A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa21*. Science 270:1804–1806

Spielmeyer W, McIntosh RA, Kolmer J, Lagudah ES (2005) Powdery mildew resistance and *Lr34/Yr18* genes for durable resistance to leaf and stripe rust cosegregate at a locus on the short arm of chromosome 7D of wheat. Theor Appl Genet 111:731–735

Srichumpa P, Brunner S, Keller B, Yahiaoui N (2005) Allelic series of four powdery mildew resistance genes at the *Pm3* locus in hexaploid bread wheat. Plant Physiol 139:885–895

Stein N, Feuillet C, Wicker T, Schlagenhauf E, Keller B (2000) 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 97:13436–13441

Stein N, Perovic D, Kumlehn J, Pelliio B, Stracke S, Streng S, Ordon F, Graner A (2005) The eukaryotic translation initiation factor 4E confer multiallelic recessive *Bymovirus* resistance in *Hordeum vulgare* (L.). Plant J 42:912–922

Sun XL, Cao YL, Yang ZF, Xu CG, Li XH, Wang SP, Zang QF (2004) *Xa26*, a gene conferring resistance to *Xanthomonas oryzae* pv. *oryzae* in rice, encodes an LRR receptor kinase-like protein. Plant J 37:517–527

Toumasini L, Yahiaoui N, Srichumpa P, Keller B (2006) Development of functional moleculer markers specific for seven *Pm 3* resistance aller and their validation in the bread wheat gene pool. Theor Appl Genet 114: 165–175

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

Wei F, Gobelman-Werner K, Morroll SM, Kurth J, Mao L, Wing R, Leister D, Schulze-Lefert P, Wise RP (1999) 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 153:1929–1948

Wei F, Wing RA, Wise RP (2002) Genome dynamics and evolution of the *Mla* (powdery mildew) resistance locus in barley. Plant Cell 14:1903–1917

Weibull J, Walther U, Sato K, Habekuss A, Kopahnke D, Proeseler G (2003) Diversity in resistance to biotic stresses. In: von Bothmer R (ed) Diversity in barley (*Hordeum vulgare*), (elsevier Science BV), Amsterdam, pp 143–178

Wicker T, Stein N, Albar L, Feuillet C, Schlagenhauf E, Keller B (2001) Analysis of a contiguous 211 kb sequence in diploid wheat (*Triticum monococcum* L.) reveals multiple mechanisms of genome evolution. Plant J 26:307–316

Wicker T, Zimmermann W, Perovic D, Paterson AH, Ganal M, Graner A, Stein N (2005) A detailed look at 7 million years of genome evolution in a 439 kb contiguous sequence at the barley *Hv-eIF4E* locus: recombination, rearrangements and repeats. Plant J 41:184–194

Xu Y, McCouch SR, Zhang Q (2005) How can we use genomics to improve cereals with rice as a reference genome? Plant Mol Biol 59:7–26

Yahiaoui N, Srichumpa P, Dudler R, Keller B (2004) Genome analysis at different ploidy levels allows cloning of the powdery mildew resistance gene *Pm3b* from hexaploid wheat. Plant J 37:528–538

Yahiaoui N, Brunner S, Keller B (2006) Rapid generation of new powdery mildew resistance genes after wheat domestication. Plant J 47:85–98

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

Yu Y, Tomkins JP, Waugh R, Frisch DA, Kudrna D, Kleinhofs A, Brueggeman RS, Muehlbauer GJ, Wise RP, Wing RA (2000) A bacterial artificial chromosome library for barley (*Hordeum vulgare* L.) and the identification of clones containing putative resistance genes. Theor Appl Genet 101:1093–1099

Zeller FJ, Hsam SLK (1998) Progress in breeding for resistance to powdery mildew in common wheat (T*riticum aestivum* L.). Proceedings of the IX international wheat genetics symposium Saskatoon, Saskatchewan, pp 178–180

Zhou FS, Kurth JC, Wei FS, Elliott C, Vale G, Yahiaoui N, Keller B, Somerville S, Wise R, Schulze-Lefert P (2001) Cell-autonomous expression of barley *Mla1* confers race-specific resistance to the powdery mildew fungus via a *Rar1*-independent signaling pathway. Plant Cell 13:337–350

CHAPTER 6

MAIZE BREEDING AND GENOMICS: AN HISTORICAL OVERVIEW AND PERSPECTIVES

MICHAEL LEE*
Iowa State University, Ames, Iowa USA

Abstract: The 1909 publication of G. H. Shull was the dawn of the modern era of maize breeding. Since that time maize has been unique as a major crop of the world and a very important species for basic research. Most of the knowledge from that research has been either worthless or inaccessible to maize breeding programs and actual improvement of maize. Recently, this relationship has changed and the connections between basic research and breeding have improved. The previous 10-15 years have been especially eventful with the advent of transgenic maize and DNA sequence information for maize germplasm. There is much to learn about the maize genome. Only a small fraction of the 30,000 to 60,000 genes and other sequences have functions assigned to them on the basis of direct experimentation. The advent of the genomics era of maize breeding, even at such an early stage, has clearly underlined our abilities to empirically assess all of the promising genetic options and questions that are raised by new sources of information. Field-based phenotyping and the prioritization of phenotyping have become much more important in the genomics era of maize breeding**.** The gains and pains of the initial decades of maize breeding have been well documented for some traits and production environments. But, with the exception of improved resistance to biotic and abiotic stress as a basis for genetic gains in grain yield, very little fundamental information and few validated mechanisms have been identified from that era. The next era of maize breeding should be fundamentally different in the sense that the strategies will rely more on a scientific method; consequently, some of the reasons for success or failure will be known and considered for devising more efficient methods of maize improvement.

1. INTRODUCTION

This review presents a summary and critical assessment of the achievements, prospects and limitations of genomics research for maize improvement, i.e., genomics-assisted breeding (GAB) and genomics-assisted selection (GAS). This has

*Address for correspondence: mlee@iastate.edu

R.K. Varshney and R. Tuberosa (eds.), Genomics Assisted Crop Improvement: Vol. 2: Genomics Applications in Crops, 129–146.

become a reality for maize improvement in some areas of the world and it has introduced new scientific and social considerations. Presently, science-based breeding of maize still occurs within the context of a mostly unknown genome and dynamic environments and furthermore, it is expensive at some stages of implementation and maintenance.

The origin of the modern era of maize breeding began roughly 100 years ago with Shull's experiments (Shull 1909). For much of the past century, maize has played a unique role among plant species because of its pivotal role as a major crop worldwide and a very important species for basic research. In particular, the previous 15 years have been especially eventful with the advent of transgenic maize and "low cost" DNA sequence information for maize germplasm. This is only the beginning as we are just starting to understand the content and organization of the maize genome, the interconnecting circuitries among genotype, environment and phenotype, and how to translate such intellectual wealth into beneficial forms.

Breeding and genomics are rather broad and encompassing topics. Therefore, the only certainty of this review is that it will miss or omit some important information or perspectives. Under this respect, some definitions may be helpful in order to reveal my perspective and bias. Plant breeding has always been the genetic adaptation of plants to the desires and needs of human societies in the context of agricultural systems. Genomics, with all of the subcategories, is the attempt to conduct comprehensive assessments and functional annotation of an organism's genome and gene products. Together, plant breeding and genomics form a complementary approach to aspects of systems biology in agriculture as the former has relied heavily on phenotypic analyses while the latter has emphasized genotypic and molecular analyses. Both approaches are necessary and increasingly synergistic.

The foundation of GAB for maize was prepared through a series of social, technical and scientific events spanning several decades. Detailed accounts of some of those events have been presented in recent publications (Charles 2001; Troyer 2004; Crosby et al. 2006). In the U.S., agricultural subsidy programs, the Plant Variety Protection Act of 1970 and the *Chakrabarty* v. *Diamond* ruling of 1980 are examples of important social events that either encouraged or enabled investments in plant breeding and genomics by corporations and venture capitalists. For maize, those investments and the security provided by the rule of law were vital to the development of GAB. In science and technology, examples of significant developments or discoveries include methods for plant regeneration from cell cultures of maize in the '70s, cloning and sequencing the first maize genes in the '80s, the introduction of the first molecular markers in the '80s and early '90s, and the advancements in computational biology and field plot equipment. Enabled by those events and developments, the first transgenic maize hybrids were first sold in the U.S. in the mid '90s (Troyer 2004). Consequently, some aspects of maize breeding have changed significantly. With respect to maize, those events are presented in their approximate chronological order in this chapter.

2. GENOMICS-ASSISTED BREEDING FOR MAIZE: TRANSFORMATION

In this section, aspects of some of the more influential developments in biotechnology and bioinformatics will be summarized and evaluated with respect to aspects of maize breeding and improvement. Generic schemes and considerations for genomic-assisted breeding have been presented (Varshney et al. 2005). When possible, reviews of methods and technology have been provided.

New and superior germplasm has always been a disruptive and driving force in a competitive seed market. For example, when inbred line B73 was developed, the research directors of the two leading seed companies in the U.S. informed Wilbert A. Russell, who with Lowell Penny developed that inbred, that their companies could not use B73 in their programs (W.A. Russell, personnel communication). However, soon thereafter, other companies used B73 to produce superior hybrids and all other companies were forced to do the same. Decades later, the value and impact of a few transgenes were similarly underestimated (Charles 2001).

Methods for genetic transformation and a few transgenes for resistance to insects and herbicides have been major assets and disruptive forces for maize breeding programs for various reasons which will be discussed in the section on maize breeding (Lee 1999; Crosbie et al. 2006; Table 1). Here, the methods and their consequences will be discussed. Technical reviews of maize transformation methods for the nuclear genome have been published (Armstrong 1999; Kaeppler 2004; Torney et al. 2006). Transformation of the mitochondrial and chloroplast genomes, while desirable for some circumstances is less advanced and will not be considered herein (Maliga 2004).

In the public sector, Agrobacterium-based transformation of immature zygotic embryos has become a method of choice; although biolistic methods of transformation are still used (Torney et al. 2006). An important advantage of Agrobacterium-based is a higher frequency of apparently unrearranged, single-copy insertions of the transgene(s) into the maize nuclear genome. That advantage is important for GAB because it facilitates assessments and predictability of the relationship between genotype and phenotype (e.g. fewer problems with silencing) with respect to the transgene and it should simplify evaluations required by regulatory agencies.

One unique problem with that method is unintentional transfer and integration of portions of the bacterial plasmid into the maize genome: up to 70% of the events may result in stable integration of the intended transgene and portions of the plasmid (Shou et al. 2004). The effects of such unwanted DNA sequences on gene expression and phenotype were not determined but they would be a consideration in subsequent research, development and regulatory review.

Although Agrobacterium-mediated transformation of maize represents a significant advance, several serious limitations remain with respect to the target tissue, length of the DNA sequence delivered, the sites of integration and genotypic specificity (Lee 1999; Kaeppler 2004; Cheng et al. 2004; Torney et al. 2006). The tissue culture-based methods of transformation are labor intensive, may result in

Table 1. Assessment of genomics platforms and related infrastructure for their utility in maize breeding[1]

Infrastructure	Not Important	Important	Very Important	Comments[2]
DNA sequence analysis	0	0	15	
Transcript analysis	12	3	0	Need more information on actual gene function.
Proteomics	12	3	0	
Analytical services[3]	0	1	14	
TILLING	13	2	0	Maize has a naturally high level of DNA sequence polymorphism.
Insertional Mutagenesis	9	3	3	Concerns about intellectual property and patent claims.
Transformation	0	1	14	Not always able to transform the desired genotype but vital for validation and product development.
Genetic Map	4	3	7	
Complete Genomic sequence	2	5	8	
Physical Map	8	4	3	
Informatics	0	0	15	
Comparative Genomics	3	6	6	
Modeling and Simulation of phenotype	7	4	4	Need more information on actual gene functions and interactions between internal and external factors for effective modeling.
Off-season breeding facilities	0	0	15	
Wide-area field testing network	0	0	15	
Twin-plot field combines	0	0	15	

[1] Survey conducted May through October 2006 of 15 research managers from six multinational maize breeding programs.

[2] Text in the "comments" column indicates that more than half of the responses included a similar statement; thus, there is some evidence of consensus for that comment.

[3] Analytical services are high-throughput analyses of chemical composition of plant parts.

somaclonal variation and require considerable time to proceed from transformed cell to seed; essentially, the methods are very similar to those established in the 1970s. Pollen-based methods would greatly improve this aspect of GAB but they have not been developed despite numerous efforts in this direction.

There has been, however, more progress in other components of transformation technology. The length of the delivered DNA sequence, while routinely in the range of a few kilobase pairs, has the potential to be extended to several tens of kilobases (Hamilton et al. 1996). The ability to introduce longer sequences of DNA

would facilitate GAB by accelerating the introduction and assessment of novel DNA sequences, combinations of genes and regulatory sequences and the establishment of transgene linkage blocks or artificial chromosomes (Lee 1995).

Most analyses of transgene integration in maize indicate that the desired sequences insert into and are transmitted by the nuclear genome (as opposed to the mitochondrial or chloroplast genomes). However, the ability to direct the transgenes to specific sites, or to eliminate certain sequences via site-specific recombination has not become routine for maize; but, incipient and potentially transferable technology has been reported for some plant species. Working models of site-specific recombination based on Cre-lox were demonstrated several years ago (Ow 2002). Recently, zinc-finger nucleases have been used to achieve a higher frequency of site-specific recombination in tobacco but tests in maize have not been reported (Wright et al. 2005). Given that so much of the maize nuclear genome consists of frequently methylated sequences derived from transposable elements which may vary greatly among maize inbred lines (Fu and Dooner 2001), it seems reasonable to be concerned about the integration sites of the potentially valuable transgenes and the costs of evaluating such seemingly random events. But at this time, it appears as if each event integrates into a unique site in the maize nuclear genome and hence, potentially introduces a unique set of variables for that given chromosomal context in a given genotype. Methods that may minimize one aspect of such variation, position effect variegation, have been developed by including putative matrix attachment regions in the constructs (Brouwer et al. 2002). The lack of control on the site of integration is certainly a source of inefficiency for GAB of maize since it necessitates evaluating additional constructs and events.

A narrow spectrum of maize genotypes are directly amenable to current methods of transformation (Torney et al. 2006). Such genotypic specificity hinders complementation tests needed for functional genomics. For maize GAB, while it is always desirable to conduct genetic investigations directly with a choice of elite and commercially valuable germplasm, meaningful assessments can be made with less elite but transformable germplasm and promising candidate events are subsequently and quickly introduced into the appropriate gene pool (Armstrong et al. 1995). Attempts to broaden the range of transformable genotypes have achieved some success through QTL mapping and breeding (Armstrong et al. 1992) and bioinformatic-assisted transformation with genes known to influence developmental processes (Lowe and Gordon-Kamm 2004).

Maize transformation is far from ideal for GAB. The time and cost required to proceed from idea to commercial varieties may exceed 10 years and US$ 100 million; with about half of that time involved in transformation work. Currently, depending on the level and specificity of gene product needed for the desired phenotypic effect, it is necessary to produce and evaluate several constructs and hundreds of events to identify candidates for breeding and commercialization. The demands of product reliability are greater than that required for most aspects of basic research and the number of options to assess are increasing with the knowledge of genomes and their interactions with the environment. Perhaps with better understanding, a

higher form of knowledge, some of the assessments may be conducted through computational biology and simulation. Although maize transformation alone may never be completely adequate to efficiently evaluate all of the leads emerging from genomics, it will be the best option for assessing the most important ones.

3. GENOMICS-ASSISTED BREEDING FOR MAIZE: MOLECULR MARKERS AND DIAGNOSTICS

In contrast with maize transformation methods, the detection and utilization of nuclear DNA polymorphism is hardly a limiting step in some maize breeding programs (Ragot and Lee 2006; Table 1). Single nucleotide polymorphism has become the method of choice for GAB of maize. The maize nuclear genomes of elite gene pools seem to contain a sufficiently high frequency of variation for SNPs and the improving annotation of maize nuclear genomes provide enough sites to survey for nearly all applications of marker-assisted selection (MAS) and GAB of maize (Ching et al. 2002). In most breeding programs, DNA markers are used to some degree and in the largest maize breeding programs, the information is used at all stages of research and development for fingerprinting, selection for simple and complex traits and predictions of parents and their progeny (Crosbie et al. 2006; Ragot and Lee 2006).

Relative to indels and simple sequence repeats, SNPs may require assessments at more sites to provide the same degree of discrimination but that may be easily accommodated by the next generation of PCR-free methods of detection. Of course, such platforms for detecting DNA polymorphism are costly and technically dynamic (Syvanen 2001). This has required some rather expensive experimentation with platform-optimization that may be reasonable and possible only for the largest and more competitive programs and markets.

The current SNPs should only become better as their information content improves through functional genomics and genome annotation. Such ‘functional’ markers, derived from characterized sequence motifs and related validation will be an advance since many of the current SNPs have not been assigned a physiologically significant purpose through direct experimentation (Andersen and Lubberstedt 2003). For maize, the prospects for functional markers have been recently reviewed (Shi et al. 2005).

A potential, perhaps temporary, limitation of functional markers is that they focus on genic DNA sequences, and only the primary DNA sequence. Therefore, variation due to changes in methylation, local and long-distance regulatory sequences and relatively new categories of sequence (e.g. miRNA) will be missed (Lee 1999; Lee 2006). Components of the maize nuclear genome are frequently (de)methylated and the sequence space between genes is commonly tens of thousands of base pairs. Consequently, much functional information could reside in a large sequence space that is recalcitrant to superficial analyses, as indicated by the results recently reported in the cloning of *Vgt1* , a QTL that influences the transition from the

vegetative to the reproductive stage of maize (Salvi et al. 2007; see also Salvi and Tuberosa in Volume 1).

Advances in the development of functional markers will be achieved through the various platforms established for functional genomics and annotation. In maize, systems for transposon mutagenesis, EST resources, microarrays and TILLING have been established to complement the Maize Genetic Stock Center and other germplasm repositories. These resources, primarily playgrounds for basic research removed from GAB, have been described and listed (Shi et al. 2005; Lawrence et al. 2005). The current utility of those platforms for maize breeding suggests that further development is needed (Table 1). Also, association mapping has emerged as a potential complement of other methods although its utility in reference populations of elite germplasm will be diminished by the limited generations for recombination and potentially, by population structure (Yu and Buckler 2006). Clearly, association mapping must be allied with populations of simpler family structure (Laird and Lange 2006). The value of such resources and methods for GAB of maize, while yet to be realized, has been demonstrated in principle. For example, positional cloning of sequences underlying QTL has become a reality for well-integrated maize programs (see Salvi and Tuberosa, Volume 1) 2002). Also, a very promising technique for directed allele-silencing, RNAi has been used to produce a potentially simpler version of Quality Protein Maize (Segal et al. 2003). The development of DNA sequence databases, deeper and with improving annotation, should make RNAi and related approaches a more important option for functional genomics and GAB of maize, both of which are huge challenges when one considers that a complete DNA sequence of bacteriophage lambda was available in 1982 and functional analysis of such a simple system is still in progress (Lee 1999).

But, all journeys begin with a single step. With the genomic sequence of gene-enriched regions and ultimately most of a single maize genotype (B73) due in 2008, the informational infrastructure should be a relatively rich resource. The value of such a resource was demonstrated long ago in one of the first significant events of primordial GAB for maize, modification of the codons of *cryIA(b)* and *cryIA(c)* (Perlak et al. 1991). This is one of the first examples of bioinformatics and computational biology for GAB and it contributed to a significant achievement in maize improvement (Armstrong et al. 1995). Also, simple bioinformatic tools demonstrated that many maize ESTs contained simple sequence repeats and could be readily converted to functional markers (Senior et al. 1996). The value of bioinformatics and comparative genomics has been demonstrated in recent advances in maize transformation based on research with *Arabidopsis thaliana* (Lowe and Gordon-Kamm 2004). The *LEC1* gene and its relationship to ectopic development of somatic embryos was initially described in *Arabidopsis*. That observation led to the hypothesis that a similar gene with a similar function could exist in maize. The *Arabidopsis LEC1* DNA sequence helped identify a maize homologue which was demonstrated to elicit a positive tissue-culture response from previously recalcitrant maize inbreds. In the private sector, there is a steady stream of candidate genes from model systems, other plants and their maize homologues that are under

evaluation for efficacy in GAB and critical stages of those investigations are enabled and expedited by contemporary genomic platforms, bioinformatic tools and well-annotated databases such as those listed at MaizeGDB and similar sites (Lawrence et al. 2005; Table 1).

4. THE EMERGING MAIZE NUCLEAR GENOME AND GAB

Since 1997, I have reviewed features of the maize genome that present challenges for GAB and each time, as more information emerges, the list of challenges increases (Lee 1999a; 1999b; 2006). As we know, the maize genome is large, an ancient and fractionated tetraploid, volatile and contains somewhere between 30,000 to 60,000 genes embedded in various categories of DNA sequences, most of which were derived from transposable elements (San Miguel et al. 1996; Gaut and Doebley 1997; Haberer et al. 2005). The duplicate sequences and their products present a degree of structural and functional redundancy that hinders many attempts to link genotype with phenotype. In addition, noncollinearity of gene-like sequences and transposons are commonly observed in the elite maize gene pool (Fu and Dooner 2002; Brunner et al. 2005). If such noncollinearity proves to be of much functional significance then it may suggest a more informative basis for establishing maize gene pools and it may be related to phenomenon such as hybrid vigor (M. Lee, in preparation).

There is much to learn about the maize genome. Only a small fraction of the 30,000 to 60,000 genes has been assigned functions on the basis of direct experimentation. Many of the clues to function will come from comparative genomics and computational biology; but in the end, they must be validated in the right genotypes of the right species in the right environments. Also, there is a huge difference in the estimated number of maize genes and the actual number of maize ESTs in public and private repositories: most EST collections contain a few hundred thousand 'unique' sequences. What could explain the difference between the estimates and the collections? One explanation may be that a portion of maize genes, like those of humans, produce more than one transcript through alternate splicing and start/stop sites and antisense transcription (Lee 2006). Recent preliminary investigations in maize suggest that antisense transcripts are frequently detected (Morrow et al. 2006; Jia et al. 2006). Of course, assessing the function(s) of such transcripts will be the next step and it will more difficult than their initial identification. Other categories of transcripts (e.g. miRNA and siRNA) may also be contributing to the difference between gene number estimates and the number of ESTs of maize (Brodersen and Voinnet 2006). For maize, this is a largely unexplored and poorly-annotated aspect of the genome but there is at least one case in which miRNA has a clearly established function in development (Lauter et al. 2005).

The nonexpressed (to our knowledge) components of the maize genome are also mostly unexplored as well. The large distances between genes and gene islands of maize are mostly composed of highly variable sequences derived from transposons; all of which have the potential to contain regulatory sequences that could influence

expression of native genes and transgenes. Embedded in that mixture of intergenic DNA sequence are the known (i.e. consensus) and some unknown regulatory sequences. Identifying such sequences and their functions has been difficult for reasons that have been described (Lee 2006). A good example of the strategy and effort required to discover such sequences has been described for the *B* locus (Stam et al. 2002) and the *Vgt1* QTL (Salvi et al. 2007). In both cases, sequences involved in the regulation of transcription of the coding region were located more than 50 kilobase pairs upstream from the start of transcription. Given the structure of the maize genome, such long-distance regulation may be common but difficult to discern. Recently, a preliminary investigation at the *tb1* locus has also identified some evidence of long-distance regulation of expression (Clark et al. 2006). No doubt, if the lessons from careful investigation of regulatory networks in other species are of any predictive value for plants, these types of sequences and their regulatory effects will be important for GAB of maize (Davidson 2006).

Other nonexpressed DNA sequences may also influence genetic variation and effects in GAB era. For example, matrix or scaffold attachment regions (MARs or SARs) of various types have been shown to be important variables in mammalian gene expression. In maize, putative MARs have been identified and their effects on transgene expression have been assessed to a limited degree (Brouwer et al. 2002). The extent to which such sites vary among genotypes and how they influence gene expression and contribute to phenotype remains to be determined. Perhaps, the foundation for addressing some of those mysteries in a systematic manner will become apparent with the acquisition of an increasing amount of gene-related genomic sequence information once the sequencing of the maize genome is completed.

The phenotypic and genotypic volatility of the maize genome has been documented repeatedly; yet, parsimonious underlying mechanisms and explanations have not been validated. The phenotypic instability of some inbred lines has been described and suggests that, even when inbreds originate from doubled haploids (i.e. monoploids) significant changes in the genome may and do occur each sexual generation (Bogenschutz and Russell 1986). Facets of the potential genomic changes have been documented for a sample of inbred lines (Gethi et al. 2002). Such volatility could be significant for GAB of maize with respect to germplasm registration, aspects of intellectual property and maintenance of some germplasm collections.

Among the many gaps in knowledge and areas of limited exchange between maize breeding and maize genomics, the most important topic that has received the least attention is the analysis of interactions between the genotype and the environment (i.e. G x E). Perhaps, the platforms for such analyses have been only recently established or need to be developed. Admittedly, such analyses are difficult, risky and time-consuming relative to more focused investigations on molecules and their possible influences on phenotype. By their nature, basic geneticists and genomicists usually try to conduct experiments with materials and methods that maximize the genetic and minimize the environmental effects on the phenotype.

Plant breeders will seek similar information but in a much wider set of environments, the target environment of the cultivars and hybrids. For maize GAB, an improved understanding of the plant's receptors and 'senses' and subsequent responses will be important for modeling, prediction methods, gene design, testing strategies and data interpretation.

Two simple examples of the effects of temperature illustrate the potential significance of considerations of GxE for GAB. In maize, the activity of the transposon Ac2 seems to increase in ears exposed to higher temperature (Osterman 1991). In *Antirrhinum*, the activity of *Tam* transposons decrease 10-1000 times at higher temperatures (Coen et al. 1989). No doubt, other fundamental processes that influence genome content and stability are influenced by signals from the environment. In maize and other plants, there are many temperature-sensitive alleles and phenotypes but the basis of the sensitivity has been determined in only a few instances. Research on alleles of the *APETALA3 (APT3)* gene of *Arabidopsis* has shown that temperature sensitivity of the phenotype of the *ap3-1* allele, which ranges from mutant to near wild type, is determined by RNA processing and the splice variants produced under different conditions (Sablowski and Meyerowitz 1998). Unfortunately for maize GAB, the generation time, size and genetic nature of the plant hinder such carefully-controlled studies. Those circumstances will be a significant problem for the next phase of research in maize genomics and explorations of the maize genome; genotype by environment interactions.

5. SOME CHANGES IN MAIZE BREEDING WITH GENOMICS-ASSISTED BREEDING

The recent changes in maize breeding procedures and structure have been discussed (Lee 1995; Lee 1999; Betran et al. 2004; Kaeppler 2004; Troyer 2004; Crosbie et al. 2006; Ragot and Lee 2006). For maize breeding programs in competitive markets and cash-based economies supported by clear 'rule of law' the changes have been extensive. Some of the changes in personnel and infrastructure have been discussed in a previous section herein. Gradually, some of the agents of change have been adapted to and adopted by other markets, economies and regions in accordance with their conditions.

Perhaps the most obvious cause of change has been transgenic maize hybrids (aka GMO maize). Transgenic maize not only elevated societies' awareness of maize breeding, it also required maize breeding programs to develop parallel breeding programs, transgenic and nontransgenic (Lee 1999; Crosbie et al. 2006; Ragot and Lee 2007). Each transgene, and perhaps combinations of transgenes, must be reviewed by various national trade and regulatory agencies and the acceptance of transgenic maize has varied greatly among regions and countries. Some of the potential and actual risks of unintentional or poorly coordinated release of transgenic maize were clearly demonstrated in the late 1990s when StarLink maize hybrids were sold and grown prior to the completion of some regulatory procedures and approval (Troyer 2004). Consequently, maize breeding programs have adapted by

backcrossing transgenes into commercial and pre-commercial inbred lines in order to produce near-isogenic versions, transgenic and nontransgenic, of a given hybrid. The rapid and relatively precise backcrossing has been enabled by the development of suitable platforms for collecting DNA marker data and the use of either winter nurseries or greenhouses for producing several generations of maize per year. When functioning as an integrated system, such backcrossing has been invaluable by reducing the time required for product assessment and development.

Exotic germplasm has acquired an additional meaning in maize breeding. The potential gene pool of maize has truly expanded to include virtually any organism or design imaginable. In the private sector, assessing and sorting the many options presented by comparative genomics and model systems have become significant components of maize breeding. Considerations such as position effects, genetic background (Fu and Dooner 2002) and possible interactions of new sequences in old genomes have become routine for some maize breeding programs.

The advent of transgenic maize and some transgenes have influenced maize breeding programs in various ways. Their impact on sales and market share has been mentioned previously. The transgenes for resistance to *Ostrinia* and *Diabrotica* are good examples of how a few genes may influence an entire program. When the transgenes for insect resistance were initially backcrossed into a relatively wide array of elite germplasm, breeders observed that plant growth and development in some families of inbreds or genetic backgrounds declined when the transgene was introgressed. So, the genomes of those inbreds, or components of their genomes may be excluded from some aspects of breeding and product development. The initial backcrossing of those few transgenes may have revealed features of the genetic structure within maize gene pools that were unpredictable because very new genes and gene products were being introduced into relatively old genomes. Once the transcriptomics and metabolomics platforms mature then it should be worthwhile to establish large-scale profiles of maize gene pools based on transcript and metabolites profiles (Table 1). Such information may minimize the surprises related to introduction of truly new transgenes and it should enable modeling and simulation methods that have better predictive ability.

The primary effect of the transgenes for resistance to *Ostrinia* and *Diabrotica* provided indirect benefits such as a lower incidence of disease in the stalk and grain and a root system more capable of maximum use of soil water. Such indirect benefits have partially enabled an increase in the planting density used by farmers. As planting densities continue to increase, breeders have discovered that certain families of inbreds have root systems that fail under some conditions. Actually, breeders have been making such discoveries for several decades prior to the advent of transgenic maize but at least some of the recent observations and changes in germplasm have been accelerated by the introduction of a few transgenes that have minimized some sources of variation so that maize breeders could identify previously obscured sources of variation.

Transgenes have also renewed the debate regarding the type of inbred line to use as a tester or common parent when evaluating populations and candidate inbreds

for hybrid performance. The direct and indirect effects of some transgenes could obscure some weaknesses in the stalk, as well as resistance to biotic stress and root systems of the candidate inbreds and populations. That would reduce the selection pressure and the opportunities to observe the native features of the germplasm.

Another important change has been the utilization of DNA sequence polymorphism and associated information. Several programs in the private sector have established centralized service laboratories capable of rapid, large-scale and cost-effective (per data point) collection of DNA marker data so that such information may be collected for all inbred lines and populations at various stages of breeding (Ragot and Lee 2006). Such information has obvious utility for quality control, assessments of genetic relationships, establishment of intellectual property and some stages of selection (Smith and Smith 1992; Niebur et al. 2004).

The capacity to screen tens of thousands of accessions at a similar number of sites in the genome provides a new and targeted approach to germplasm evaluation and use of exotic germplasm within the maize gene pool. On the basis of knowledge about relationships between DNA sequence and properties of gene products, it should be possible to conduct sequence searches or design assays to search for specific sequences within certain genes to determine if such a motif exists in the native gene pool. Then, the few candidate alleles could be quickly and precisely introgressed into the elite, reference population(s) (Lee 1995). For some phenotypes, this approach could be a viable alternative to transgenic approaches and it would reduce some of the concerns and costs of that option.

Improvements in the acquisition of DNA marker data have revolutionized two old and reliable methods of maize breeding, backcrossing and recurrent selection. Marker-assisted backcrossing, already described and reviewed elsewhere (Ragot and Lee 2006), has become a standard component of breeding programs that need to introgress transgenes, native genes and QTL. That development was quite predictable (Lee 1995). In contrast, the ability to conduct marker-assisted recurrent selection (MARS) for QTL and complex traits such as grain yield has been a surprise of GAB era. There are several good reasons why MAS should not succeed in such situations (Holland 2004). The initial reports of MAS for complex traits and QTL in maize were mostly negative with the exception of the special case of sweet corn (Moreau et al. 2004; Edwards and Johnson 1994). Subsequent studies on a larger scale in several elite reference populations indicate that MARS for QTL and complex traits was an effective and competitive strategy relative to phenotypic selection (Johnson 2004; Crosbie et al. 2006). The changing fortunes of MAS should encourage reviews and reconsideration of other failed attempts to use DNA markers at other stages of selection such as choice of parents and methods for predicting and testing hybrids (Bernardo 1994). The advent of functional markers and new insights into patterns of gene expression of inbreds and hybrids could be important components of improved methods to those significant challenges of maize breeding (Lee 1999; Swanson-Wagner et al. 2006).

Two final components for GAB of maize should be briefly described, namely the places where maize breeding occurs and the people who conduct the breeding.

The personnel involved and their roles in some of today's maize breeding programs would not be recognizable to an experienced maize breeder of 30 years ago. GAB of maize truly depends on the contributions of lawyers, computational biologists, bioinformaticians, information technologists, biochemists, cell biologists, developmental biologists, molecular biologists as well as the previously recognizable collection of statisticians, pathologists, entomologists, geneticists, physiologists and plant breeders. The intellectual skill sets of a contemporary maize breeder have expanded to include association mapping and Bayesian-based statistics while the resources for educating the next generation of plant breeders have decreased. The stages of maize breeding have been organized into a series of activities resembling an assembly line of a factory in which a breeder may be responsible for the development of new populations and the early generations of selection while other colleagues would focus on later stages of hybrid testing and commercialization. The extent, to which the various specialists collaborate, especially in the private sector, has been a key to some of the early success of GAB of maize.

The places in which maize breeding occurs have also changed with GAB (Ragot and Lee 2006). Structures and climates that facilitate the production of 3-4 generations per year have become much more important components of infrastructure. In several programs, methods that produce doubled-haploids have been implemented to various degrees as a means of reaching homozygosity as quickly as possible (Melchinger et al. 2005). With DNA markers and associated information, it has become possible to conduct effective selection in environments other than the target environment; a huge change for maize breeding. Also, components of valuable germplasm may be created in any laboratory. The next great commercially valuable transgene will probably originate from a thermal cycler and a method based on DNA shuffling (Stemmer 1994). Maize breeding has changed and the roles of the maize breeder have also changed.

With this collection of new tools, information and personnel, maize breeding programs have initiated projects to address some of the most challenging limitations encountered by crop-based agriculture. Indications from the private sector suggest that genes for improved tolerance to water deficits have been isolated from *Arabidopsis* and have shown promising results in maize. In parallel, the CGIAR's Generation Challenge program has been supporting genomics-based investigations towards improved drought tolerance of maize (http://www.generationcp.org/vw/index.php). Also, the HarvestPlus program has initiated genomics-based and MAS programs for biofortification of maize grain (Pixley et al. 2006). In the near future, GAB and GAS of maize should deliver germplasm with improved protein composition, resistance to parasites such as *Striga* and the capacity to eliminate toxins (e.g. mycotoxins) produced by pathogens. Decades of well-intentioned conventional selection have achieved limited success for those phenotypes.

The arrival of the genomics era of maize breeding was immediately preceded by a significant increase in the rate and volume of field-based assessments of candidate hybrids. The increase was driven by advances in computational power,

plot equipment, modifications to off-season nurseries and an overall expansion in the infrastructure that supports maize breeding (Smith et al. 1999 and Table 1). Maize breeding had become a year-round activity by the late 1980s. The advent of the genomics era of maize breeding, even at such an early stage, has clearly stressed programs' abilities to empirically assess all of the promising genetic options and questions that are raised by new sources of information. So, field-based phenotyping and the prioritization of phenotyping has become much more important in the genomics era of maize breeding (Table 1). In some regards, this is not new for some programs because testing combinations of inbred lines in hybrids has been a challenge for decades (Smith et al. 1999). As with the challenge of predicting performance of untested hybrids, some of the many questions and opportunities raised by the early stages of genomics will be addressed through modeling and simulation to predict aspects of plant growth and development in response to changes in genetic and environmental variables (Hammer et al. 2006). It is certainly premature to assess the impact of such efforts on maize breeding because the current models are based on limited information (Table 1). However, it is clear that such activity will become more important to maize breeding because it is simply too expensive and probably impossible to directly assess a fraction of the many options that will be enabled by the maturation of the genomics era.

In some ways, maize breeding has come a long way since the initial reports of Shull and colleagues nearly 100 years ago. The gains and pains of artificial selection in maize have been well documented for some traits and production environments (Duvick et al. 2004; Duvick 2006). But, with the exception of improved resistance to biotic and abiotic stress as a basis for genetic gains in grain yield, very little fundamental information and few (if any) validated mechanisms have been reported to account for those gains. The next era of maize breeding should be fundamentally different in the sense that the strategies will be based more on a scientific method and some of the reasons for success or failure should be known and incorporated into more efficient methods of maize improvement.

ACKNOWLEDGEMENTS

The author wishes to thank the editors and reviewers for their helpful suggestions for this chapter, his students and colleagues for the chance to learn together, the international agencies for the chance to learn about maize breeding in so many regions of the world, colleagues in the private sector for the opportunities to learn about commercial maize improvement and problems of contemporary genomics and the many who have made maize breeding and genetics so stimulating and productive since Schull's reports in 1909.

REFERENCES

Andersen JR, Lubberstedt T (2003) Functional markers in plants. Trends Plant Sci 11:554–560

Armstrong CL (1999) The first decade of maize transformation: a review and future perspective. Maydica 44:101–109

Armstrong CL, Romero-Severson J, Hodges TK (1992) Improved tissue culture response of an elite maize inbred through backcross breeding, and identification of chromosomal regions important for regeneration by RFLP analysis. Theor Appl Genet 84:755–762

Armstrong CL, Parker GB, Pershing JC, Brown SM, Sanders PR, Duncan DR, Stone T, Dean DA, Deboer DL, Hart J, Howe AR, Morrish FM, Pajeau ME, Petersen WL, Reich BJ, Rodriguez R, Santino CG, Sato SJ, Schuler W, Sims SR, Stehling S, Tarochione LJ, Fromm ME (1995) Field evaluation of European corn borer control in progeny of 173 transgenic corn events expressing an insecticidal protein from *Bacillus thuringiensis*. Crop Sci 35:550–557

Bernardo R (1994) Prediction of maize single-cross performance using RFLPs and information from related hybrids. Crop Sci 34:20–25

Betran FJ, Menz M, Banziger M (2004) Corn breeding. In: Smith CW, Betran FJ, Runge ECA, (eds) Corn: origin, history, technology, and production. Wiley, Hoboken, NJ

Bogenschutz TG, Russell WA (1986) An evaluation for genetic variation within inbred lines maintained by sib-mating and self-pollination. Euphytica 35:403–412

Brodersen P, Voinnet O (2006) The diversity of RNA silencing pathways in plants. Trends Plant Sci 22:268–280

Brouwer C, Bruce W, Maddock S, Avramova Z, Bowen B (2002). Suppression of transgene silencing by matrix attachment regions in maize: Dual role for the Maize 5′ ADH1 matrix attachment region. Plant Cell 14:2251–2264

Brunner S, Fengler K, Morgante M, Tingey S, Rafalski A (2005). Evolution of DNA sequence nonhomologies among maize inbreds. Plant Cell 17:343–360

Charles D (2001) Lords of the Harvest. Perseus Publishing, Cambridge, MA

Cheng M, Lowe BA, Spencer TM, Xe Y, Armstrong CL (2004). Invited review: factors influencing agrobacterium-mediated transformation of monocotyledonous species. In Vitro Cell Dev Biol 40: 31–45

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

Clark R, Nussbaum-Wagler T, Quijada P, Doebley JF (2006) A distant upstream enhancer at the maize domestication gene, *tb1*, has pleiotropic effects on plant and inflorescent architecture. Nature Gen 38:594–597

Coen E, Robbins TP, Almeida J, Hudson A, Carpenter R (1989) Consequences and mechanisms of transposition in *Antirrhinum majus*. In: Berg DE, Howe MM (eds) Mobile genetic elements. American Society for Microbiology, Washington DC, pp 413–436

Crosbie TM, Eathington SR, Johnson GR, Edwards M, Reiter R, Stark S, Mohanty RG, Oyervides M, Buehler RE, Walker AK, Dobert R, Delannay X, Pershing JC, Hall MA, Lamkey KR (2006) Plant breeding: past, present, and future. In: Lamkey KR, Lee M (eds) Plant breeding: the Arnel R. Hallauer international symposium Blackwell Publishing, Ames, IA

Davidson EH (2006) The regulatory genome. Gene regulatory networks in development and evolution. Elsevier, New York

Duvick DN (2005) Contribution of breeding to yield advances in maize. In: Sparks DN (ed), Advances in agronomy, Vol. 86, Academic Press, San Diego, CA pp 83–145

Duvick DN, Smith JSC, Cooper M (2004) Changes in performance, parentage, and genetic diversity of successful corn hybrids, 1930–2000. In: Smith CW, Betran FJ, Runge ECA (eds) Corn: origin, history, technology, and production. Wiley, Hoboken, NJ

Eathington SR (2005) Practical applications of molecular technology in the development of commercial maize hybrids. Proceedings of the 60th annual corn and sorghum seed research conferences. American Seed Trade Association, Washington, DC

Edwards M, Johnson L (1994) RFLPs for rapid recurrent selection. In: "Proceedings of symposium on analysis of molecular marker data" Am Soc Hort Sci and CSSA, Corvallis, OR pp 33–40.

Frisch M, Bohn M, Melchinger AE (1999) Comparison of selection strategies for marker-assisted backcrossing of a gene. Crop Sci 39:1295–1301

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

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

Gethi JG, Labate JA, Lamkey KR, Smith ME, Kresovich S (2002) SSR variation in important US maize inbred lines. Crop Sci 42:951–957

Haberer G, Young S, Bharti AK, Gundlach H, Raymond C, Fuks G, Butler E, Wing RA, Rounsley S, Birren B, Nusbaum C, Mayer KFX, Messing J (2005) Structure and architecture of the maize genome. Plant Physiol 139:1612–1624

Hamilton CM, Frary A, Lewis C, Tanksley SD (1996) Stable transfer of intact high molecular weight DNA into plant chromosomes. Proc Natl Acad Sci 93:9975–9979

Hammer G, Cooper M, Tardieu F, Welch S, Walsh B, van Eeuwijk F, Chapman S, Podlich D (2006). Models for navigating biological complexity in breeding improved crop plants. Trends Plant Sci 11:587–593

Holland, JB (2004) Implementation of molecular markers for quantitative traits in breeding programs – challenges and opportunities. In: Fischer T et al (eds) New directions for a diverse planet. Proceedings of the 4th International Crop Science Congress, Brisbane, Australia, 26 September–1 October 2004, www.cropscience.org.au

Hospital F, Chevalet C, Mulsant P (1992) Using markers in gene introgression breeding programs. Genetics 132:1199–1210

Jia Y, Swanson-Wagner R, Emrich S, DeCook R, Fu Y, Guo L, Borsuk L, Ashlock D, Nettleton D, Schnable P (2006) Widespread maize natural antisense transcripts are frequently located in UTR repeat regions and conserved in other cereal transciptome using strand-specific microarray hybridized with two maize inbreds In: Smith CW, Betrán J, Runge ECA, (eds). Poster 208. 48th Annual Maize Genetics Conference

Johnson GR (2004) Marker assisted selection. In: Janick J (ed) Plant breeding reviews. Vol. 24 part 1, Wiley, Hoboken, NJ, pp 293–310

Kaeppler S (2004) Biotechnology: new horizons. In: Smith CW, Betran FJ, Runge ECA (eds) Corn: origin, history, technology, and production. Wiley, Hoboken, NJ pp 399-425

Laird NM, Lange C (2006) Family-based designs in the age of large-scale gene-association studies. Nature Rev Genet 7:385–394

Lauter N, Kampani A, Carlson S, Goebel M, Moose SP (2005) MicroRNA172 downregulates glossy15 to promote vegetative phase change in maize. Proc Natl Acad Sci USA 102:9412–9417

Lawrence CJ, Seigfried T, Brendel V (2005) The maize genetics and genomics database. The community resource for access to diverse maize data. Plant Physiol 138:55–58

Lee M (1995) DNA markers in plant breeding programs. Adv Agron 55:265–344

Lee M (1999a) Genomic technologies and plant breeding. In: MX dos Santos (ed) XVIII Reunion Latinoamericana del Maiz. International center for maize and wheat improvement, El Batan, Mexico, pp 95–114

Lee M (1999b) Towards understanding and manipulating heterosis in crop plants: can molecular genetics help? In: Coors JG, Pandey S (eds) The genetics and exploitation of heterosis in crops. American Society of Agronomy, Madison, WI, pp 185–194

Lee M (2006) The phenotypic and genotypic eras of plant breeding. In: Lamkey KR, Lee M (eds) Plant breeding: the Arnel R Hallauer international symposium. Blackwell Publishing, Ames, I.

Lowe KS, Gordon-Kamm WJ (2004) Methods of use of LEC1 polynucleotides and polypeptides. United States Patent Application 20040016022

Maliga, P (2004) Plastid transformation in higher plants. Ann Rev Plant Biol 55:289–313.

Melchinger AE, Longin CF, Utz HF, Reif JC (2005) Hybrid maize breeding with doubled haploid lines: quantitative genetic and selection theory for optimum allocation of resources. 41st Illinois Corn Breeders' School, University of Illinois at Urbana-Champaign, pp 8–21

Moreau L, Charcosset A, Gallais A (2004) Experimental evaluation of several cycles of marker-assisted selection in maize. Euphytica 137:111–118

Morrow D, Ma J, Fernandes J, Walbot V (2006) Comparative profiling of the sense and antisense transcriptome of maize. Poster 162. 48th Annual Maize Genetics Conference

Niebur WS, Rafalski JA, Smith OS, Cooper M (2004) Applications of genomics technologies to enhance rate of genetic progress for yield of maize within a commercial breeding program. In: Fischer T et al (eds) New directions for a diverse planet. Proceedings of the 4th International Crop Science Congress, Brisbane, Australia, 26 September–1 October 2004. www.cropscience.org.au

Openshaw S, Frascaroli E (1997) QTL detection and marker-assisted selection for complex traits in maize. Proceedings of the 52nd annual corn and sorghum research conference. American Seed Trade Association, Washington, DC, pp 44–53

Osterman JC (1991) Transposition of *Ac-2* in response to temperature. Maydica 36:147–151

Ow DW (2002) Recombinase-directed plant transformation for the post-genomic era. Plant Mol Biol 48:183–200

Perlak FJ, Fuchs RL, Dean DA, McPherson SL, Fischoff DA (1991) Modification of the coding sequence enhances plant expression of insect control proteins. Proc Natl Acad Sci 88: 3324–3328.

Pixley K, Beck D, Palacios N, Gunaratna N, Guimaraes PE, Menkir A, White WS, Nestel P, Rocheford T (2007) Opportunities and strategies for biofortified maize (in review)

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

Ragot M, Lee M (2007) Marker Assisted Selection in maize: current status, potential, limitations, and perspectives from the private and public sectors. In: ElcioG (ed) Marker-assisted selection in agriculture. Food and Agriculture Organization (FAO), United Nations Rome, Italy, pp. 117–150.

Sablowski RWM, Meyerowitz EM (1998) Temperature-sensitive splicing in the floral homeotic mutant apetala3-1. Plant Cell 10:1453–1463

Salvi S, Tuberosa R (2005) To clone or not to clone plant QTLs: present and future challenges. Trends Plant Sci 10:297–304

Salvi S, Sponza G, Morgante M, Fengler K, Meeley R, Ananiev E, Svitashev S, Bruggemann E, Niu X, Li B, Hainey CF, Rafalski A, Tingey SV, Tomes D, Miao G-H, Phillips RL, Tuberosa R (2007) Conserved non-coding genomic sequences controlling flowering time differences in maize. Proc Natl Acad Sci USA 104:11376–11381

Segal G, Song R, Messing J (2003) A new opaque variant of maize by single dominant RNA-interference-inducing transgene. Genetics 165:387–397

Senior ML, Chin ECL, Lee M, Smith JSC, Stuber CW (1996) Simple sequence repeat markers developed from maize sequences found in the GENEBANK database: map construction. Crop Sci 36: 1676–1683

Shi C, Wenzel G, Frei U, Lübberstedt T (2005) From genomics to functional markers in maize. Prog Bot 67:53–70

Shou H, Frame BR, Whitham SA, Wang K (2004) Assessment of transgenic events produced by particle bombardment or *Agrobacterium*-mediated transformation. Mol Breed 13:201–208

Shull GH (1909) The composition of a field of maize. Rep Am Breeders' Assoc 4:296–301

Smith JSC, Smith OS (1992) Fingerprinting crop varieties. Adv Agron 47:85–140

Smith OS, Hoard K, Shaw F, Shaw R (1999) Prediction of single-cross performance. In: Coors JG and Pandey S (eds) The genetics and exploitation of heterosis in crops. American Society of Agronomy, Madison, WI pp 277–285

Stam M, Belele C, Ramakrishna W, Dorweiler JE, Bennetzen JL, Chandler VL (2002) The regulatory regions required for B′ paramutation and expression are located far upstream of the maize b1 transcribed sequences. Genetics 162:917–930

Stemmer WPC (1994) DNA shuffling by random fragmentation and reassembly: *In vitro* recombination for molecular evolution. Proc Natl Acad Sci 91:10747–10751

Stromberg LD, Dudley JW, Rufener GK (1994) Comparing conventional early generation selection with molecular marker assisted selection in maize. Crop Sci 34:1221–1225

Swanson-Wagner R, Jia Yi, DeCook R, Borsuk L, Nettleton D, Schnable P (2006) All possible modes of gene action are observed in a global comparison of gene expression in a maize F1 hybrid and its inbred parents. Proc Natl Acad Sci 103:6805–6810

Syvanen AC (2001) Accessing genetic variation: genotyping single nucleotide polymorphisms. Nat Rev Genet 2:930–942

Torney F, Frame B, Wang K (2007) Maize. In: Pua EC, Davey MR Biotechnology in Agriculture and Forestry Transgenic Crops IV. Springer, Berlin, pp. 73–98

Troyer AF (2004) Persistent and popular germplasm in seventy centuries of corn evolution. In: Smith CW, Betrán J, Runge ECA (eds) Corn: origin, history, technology, and production Wiley, Hoboken, NJ, pp 133–232

Varshney RK, Graner A, Sorrels ME (2005) Genomics-assisted breeding for crop improvement. Trends Plant Sci 10:621–630

Wright DA, Townsend V, Winfrey RJ Jr, Irwin PA, Rajagopol J, Lonosky PM, Hall BD, Jondle MD, Voytas DF (2005) High-frequency recombination in plants mediated by zinc-finger nucleases. Plant 44:693–705

Yu J, Buckler ES (2006) Genetic association mapping and genome organization of maize. Curr Opin Biotechnol 17:1–6

CHAPTER 7

MOLECULAR MARKERS AND MARKER-ASSISTED SELECTION IN RICE

DAVID J. MACKILL*
International Rice Research Institute, DAPO Box 7777, Metro Manila, Philippines

Abstract: The status of rice as a model crop and the sequencing of the *indica* and *japonica* genomes have provided breeders with the necessary tools for marker assisted breeding. Simple Sequence Repeat (SSR) markers are easily available for any region of the genome, and candidate gene markers are being developed rapidly. The likely targets of MAS include yield and agronomic traits, cooking and nutritional quality, and resistances to abiotic and biotic stresses. MAS for gene pyramiding for disease and insect resistances is being widely used. For major genes and QTLs of larger effect, marker assisted backcrossing (MAB) is an effective method for developing improved versions of widely-grown "mega" varieties. Developing submergence tolerant mega varieties is a good example of how the MAB approach can result in significantly improved mega varieties within two to three years. The use of markers in more conventional breeding nurseries has been limited by cost, but is beginning to be applied for some traits like grain quality. Lower-cost marker methods combined with large-scale gene discovery will increase the use of MAS over the next decade.

1. INTRODUCTION

As a model cereal crop, rice has benefited greatly from the advances in plant genomics. The completed genome sequence of *japonica* cultivar Nipponbare and the complete draft DNA sequence of *indica* cultivar 93-11 have provided the basis for advanced gene discovery, complementing the detailed investigations made for the dicot Arabidopsis. The existence of multiple mutant collections, cDNA and EST databases, and whole-genome chips has enabled large-scale functional genomics applications (Leung and An 2004). Some cultivars, particularly of the *japonica* type, can be readily transformed through the Agrobacterium approach, permitting gene validation through complementation and overexpression studies.

*Address for correspondence: d.mackill@cgiar.org

R.K. Varshney and R. Tuberosa (eds.), Genomics Assisted Crop Improvement: Vol. 2: Genomics Applications in Crops, 147–168.

The mapping of genes controlling agronomic traits coupled with the widespread availability of easy to use simple sequence repeat (SSR) markers (McCouch et al. 2002) and quick DNA extraction methods (Xu et al. 2005b) has provided breeders with an excellent opportunity to apply marker assisted selection (MAS) methods in rice. However, most rice breeders have not taken advantage of this new approach, due to insufficient laboratory facilities, lack of trained staff, high costs, or lack of clear objectives amenable to MAS. With continued cost reductions, and a focus on introgressing useful genes into elite backgrounds, the wider adoption of MAS, and marker assisted backcrossing (MAB) in particular, is expected for rice. The following review will focus on general strategies for MAS in rice, and will emphasize traits where there is a clear advantage for using the technique.

2. MOLECULAR MARKERS FOR RICE BREEDING

Molecular markers can have a number of applications in agriculture, and their application in rice improvement has been recently reviewed (Mackill and McNally 2004; Collard and Mackill 2007). Rice scientists have greatly benefited by the high quality genome sequence of *japonica* cultivar Nipponbare (IRGSP 2005), and the less-refined sequence of the *indica* cultivar 93-11 (Yu et al. 2002). In addition to the large number of validated SSR markers (McCouch et al. 2002 and at http://www.gramene.org/), the IRGSP identified 18,828 class I SSRs throughout the genome (IRGSP 2005). These markers are also effective with most wild species used as sources of new genes or alleles (Gao et al. 2005). The costs and ease of high-throughput DNA extraction are declining (Xu et al. 2005b), making it more practical to apply MAS methods in breeding populations.

The major requirement for application of molecular markers in rice improvement is that a gene or QTL of significant effect be mapped with a high degree of accuracy, and that the gene is effective in the desired genetic background. Furthermore, the gene should not have negative effects on other traits. In initial mapping studies, a gene is generally mapped at low resolution using small populations (<200). For MAS, more tightly linked markers are required. Hence, the gene should be fine-mapped at a higher resolution. With populations typically higher than 400, genes can be mapped at less than 1 cM resolution. This will usually yield markers that are sufficient for MAS. However, further analysis resulting in positional cloning of the gene will be beneficial for MAS. If the candidate gene is identified, markers within or closely linked to the gene can be developed. These markers will not undergo recombination in breeding populations, and they are usually diagnostic of the favorable allele in a wide range of germplasm. If the functional nucleotide polymorphism is identified this will lead to a more powerful, gene-based or functional marker (FM) (Andersen and Lubberstedt 2003), sometimes referred to as a "perfect marker". However, with SSR markers still in wide use, the SSR markers within or adjacent to the gene will remain valuable for MAS.

A marker that is non-recombinant with the target gene can be used to reliably predict the presence of the gene in segregating populations. A codominant marker is preferred so that heterozygotes can be detected in backcross generations. The marker can be used in combination with flanking markers to reduce the effect of linkage drag when transferring only the desired QTLs or genes into an elite genetic background (Tanksley et al. 1989; Frisch et al. 1999b; Hospital 2001). In combination with selection against unlinked donor markers, this method can be used to quickly and cleanly transfer QTLs or genes into elite cultivars, providing the best assurance that the traits of the recurrent parent are accurately recovered. This is the most practical application of markers in rice improvement. Markers can also be used in normal breeding nurseries, although the number of plants or progeny grown in such nurseries is usually much higher, making it more costly.

New types of markers promise to bring down the cost of MAS and make it more attractive for breeders to use. Single Nucleotide Polymorphisms (SNPs) are the most common type of polymorphism in plants. For the two rice subspecies, 80,127 polymorphic sites were identified (IRGSP 2005). High throughput assays for SNPs are being developed to allow more widespread use of these markers. Another promising marker is the Single Feature Polymorphism (SFP), short deletions that are detected using oligo arrays hybridized with genomic DNA to perform whole-genome scans (Hazen and Kay 2003).

3. TARGET TRAITS FOR MARKER ASSISTED SELECTION

The following discussion will cover only those genes or traits where MAS is a realistic option in current breeding programs. This requires that the trait is generally an important objective in rice breeding programs. It also requires that valuable genes or QTLs have been identified, and that MAS offers a significant advance over what could be achieved without the use of markers (i.e. through strictly phenotypic selection). This survey is not considered exhaustive, and emphasis is on the more important breeding objectives and on recent literature. More detailed information on mapped genes is available at the Gramene web site (http://www.gramene.org/).

The decision on which traits are of relevance for MAS is not trivial and has been made in this article based on which traits seem likely to benefit most from using markers. This means that a number of traits are left out, but might be of particular relevance for specific breeding objectives. An example would be flowering duration, which is considered an important agronomic trait. The elegant work by Yano and colleagues has identified the most important genes underlying QTLs for flowering duration and photoperiod sensitivity (Izawa et al. 2003). Most breeders can easily and accurately measure flowering duration in their breeding materials and would not consider this trait a high priority for MAS. However, where whole-genome monitoring is used to combine the essential requirements for a suitable variety, QTLs for flowering duration could be selected

for. MAB methods might also be appropriate if a breeder would like to change the duration of a very desirable cultivar by introducing a different heading date allele.

3.1. Grain Yield and Agronomic Traits

3.1.1. Grain yield

Grain yield is one of the most important breeding objectives, but in practice it is often the most difficult to improve on. Historically, the major advances in increasing yield of rice came from the increase in harvest index brought about largely because of the development of semidwarf cultivars in the 1960s and 1970s (Peng et al. 2000). Some increases in biomass have been observed in more recent cultivars (Peng et al. 2000), but increasing the yields further for tropical rices has been very difficult. In general, hybrid rice is thought to be a more promising approach to increasing the yield potential in rice over the existing inbred cultivars. There have been some recent advances in raising the yield potential through the 'super hybrid' rice developed in China (Peng et al. 2004).

Grain yield does not appear to be a very promising target for marker assisted selection due to the complicated genetic inheritance. The QTLs almost always have small effects and are specific to genetic background (Guo et al. 2005). It is not uncommon to find larger effects of QTLs where one of the parents is a low-yielding type, but it is simplistic to think that these alleles would confer a significant advantage in the genetic background of modern elite varieties. Some researchers have noted QTLs of larger effect for specific yield components. For example Mei et al. (2006) found some useful QTLs for number of spikelets per panicle, which was related to number of secondary panicle branches. Takai et al. (2005) detected QTLs of moderate effect for grain filling percentage. A QTL for yield potential on chromosome 5, *rg5*, which represented an allele from the *indica* cultivar Kasalath studied in the background of *japonica* Nipponbare, was thought to be of use in different genetic backgrounds (Ishimaru et al. 2005).

Many researchers have applied the Advance Backcross QTL (ABQTL) approach (Tanksley and Nelson 1996) to rice for discovering QTLs with positive effects on agronomic traits derived from wild species (Xiao et al. 1998; Moncada et al. 2001; Septiningsih et al. 2003; Thomson et al. 2003; Lee et al. 2005b). Tian et al. (2006) used an accession of 'Dongxiang' wild rice (*O. rufipogon*) to develop intgrogression lines with small segments transferred to an elite cultivar. In general, the lines with more introgressed segments from wild rice performed poorly; however, some favorable alleles were identified for grain number on chromosomes 5, 8 and 9. They found that QTL analysis in the BC2F2 generation was only a moderate predictor of the effects in the introgression lines.

By map-based cloning Ashikari et al. (2005) isolated a cytokinin oxidase conferring higher grain number and yield. The Null allele for this gene (OsCKX2) conferred higher grain number, and they developed NILs through a MAB approach.

This gene was combined with the major semidwarf gene *Sd1*, corresponding to a mutant allele at a GA_{20} oxidase gene (Sasaki et al. 2002), in a process they termed "tailor-made crop improvement".

3.1.2. Heterosis and hybrid rice

Attempts to identify QTLs affecting heterosis have had some success, although some authors have noted that it is difficult to find a correlation between F1 performance and parental genotypes (Li et al. 2001; Luo et al. 2001). You et al. (2006) observed that main effect QTLs for agronomic traits could contribute cumulatively to genetic variation, and larger effects could be observed in the hybrids, with relatively lower contributions of epistatic effects. A region near the *Sd1* locus on chromosome 1 seemed to have the largest effect. Mei et al. (2005), in two BC and a RIL population using Lemont (*japonica*) and Teqing (*indica*) indicated that the primary genetic basis for heterosis was overdominance from epistatic loci, while Abdelkhalik et al. (2005) suggested it was from the cumulative effects of partially dominant loci. Hua et al. (2003) found little overlap between heterotic effects and trait performance. They found that heterosis could be explained by single-locus effects in combination with dominance by dominance interaction at the two-locus level.

The relatively complicated genetic control of yield and heterosis has not allowed the identification of promising markers for MAS; however, the production of hybrid rice involves several concomitant traits that are readily amenable to MAS. Most currently grown hybrids rely on the cytoplasmic male sterility (cms) system, which requires that the male parent possess genes for restoration ability. These restorer genes are very tedious to manipulate in breeding because they can only be assayed in mature plants from testcross progeny. For the widely used WA type cms, the *Rf3* gene has been mapped on chromosome 1 (Zhang et al. 1997) and an additional locus has been mapped on chromosome 10 (Yao et al. 1997). For the Chinsurah Boro II type cms, the *Rf1* gene has been cloned (Yao et al. 1997), and Wang et al. (2006) found that the locus involved two related PPR (pentatricopeptide repeat) genes (*Rf1a* and *Rf1b*).

The 2-line system of hybrid rice production involves genetic male sterility genes that are sensitive to temperature and/or photoperiod. This includes the genes designated *tms2* (Lopez et al. 2003), *tms3* (Lang et al. 1999), *tms4* (Dong et al. 2000), *tms5* (Wang et al. 2003b), and *tms6* (Wang et al. 2004; Lee et al. 2005a). With one or a combination of these genes, a breeding line is developed that will be fertile under low-temperatures and/or short daylength for seed maintenance, but will be sterile under higher temperatures or longer daylengths for hybrid seed production.

Wide compatibility (WC) is a very useful trait for conferring spikelet fertility to the hybrids from wide crosses (usually *indica* × *japonica* crosses). Several WC genes have been identified, but the *S5* gene on chromosome 6 has been most widely used (Ikehashi and Araki 1986). Candidate genes for this locus have been reported (Ji et al. 2005; Qiu et al. 2005). Several other QTLs in the WC cultivar Dular have been mapped and targeted for MAS (Wang et al. 2005). Another useful gene for hybrid rice is elongated uppermost internote (*eui*), a recessive tall allele that can be

used in the female parent to increase panicle exsertion for hybrid rice production. Several groups have reported the *eui1* gene to be a putative cytochrome P450 gene (Luo et al. 2006; Ma et al. 2006; Zhu et al. 2006).

3.1.3. Other agronomic traits

A number of agronomic traits of value in breeding programs have been identified. These include seedling vigor (Kanbar et al. 2006), seed longevity or dormancy (tolerance of pre-harvest sprouting) (Wan et al. 2006; Yamaguchi et al. 2006), and grain shattering (Li et al. 2006). While these are all useful traits, it is not clear if they would be high-priority targets for MAS in rice breeding programs.

3.2. Grain Quality and Nutrition

3.2.1. Cooking quality and grain appearance

Both cooking quality (taste and texture, largely) as well as grain appearance are high-priority traits in nearly all rice breeding programs. Rice starch is composed of amylose and amylopectin, and the percentage of amylose is an important characteristic of cooking quality. Amylose content is largely determined by the *wx* locus on chromosome 6, and in particular on the amount of Wx protein (granule-bound starch synthase, GBSS) present (Wang et al. 1995). Varieties with near zero amylose and no Wx protein are called glutinous or waxy, and are used for special foods such as desserts and snacks. High amylose cultivars (>25%), with high levels of Wx protein, are common in *indica* rices, and cook dry and fluffy, often becoming hard after cooling. Low amylose cultivars (15–20%) is soft and sticky; nearly all temperate *japonica* cultivars have low amylose. Intermediate amylose (20–25%) is widely preferred in many parts of the tropics. These rices are soft but not sticky. The level of Wx protein is related to the level of mature Wx mRNA, which is related to the stage of intron 1 excision from pre-mRNA (Cai et al. 1998; Isshiki et al. 1998). A single nucleotide polymorphism in the slice site of the leader intron of *Wx* in the amplified fragment explained 80% of the variation in apparent amylose content in a diverse sample of rice varieties (Ayres et al. 1997). A microsatellite marker closely linked to *wx*, containing a variable number of CT repeats, is closely associated with amylose content and grain quality. Larkin and Park (2003) found that two SNPs in exons 6 and 10 were associated with differences in apparent amylose content and viscosity properties of rice cultivars.

Gelatinization temperature (GT) is another important quality predictor, with intermediate GT being a preferred objective for many consumers. Waters et al. (2006) found two polymorphisms in the starch synthase II (SSIIa) gene (*Alk* on chromosome 6), that explain differences in GT. Umemoto and Aoki (2005) found SNPs in this gene that affected activity and starch granule association of SSIIa. In QTL studies of quality parameters, the *Wx* locus appears to have a larger effect, but a QTL at *Alk* is also observed (Fan et al. 2005). In a QTL analysis by Ge et al. (2005) QTLs associated with the *Wx* locus affected water absorption of cooked rice.

Fragrance is an important trait for many breeding programs, and a major gene on chromosome 8 (betaine aldehyde dehydrogenase 2, *BAD2*) has been positionally cloned (Bradbury et al. 2005a), with fragrance representing a loss of function caused by an 8-bp deletion. A perfect marker has been developed from this polymorphism for MAS (Bradbury et al. 2005b). Cooked kernel elongation is a desired trait of some consumers, especially for Basmati rices, and a gene has been mapped to chromosome 8 (Ahn et al. 1993).

Grain size, shape and appearance are very important quality factors, with most consumers preferring specific size and shape characteristics. These factors are generally under the control of QTLs of small effect, although a few with higher effect have been observed. For example, a QTL on chromosome 3 for grain length and weight has been noted in several studies (Li et al. 2004; Ge et al. 2005) and a gene underlying a QTL in the same region was recently cloned (Fan et al. 2006). Grain chalkiness is nearly always considered an undesirable trait; again its genetic control is complicated. But a QTL on chromosome 8 was found to be important in chalkiness and some other quality traits (Wan et al. 2005). While the genes for starch properties seem to be excellent candidates for MAS, the more complicated genetic control of appearance traits, and the fact that they are more easily assessed visually, has meant that MAS is not as useful.

3.2.2. Nutritional properties

As of this writing there are few nutritional traits that have been mapped in rice. QTLs for cadmium concentration in brown rice when grown in Cd-affected soil, have been identified (Ishikawa et al. 2005). This has been proposed as a means of developing cultivars with tolerance to soils high in Cd. Low phytic acid mutants have been studied because of their potential for improved quality animal feed, as well as the potential for low phytic acid cultivars to improve uptake of Fe in micronutrient deficient populations. One low phytic acid mutant *Lpa1* was mapped to a 47-kb region on chromosome 5 (Andaya and Tai 2005).

Iron deficiency anemia is a major health problem of rice consumers in Asia, and cultivar differences have been observed (Haas et al. 2005). One of the complications of this trait is that differences are much more pronounced in brown rice than in milled rice, which is the major form consumed. Identification of QTLs controlling Fe and Zn content is a high priority and would greatly aid in developing cultivars with higher micronutrient density.

3.3. Resistance to Abiotic Stresses

3.3.1. Drought tolerance

Tolerance of abiotic stresses are a primary target for MAS, because of the difficulty in screening for them, and also because a number of QTLs with fairly large effect have been found (Mackill 2006). Among the abiotic stresses, drought is by far the most widely prevalent in rice, and it is also one of the most complicated genetically and physiologically. There has been considerable effort to map QTLs

for drought tolerance and associated traits, including root growth. In general, these have failed to identify QTLs with major effects. Some of the early work on mapping for drought tolerance component traits has been reviewed previously (Nguyen et al. 1997; Mackill et al. 1999; Price 2002). While a number of QTLs have been identified for component traits, especially root depth and thickness, these traits have not been easy to associate with drought tolerance as measured by grain yield under stress.

Zou et al. (2005) identified QTLs in a RIL population from the cross of intolerant Zhenshan 97B with IRAT109, an upland drought tolerant rice. Two QTLs for grain yield under drought were detected on chromosome 2, with LOD scores around 4.2. Yue et al. (2005) studied the same population under two soil types: sandy soil and paddy (flooded) soil. They used Drought Response Index (DRI, Pantuwan et al. 2002), that can remove the effects of differences in genetic yield potential and flowering time, as well as relative yield (RY) to estimate drought tolerance. QTLs for DRI and RY mostly had small effect, and were not strongly related to root traits. One QTL on chromosome 10 contributed 32% of the phenotypic variance for RY. Root traits played a more important role under the sandy soil conditions. The same population was studied for stress at the reproductive stage in tubes (Yue et al. 2006). QTLs on chromosomes 2 and 9 from IRAT109 consistently conferred an advantage for RY. The QTL on chromosome 9 was also associated with higher spikelet fertility, which was a good indicator of drought tolerance.

At IRRI, a backcrossing scheme using multiple donor lines of diverse origin was employed to generate populations with multiple introgressions in diverse backgrounds (Li et al. 2005b; Lafitte et al. 2006). Segregating generations were selected under severe drought stress. In progeny selected for tolerance, a number of chromosomal segments were associated with improved tolerance, as well as other traits. A number of chromosomal regions selected appeared to be similar to QTLs detected in previous studies (Li et al. 2005a). In a study using Teqing as the *indica* recurrent parent and Lemont as the tropical *japonica* donor, a number of alleles from Lemont imparted improved tolerance in random introgression lines (Xu et al. 2005a).

Most drought tolerance studies indicate that there are multiple QTLs with small effect on the trait. There has thus been an attempt to identify genes that are differentially expressed between tolerant and susceptible types, and correlate these with consistently mapped QTLs. For example, Gorantla et al. (2005) observed 13 drought response genes close to known QTLs. If the candidate genes underlying QTLs can be identified this would make selection for multiple QTLs more efficient in breeding populations. Nevertheless, a search for QTLs of larger effect has continued in order to find those that would impart a higher level of tolerance when transferred into improved lines. Some recent results indicate that selecting parents with a higher level of drought tolerance shows promise in identifying more effective QTLs (G. Atlin, personal communication). Even with QTLs of moderate effect, the benefits would be substantial if they could be transferred into high-yielding and widely accepted cultivars.

3.3.2. Salinity tolerance and other soil nutrient stresses

Tolerance to salinity is desirable at both the vegetative and reproductive stage, and tolerance at the two stages does not seem to be correlated. At the vegetative stage, large differences have been observed in tolerance levels. A major QTL named *Saltol* controlling 64–80% of the phenotypic variability from the tolerant cultivar Pokkali was identified on chromosome 1 (Bonilla et al. 2002), and the QTL seems to be present in other varieties (Takehisa et al. 2004). This QTL has been fine-mapped and is being used in MAB strategies to improve the salinity tolerance of important cultivars (G. Gregorio, personal communication). A gene from the tolerant cultivar Nona Bokra at a similar location was recently identified through positional cloning (Ren et al. 2005). This selective transporter was named *SKC1* and was involved in maintaining K+ homeostasis. Mapping of tolerance at the reproductive stage is still underway.

A number of fairly strong QTLs have been mapped for other soil-related stresses, and these show great promise for MAS (Wissuwa 2005; Mackill 2006). QTLs for tolerance to Fe toxicity (Wan et al. 2003; Shimizu et al. 2005), P deficiency (Ni et al. 1998; Wissuwa et al. 1998), and Al toxicity (Nguyen et al. 2003) are among the more promising. Among these, the P deficiency QTL has been fine-mapped (Wissuwa et al. 2002) and is being transferred in a MAS program. QTLs for ability for rice to tolerate low N levels have been identified, and the trait appears to be quite complex (Lian et al. 2005). However there has been considerable interest in this trait because of environmental concerns from over-fertilization and high cost of N fertilizer for many rice farmers.

3.3.3. Submergence tolerance

Tolerance to excess water includes tolerance of complete submergence and tolerance to gradually rising water levels that stagnate for one or more months. In the latter situation, deepwater rices are suitable where water levels go beyond 50 cm. Under these situations, rapid internode elongation is usually needed. A major QTL was mapped on chromosome 12 for early elongation ability in deepwater rice (Tang et al. 2005). However, submergence tolerance is much more useful, being appropriate for the large areas of rainfed lowland rice where short-term flooding is common.

Submergence tolerant cultivars can survive 2 weeks or more of complete submergence, whereas most cultivars die within a week. The most tolerant cultivars (e.g. FR13A, Kurkaruppan, Goda Heenati) are from Orissa, India, or Sri Lanka, and their tolerance is controlled at the *Sub1* locus on chromosome 9 (Xu et al. 2006). Three ethylene-response-factor (ERF)-like genes at this locus have been identified, two of which are induced by submergence. The *Sub1A* gene conferred tolerance when overexpressed in a variety that lacked the gene. This gene seems to be only present in *indica* cultivars. SNP markers are available from the sequence that can be used for MAS. Some other QTLs have been identified in FR13A that confer additional levels of submergence tolerance (Nandi et al. 1997), but the *Sub1* locus accounts for about 70% of the phenotypic variation for survival under submergence.

Tolerance to submergence at the germination stage (anaerobic germination) seems to be independent of tolerance during vegetative stage growth. Tolerance at germination would be desirable both for protecting germinating seeds from unexpected flooding caused by heavy rainfall and for strategies of deliberate submergence to control weeds in direct-seeded rice. A QTL study measured shoot length under water and found five QTLs, mostly of relatively small effect (Jiang et al. 2004). Studies with varieties with a higher level of tolerance are underway and may yield QTLs of larger effects.

3.3.4. Low-temperature tolerance

Cold tolerance at the booting stage (about 12 days before heading) is an important breeding objective in temperate areas or in the high-elevation tropics. *Japonica* rice cultivars are more tolerant than *indicas*. Takeuchi et al. (2001) identified three QTLs in a cross between two *japonicas*; one of the QTLs on chromosome 7 contributed 22% to the phenotypic variation. Dai et al. (2004) also detected a strong QTL on chromosome 7 from a Yunnan land race. Another QTL on chromosome 4 was detected by introgressing the chromosomal fragment from the tolerant tropical *japonica* Silewah into the *japonica* variety Hokkai241. The gene *Ctb1* has been fined mapped and candidates identified (Saito et al. 2004). Andaya and Mackill (2003b) identified eight QTLs in a RIL population between a *japonica* cold tolerant variety M-202 and an *indica* susceptible variety IR50. The two strongest QTLs were on chromosomes 2 and 3.

At the vegetative stage low temperature causes chlorosis and even death of the plants. This is a problem in some temperate areas and in the subtropics where rice is grown in the winter season as a double crop. A strong QTL for cold-induced necrosis and wilting tolerance has been identified on chromosome 12 (Andaya and Mackill 2003a). In another *indica* × *japonica* cross (Jileng 1/Milyang 23), Han et al. (2004) detected different QTLs for leaf discoloration under 12°C water temperature, and they did not detect the same strong QTL on chromosome 12. They found a larger QTL on chromosome 1.

3.4. Resistance to Biotic Stresses

3.4.1. Disease resistance

Blast (*Magneporthe grisea*) is the most widespread fungal disease of rice and has attracted major efforts to map both major gene and quantitative (partial) resistance. For most breeding objectives in the tropics, quantitative resistance seems to be the best option, because of the risk of breakdown of major gene resistance. Partial resistance is controlled by multiple QTLs of relatively small effect (Wisser et al. 2005; Wu et al. 2005). In practical breeding programs for tropical lowland rice, it has not been too difficult to develop varieties with sufficient partial resistance, but the small effects of most QTLs has made it difficult to apply MAS. Association of known PR genes with QTLs is seen as a means of finding more useful markers for selecting for partial resistance to blast (Liu et al. 2004).

At least 40 major genes have been mapped, and candidates have been found for some (Chen et al. 2005). Characterizing the individual R genes for their race specificity has provided information for pyramiding combinations of genes to obtain more broad spectrum resistance, which is thought to be more durable. Fine-mapping and identification of good PCR-based markers for the important resistance genes has provided useful tools for MAS (Fjellstrom et al. 2006).

Bacterial blight (*Xanthomonas oryzae*) has also been well studied, and at least 30 genes are known. Compared to blast, major gene resistance has been more durable for bacterial blight. A number of resistance genes have been identified by map-based cloning, beginning with the well-known *Xa21* gene (Song et al. 1995). These genes have been the focus of efforts for fine-scale mapping and MAS. Some recent examples include *Xa2* (He et al. 2006), *xa13* (Chu et al. 2006) and *Xa27* (Gu et al. 2005).

Tungro is the most serious virus disease, being caused by two viruses, spherical and bacilliform. A resistance gene for the spherical virus was mapped on chromosome 4 (Sebastian et al. 1996). Four resistance genes are reported in the Gramene database, but they have not all been documented in refereed journal articles. Rice Yellow Mottle Virus is an important disease in Africa. Two QTLs on chromosomes 7 and 12 showing complementary epistasis have been mapped and used in MAS (Ahmadi et al. 2001).

3.4.2. Insect resistance

Major attention has been given to the brown planthopper (BPH), which can cause devastating crop losses under conducive conditions. At least 18 genes have been mapped and/or fine-mapped (Jena et al. 2006). Quantitative trait loci have also been mapped (Sun et al. 2005) with the objective of attaining resistance that would not break down with evolution of new biotypes. Resistance to green leaf hopper (GLH) has also received some attention, mostly as a vector of tungro virus (Sebastian et al. 1996; Wang et al. 2003a). Green rice leaf hopper is a pest in temperate areas, and at least six resistance genes are known. CAPS markers for MAS for resistance gene *Grh3* were developed (Saka et al. 2005).

Stem borers are one of the most important insect pests of rice, and no strong sources of resistance have been identified. Efforts for developing resistant varieties is focused more on transgenics with Bt toxin genes.

4. MAS EXAMPLES IN RICE

Examples of the use of MAS in rice breeding programs have been limited as reported in the published literature. Reasons for this scarcity are numerous and include

- Cost of applying genotyping methods, especially in large breeding populations
- Limited number of major genes or QTLs that will be beneficial in local germplasm

- Lack of traits that have been fine-mapped with good selectable markers
- Novelty of methods and long-term nature of MAS projects: most of them are still ongoing

MAS activities are often not published in journals because they are perceived as applied breeding activities and in many cases documentation may be insufficient for full data analysis. This is unfortunate, because there is a need to properly document these activities to provide rice scientists with up-to-date information on how MAS can be applied in practice. One can get a better idea of the current extent of activities in MAS by browsing the abstracts presented at the 2005 International Rice Genetics Symposium, available at http://www.irri.org/rg5/.

The prerequisites to apply effective MAS can be considered generally as: mapping the gene or major QTLs for the trait, fine-scale mapping, validation of the effect of the gene/QTL in appropriate genetic background, and development of diagnostic markers for the target gene (Collard and Mackill 2007). The current applications of MAS in rice breeding can be considered under three general categories: gene pyramiding, marker assisted backcrossing, and use of markers in standard breeding populations. Molecular markers are also useful for evaluation of breeding material and cultivars (Collard and Mackill 2007), but these are not covered in the present article.

4.1. Gene Pyramiding

Pyramiding genes for resistance to diseases has been the most common application of MAS in rice. The major benefit obtained from the use of markers in gene pyramiding is the ability to detect plants with multiple genes whose effects are difficult to separate phenotypically. Collard and Mackill (2007) documented five reports in the literature, mostly for blast and bacterial blight, and many more are in progress (for examples see: http://www.irri.org/rg5/Abstracts.pdf). The chief advantage of these pyramids is that they combine the resistances for many races of the pathogen achieved from the different spectra of the component resistance genes. In blast, a number of gene combinations are thought to give a very strong level of resistance, especially combinations of *Pi1* and *Pi2*. Markers specific for the *Pi2*/*Piz*/*Pi9* gene complex on chromosome 6 are likely to be useful for breeding programs seeking higher levels of blast resistance (Fjellstrom et al. 2006).

Pyramiding major genes for bacterial blight should be valuable, because the durability of resistance genes is relatively longer, and some R genes have been very effective. *Xa4*, *xa5*, *xa13*, and *Xa21* have been commonly used (Huang et al. 1997; Joseph et al. 2004). For blast, the concept also seems to apply; however, the durability of blast resistance genes has been very low. This has led to concern that even gene pyramids could be overcome by more complex *Magnaporthe* races. At present there is not sufficient information to draw firm conclusions. In the irrigated lowland tropics, partial resistance is usually sufficient. In areas of higher blast pressure (temperate rice or the tropical uplands), major gene pyramids seem

an appropriate, although not fail-safe strategy. Narayanan et al. (2004) combined two blast resistance genes *Pi1* and *Piz*^{5} (*Pi2*) with the bacterial blight resistance transgene *Xa21*.

In many applications of gene pyramiding, the objective is actually to transfer the target genes into an existing variety, thus representing a combination of pyramiding with MAB. As described in the next section, one potential drawback to this is the compounding of linkage drag over several genes. With two to three gene pyramids this may not be a major drawback, but certainly it will be for more complex pyramids.

Gene pyramids have been proposed for other traits aside from resistance. For example, Nas et al. (2005) combined three genes controlling temperature sensitive male sterility in an attempt to derive lines with different behavior for sterility induction.

4.1.1. Marker assisted backcrossing

The MAB approach is a highly effective, yet underused strategy for applying markers in rice breeding programs. The advantages of MAB for rice breeding include:

- Lower population sizes than conventional breeding that require less genotyping
- The existence of widely-grown rice cultivars ("mega varieties") that possess the essential features of high grain quality, yield and local adaptation, but lack specific traits such as stress tolerance
- Existence of major QTLs and major genes that could add value to many elite rice cultivars

Tanksley et al (1989) pointed out the potential of using markers to precisely introduce a target gene into a desired cultivar. Theoretical considerations involved in reducing linkage drag and the contribution of background markers have been described (Frisch et al. 1999b; 1999a). Collard and Mackill (2006) divided the process of MAB into three components: use of a tightly linked marker or gene marker (perfect marker) for foreground selection, use of flanking markers for "recombinant selection", and use of background markers for selection against the donor genome at unlinked loci.

Most rice workers have employed MAS without rigorous use of background markers or flanking markers. There are several reports of combining pyramiding and MAB for bacterial blight resistance: three resistance genes in BC3F2 (Sanchez et al. 2000), three resistance genes in BC2F3 (Singh et al. 2001), and two resistance genes in BC1F3 (Joseph et al. 2004). Liu et al. (2003) transferred the blast resistance gene *Pi1* into the widely used restorer line Zhenshan 97 using background markers to remove the donor genome. Ahmadi et al. (2001) applied foreground selection for two QTLs for resistance to RYMV, in combination with background selection. After the BC3 generation, the proportion of the recipient genome was about 95% and the segment around the two QTLs was about 20 cM. Steele et al. (2006) performed a complicated assembly of multiple QTLs related to drought and fragrance. While the objective was to combine five QTLs over three backcross generations while

removing unnecessary or undesirable donor segments, only one of the four drought-related QTLs (for root length) functioned in the background of the recipient cultivar Kalinga. Toojinda et al. (2005) combined genes for submergence tolerance (*Sub1*), and resistance to BB, BPH, and blast, as well as some quality traits, into two widely grown varieties, Khao Dawk Mali 105 and RD6 over four backcrosses. The extent of donor genome still present was not mentioned.

The most systematic studies reported are those of Q. Zhang and colleagues for transferring bacterial blight resistance into widely grown cultivars. A fragment of 3.8 cM surrounding the *Xa21* gene was introduced into Minghui 63 restorer line (Chen et al. 2000) and a similar fragment was introduced into the restorer '6078' (Chen et al. 2001) within three backcross generations. The process involved selection of BC progeny possessing the target locus, selection of progeny with recombination on one side of the locus (recombinant selection), and selecting among the remaining progeny for fewer number of unlinked donor segments. Work at IRRI has shown that the same approach can be applied to QTLs that have been fine-mapped. The *Sub1* QTL for submergence tolerance was introgressed into the widely grown cultivar Swarna (Xu et al. 2006), and in the BC3 progeny the introgressed fragment was limited to 2.3 Mb of DNA (Neeraja et al., submitted).

One of the applications of MAB is the introduction of transgenes into a desirable cultivar. The transformation system is efficient for only a few cultivars and is not very easy in *indica* rice cultivars. So generation of transgenics and selection of the desirable event will usually occur in an easily transformable cultivar, and needs to be backcrossed into other cultivars. This is currently the approach being used to introduce the vitamin A trait (Golden Rice) into the mega varieties in Asia (Al-Babili and Beyer 2005; Paine et al. 2005).

One drawback to the above approach is the intensive amount of genotyping that must be done. Large backcross populations, usually on the order of 500 plants or more, must be produced to make sure that there are sufficient plants to carry out background selection after the foreground and recombinant selection have been performed. With flanking markers of 3–5 cM, The foreground/recombinant selection will limit the population to less than 5% of the total backcross progeny. If very large populations are used, the process could be completed by BC2, but the number of molecular data points required would be lower for BC3 (Frisch et al. 1999a). Despite the intensive genotyping required, the approach has much to commend it, especially considering the difficulty of developing improved varieties that can replace the existing mega varieties. These varieties have many desirable features and changes to their genetic makeup more often result in deficiencies, even when they occur on a small scale. The approach is certainly not proposed as a replacement of conventional breeding efforts for developing new varieties. On the other hand, it gives greater assurance that the intense efforts undertaken for MAB would result in improved cultivars that have the characteristics that farmers want. This is particularly important considering that most varieties released for cultivation are not ultimately adopted by a large number of farmers (Mackill 2006).

4.1.2. Use of markers in breeding nurseries

The incorporation of MAS into breeding nurseries has been problematic. In early generations, where MAS could have the greatest impact, most breeders handle hundreds of thousands, and even millions of plants. This presents a practical hurdle for DNA extraction and marker analysis. For most traits, visual or simple chemical screens are generally sufficient. More widespread use of MAS in early generations will require less expensive procedures.

In later generations, MAS can be applied for traits that are difficult to assess phenotypically. An example that appears practical at the moment is for grain quality, where a limited number of genes appear to be crucial to determining the basic properties of starch that affect texture, probably the most important quality trait. Knowing the alleles at the *waxy* and *Alk* loci for example, the breeder can obtain basic information on the desirability of breeding lines, which could normally not be obtained until advanced evaluation of fixed lines. These markers are currently being employed at IRRI in F5 to F6 lines before they are evaluated intensively in replicated yield trials.

Another example of the value of markers is in hybrid rice breeding. With the existing 3-line method of hybrid rice seed production, the male parent is required to possess restoring ability, usually controlled by two dominant genes for the widely used WA cms system. Testing for restoring ability is a laborious process that involves making test crosses to a cms line and assessing the fertility of the progeny. Having a marker system that predicted at least one of the *Rf* genes greatly reduced the amount of testing needed for this trait (M. Sattari et al., unpublished data).

For rice blast, it has been mentioned that partial resistance is probably the most effective strategy for the tropical lowlands. Many varieties have the necessary level of resistance and so it is not difficult to develop new lines that have this resistance. However, accurate assessment of partial blast resistance is not easy to obtain without repeated and careful evaluation. The genetic control of this partial resistance is complex, and the QTLs do not have large effects. For this reason, the standard approaches of MAS seem inadequate to assist breeders in their selection work. This would perhaps be different if markers that were diagnostic of partial resistance alleles were available. There is evidence that the level of partial resistance can be attributed to the number of favorable alleles of PR-related genes that are associated with the QTLs identified for partial resistance (Liu et al. 2004; Wu et al. 2004).

5. PROSPECTS

At the moment, the use of MAS in rice is occurring mostly for pyramiding resistance genes and MAB. Increasing use of these methods is expected, particularly for MAB. More widespread use of MAS is also expected with the improvement of methods for marker analysis and identification of candidate genes for economic traits. Rice gene discovery will progress rapidly. As can be seen from examples presented here, development of perfect markers is very beneficial for practical application of MAS.

Costs of DNA extraction have been reduced and it is not difficult to obtain samples for tens of thousands of plants, which should facilitate the entry of MAS into the conventional breeding process. Of particular note is the potential for lower-cost whole genome scans, which would allow more targeted assembly of genes and chromosomal regions into the predicted favorable genotypes. Low-cost array-based methods of assessing hundreds or thousands of markers will be available in the near future, making the greater use of MAS in rice inevitable.

The current focus on gene discovery should result in identification of an increasing number of genes and genetic pathways that are responsible for complex economic traits such as grain yield. One of the reasons that large population sizes are emphasized in traditional plant breeding is that the genetic control of target traits is not known, and the assembly of favorable alleles is undertaken on a trial and error basis. A more targeted allele assembly implies that population development is aimed at assembling specific alleles and fewer populations should be needed to achieve the desired result.

REFERENCES

Abdelkhalik AF, Shishido R, Nomura K, Ikehashi H (2005) QTL-based analysis of heterosis for grain shape traits and seedling characteristics in an *indica-japonica* hybrid in rice (*Oryza sativa* L.). Breed Sci 55:41–48

Ahmadi N, Albar L, Pressoir G, Pinel A, Fargette D, Ghesquiere A (2001) Genetic basis and mapping of the resistance to rice yellow mottle virus. III. Analysis of QTL efficiency in introgressed progenies confirmed the hypothesis of complementary epistasis between two resistance QTLs. Theor Appl Genet 103:1084–1092

Ahn SN, Bollich CN, McClung AM, Tanksley SD (1993) RFLP analysis of genomic regions associated with cooked-kernel elongation in rice. Theor Appl Genet 87:27–32

Al-Babili S, Beyer P (2005) Golden Rice – five years on the road – five years to go? Trends Plant Sci 10:565–573

Andaya VC, Mackill DJ (2003a) Mapping of QTLs associated with cold tolerance during the vegetative stage in rice. J Exp Bot 54:2579–2585

Andaya VC, Mackill DJ (2003b) QTLs conferring cold tolerance at the booting stage of rice using recombinant inbred lines from a *japonica* × *indica* cross. Theor Appl Genet 106:1084–1090

Andaya CB, Tai TH (2005) Fine mapping of the rice low phytic acid (*Lpa1*) locus. Theor Appl Genet 111:489–495

Andersen JR, Lubberstedt T (2003) Functional markers in plants. Trends Plant Sci 8:554–560

Ashikari M, Sakakibara H, Lin SY, Yamamoto T, Takashi T, Nishimura A, Angeles ER, Qian Q, Kitano H, Matsuoka M (2005) Cytokinin oxidase regulates rice grain production. Science 309:741–745

Ayres NM, McClung AM, Larkin PD, Bligh HFJ, Jones CA, Park WD (1997) Microsatellites and a single-nucleotide polymorphism differentiate apparent amylose classes in an extended pedigree of US rice germ plasm. Theor Appl Genet 94:773–781

Bonilla P, Dvorak J, Mackill DJ, Deal K, Gregorio G (2002) RFLP and SSLP mapping of salinity tolerance genes in chromosome 1 of rice (*Oryza sativa* L.) using recombinant inbred lines. Philipp Agric Sci 85:68–76

Bradbury LMT, Fitzgerald TL, Henry RJ, Jin QS, Waters DLE (2005a) The gene for fragrance in rice. Plant Biotechnol J 3:363–370

Bradbury LMT, Henry RJ, Jin QS, Reinke RF, Waters DLE (2005b) A perfect marker for fragrance genotyping in rice. Mol Breed 16:279–283

Cai XL, Wang ZY, Xing YY, Zhang JL, Hong MM (1998) Aberrant splicing of intron 1 leads to the heterogeneous 5′ UTR and decreased expression of *waxy* gene in rice cultivars of intermediate amylose content. Plant J 14:459–465

Chen S, Lin XH, Xu CG, Zhang QF (2000) Improvement of bacterial blight resistance of 'Minghui 63', an elite restorer line of hybrid rice, by molecular marker-assisted selection. Crop Sci 40:239–244

Chen S, Xu CG, Lin XH, Zhang Q (2001) Improving bacterial blight resistance of '6078', an elite restorer line of hybrid rice, by molecular marker-assisted selection. Plant Breed 120:133–137

Chen S, Wang L, Que ZQ, Pan RQ, Pan QH (2005) Genetic and physical mapping of *Pi37*(t), a new gene conferring resistance to rice blast in the famous cultivar St. No. 1. Theor Appl Genet 111:1563–1570

Chu ZH, Fu BY, Yang H, Xu CG, Li ZK, Sanchez A, Park YJ, Bennetzen JL, Zhang QF, Wang SP (2006) Targeting *xa13*, a recessive gene for bacterial blight resistance in rice. Theor Appl Genet 112:455–461

Collard BCY, Mackill DJ (2007) Marker-assisted selection: an approach for precision plant breeding in the 21st century. Phil Trans R Soc B Rev doi:10.1098/rstb.2007.2170

Dai LY, Lin XH, Ye CR, Ise KZ, Saito K, Kato A, Xu FR, Yu TQ, Zhang DP (2004) Identification of quantitative trait loci controlling cold tolerance at the reproductive stage in Yunnan landrace of rice, Kunmingxiaobaigu. Breed Sci 54:253–258

Dong NV, Subudhi PK, Luong PN, Quang VD, Quy TD, Zheng HG, Wang B, Nguyen HT (2000) Molecular mapping of a rice gene conditioning thermosensitive genic male sterility using AFLP, RFLP and SSR techniques. Theor Appl Genet 100:727–734

Fan CC, Yu XQ, Xing YZ, Xu CG, Luo LJ, Zhang QF (2005) The main effects, epistatic effects and environmental interactions of QTLs on the cooking and eating quality of rice in a doubled-haploid line population. Theor Appl Genet 110:1445–1452

Fan CH, Xing YZ, Mao HL, Lu TT, Han B, Xu CG, Li XH, Zhang QF (2006) GS3, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. Theor Appl Genet 112:1164–1171

Fjellstrom R, McClung AM, Shank AR (2006) SSR markers closely linked to the *Pi-z* locus are useful for selection of blast resistance in a broad array of rice germplasm. Mol Breed 17:149–157

Frisch M, Bohn M, Melchinger AE (1999a) Comparison of selection strategies for marker-assisted backcrossing of a gene. Crop Sci 39:1295–1301

Frisch M, Bohn M, Melchinger AE (1999b) Minimum sample size and optimal positioning of flanking markers in marker-assisted backcrossing for transfer of a target gene. Crop Sci 39:967–975

Gao LZ, Zhang CH, Jia JZ (2005) Cross-species transferability of rice microsatellites in its wild relatives and the potential for conservation genetic studies. Genet Resour Crop Evol 52:931–940

Ge XJ, Xing YZ, Xu CG, He YQ (2005) QTL analysis of cooked rice grain elongation, volume expansion, and water absorption using a recombinant inbred population. Plant Breed 124:121–126

Gorantla M, Babu PR, Lachagari VBR, Feltus FA, Paterson AH, Reddy AR (2005) Functional genomics of drought stress response in rice: Transcript mapping of annotated unigenes of an indica rice (*Oryza sativa* L. cv. Nagina 22). Curr Sci 89:496–514

Gu KY, Yang B, Tian DS, Wu LF, Wang DJ, Sreekala C, Yang F, Chu ZQ, Wang GL, White FF, Yin ZC (2005) *R* gene expression induced by a type-III effector triggers disease resistance in rice. Nature 435:1122–1125

Guo LB, Xing YZ, Mei HW, Xu CG, Shi CH, Wu P, Luo LJ (2005) Dissection of component QTL expression in yield formation in rice. Plant Breed 124:127–132

Haas JD, Beard JL, Murray-Kolb LE, del Mundo AM, Felix A, Gregorio GB (2005) Iron-biofortified rice improves the iron stores of nonanemic Filipino women. J Nutr 135:2823–2830

Han L-Z, Qiao Y-L, Cao G-L, Zhang Y-Y, An Y-P, Ye J-D, Koh H-J (2004) QTLs analysis of cold tolerance during early growth period for rice. Rice Sci 11:245–250

Hazen SP, Kay SA (2003) Gene arrays are not just for measuring gene expression. Trends Plant Sci 8:413–416

He Q, Li DB, Zhu YS, Tan MP, Zhang DP, Lin XH (2006) Fine mapping of *Xa2*, a bacterial blight resistance gene in rice. Mol Breed 17:1–6

Hospital F (2001) Size of donor chromosome segments around introgressed loci and reduction of linkage drag in marker-assisted backcross programs. Genetics 158:1363–1379

Hua J, Xing Y, Wu W, Xu C, Sun X, Yu S, Zhang Q (2003) Single-locus heterotic effects and dominance by dominance interactions can adequately explain the genetic basis of heterosis in an elite rice hybrid. Proc Natl Acad Sci USA 100:2574–2579

Huang N, Angeles ER, Domingo J, Magpantay G, Singh S, Zhang G, Kumaravadivel N, Bennett J, Khush GS (1997) Pyramiding of bacterial blight resistance genes in rice: marker-assisted selection using RFLP and PCR. Theor Appl Genet 95:313–320

Ikehashi H, Araki H (1986) Genetics of F_1 sterility in rice. In: Rice genetics. International rice research institute, Los Baños, Philippines, pp 119–132

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

Ishikawa S, Ae N, Yano M (2005) Chromosomal regions with quantitative trait loci controlling cadmium concentration in brown rice (*Oryza sativa*). New Phytol 168:345–350

Ishimaru K, Kashiwagi T, Hirotsu N, Madoka Y (2005) Identification and physiological analyses of a locus for rice yield potential across the genetic background. J Exp Bot 56:2745–2753

Isshiki M, Morino K, Nakajima M, Okagaki RJ, Wessler SR, Izawa T, Shimamoto K (1998) A naturally occurring functional allele of the rice waxy locus has a GT to TT mutation at the 5′ splice site of the first intron. Plant J 15:133–138

Izawa T, Takahashi Y, Yano M (2003) Comparative biology comes into bloom: genomic and genetic comparison of flowering pathways in rice and *Arabidopsis*. Curr Opin Plant Biol 6:113–120

Jena KK, Jeung JU, Lee JH, Choi HC, Brar DS (2006) High-resolution mapping of a new brown planthopper (BPH) resistance gene, *Bph18*(t), and marker-assisted selection for BPH resistance in rice (*Oryza sativa* L.). Theor Appl Genet 112:288–297

Ji Q, Lu JF, Chao Q, Gu MH, Xu ML (2005) Delimiting a rice wide-compatibility gene S^{n5} to a 50 kb region. Theor Appl Genet 111:1495–1503

Jiang L, Hou M-Y, Wang C-M, Wan J-M (2004) Quantitative trait loci and epistatic analysis of seed anoxia germinability in rice (*Oryza sativa*). Rice Sci 11:238–244

Joseph M, Gopalakrishnan S, Sharma RK, Singh VP, Singh AK, Singh NK, Mohapatra T (2004) Combining bacterial blight resistance and Basmati quality characteristics by phenotypic and molecular marker-assisted selection in rice. Mol Breed 13:377–387

Kanbar A, Janamatti M, Sudheer E, Vinod MS, Shashidhar HE (2006) Mapping QTLs underlying seedling vigour traits in rice (*Oryza sativa* L.). Curr Sci 90:24–26

Lafitte HR, Li ZK, Vijayakumar CHM, Gao YM, Shi Y, Xu JL, Fu BY, Ali AJ, Domingo J, Maghirang R, Torres R, Mackill D (2006) Improvement of rice drought tolerance through backcross breeding: evaluation of donors and selection in drought nurseries. Field Crop Res 97:77–86

Lang NT, Subudhi PK, Virmani SS, Brar DS, Khush GS, Li ZK, Huang N (1999) Development of PCR-based markers for thermosensitive genetic male sterility gene *tms3* (t) in rice (*Oryza sativa* L.). Hereditas 131:121–127

Larkin PD, Park WD (2003) Association of waxy gene single nucleotide polymorphisms with starch characteristics in rice (*Oryza sativa* L.). Mol Breed 12:335–339

Lee DS, Chen LJ, Suh HS (2005a) Genetic characterization and fine mapping of a novel thermo-sensitive genic male-sterile gene *tms6* in rice (*Oryza sativa* L.). Theor Appl Genet 111: 1271–1277

Lee SJ, Oh CS, Suh JP, McCouch SR, Ahn SN (2005b) Identification of QTLs for domestication-related and agronomic traits in an *Oryza sativa* × *O. rufipogon* BC1F7 population. Plant Breed 124:209–219

Leung H, An GL (2004) Rice functional genomics: large-scale gene discovery and applications to crop improvement. Adv Agron 82:55–111

Li Z-K, Luo LJ, Mei HW, Wang DL, Shu QY, Tabien R, Zhong DB, Ying CS, Stansel JW, Khush GS, Paterson AH (2001) Overdominant epistatic loci are the primary genetic basis of inbreeding depression and heterosis in rice. I. Biomass and grain yield. Genetics 158:1737–1753

Li JM, Thomson W, McCouch SR (2004) Fine mapping of a grain-weight quantitative trait locus in the pericentromeric region of rice chromosome 3. Genetics 168:2187–2195

Li ZC, Mu P, Li CP, Zhang HL, Li ZK, Gao YM, Wang XK (2005a) QTL mapping of root traits in a doubled haploid population from a cross between upland and lowland japonica rice in three environments. Theor Appl Genet 110:1244–1252

Li ZK, Fu BY, Gao YM, Xu JL, Ali J, Lafitte HR, Jiang YZ, Rey JD, Vijayakumar CHM, Maghirang R, Zheng TQ, Zhu LH (2005b) Genome-wide introgression lines and their use in genetic and molecular dissection of complex phenotypes in rice (*Oryza sativa* L.). Plant Mol Biol 59:33–52

Li C, Zhou A, Sang T (2006) Rice domestication by reducing shattering. Science 311:1936–1939

Lian XM, Xing YZ, Yan H, Xu CG, Li XH, Zhang QF (2005) QTLs for low nitrogen tolerance at seedling stage identified using a recombinant inbred line population derived from an elite rice hybrid. Theor Appl Genet 112:85–96

Liu S-P, Li X, Wang C-Y, Li X-H, He Y-Q (2003) Improvement of resistance to rice blast in Zhenzhan 97 by molecular marker-aided selection. Acta Bot Sinica 45:1346–1350

Liu B, Zhang SH, Zhu XY, Yang QY, Wu SZ, Mei MT, Mauleon R, Leach J, Mew T, Leung H (2004) Candidate defense genes as predictors of quantitative blast resistance in rice. Mol Plant Microbe Interact 17:1146–1152

Lopez MT, Toojinda T, Vanavichit A, Tragoonrung S (2003) Microsatellite markers flanking the *tms2* gene facilitated tropical TGMS rice line development. Crop Sci 43:2267–2271

Luo LJ, Li ZK, Mei HW, Shu QY, Tabien R, Zhong DB, Ying CS, Stansel JW, Khush GS, Paterson AH (2001) Overdominant epistatic loci are the primary genetic basis of inbreeding depression and heterosis in rice. II. Grain yield components. Genetics 158:1755–1771

Luo AD, Qian Q, Yin HF, Liu XQ, Yin CX, Lan Y, Tang JY, Tang ZS, Cao SY, Wang XJ, Xia K, Fu XD, Luo D, Chu CC (2006) EUI1, encoding a putative cytochrome P450 monooxygenase, regulates internode elongation by modulating gibberellin responses in rice. Plant Cell Physiol 47:181–191

Ma HK, Zhang SB, Ji L, Zhu HB, Yang SL, Fang XJ, Yang RC (2006) Fine mapping and in silico isolation of the EUI1 gene controlling upper internode elongation in rice. Plant Mol Biol 60:87–94

Mackill DJ (2006) Breeding for resistance to abiotic stresses in rice: the value of quantitative trait loci. In: Lamkey KR, Lee M (eds) Plant breeding: The Arnel R Hallauer international symposium. Blackwell Publishing Ames, IA, pp 201–212

Mackill DJ, McNally KL (2004) A model crop species: molecular markers in rice. In: Lörz H, Wenzel G (eds) Molecular marker systems in plant breeding and crop improvement, Vol. 55 Biotechnology in agriculture and forestry. Springer, Heidelberg, pp 39–54

Mackill DJ, Nguyen HT, Zhang J (1999) Use of molecular markers in plant improvement programs for rainfed lowland rice. Field Crop Res 64:177–185

McCouch SR, Teytelman L, Xu Y, Lobos KB, Clare K, Walton M, Fu B, Maghirang R, Li Z, Zing Y, Zhang Q, Kono I, Yano M, Fjellstrom R, DeClerck G, Schneider D, Cartinhour S, Ware D, Stein L (2002) Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.). DNA Res 9:199–207

Mei HW, Li ZK, Shu QY, Guo LB, Wang YP, Yu XQ, Ying CS, Luo LJ (2005) Gene actions of QTLs affecting several agronomic traits resolved in a recombinant inbred rice population and two backcross populations. Theor Appl Genet 110:649–659

Mei HW, Xu JL, Li ZK, Yu XQ, Guo LB, Wang YP, Ying CS, Luo LJ (2006) QTLs influencing panicle size detected in two reciprocal introgressive line (IL) populations in rice (*Oryza sativa* L.). Theor Appl Genet 112:648–656

Moncada P, Martinez CP, Borrero J, Chatel M, Gauch H, Guimaraes E, Tohme J, McCouch SR (2001) Quantitative trait loci for yield and yield components in an *Oryza sativa* × *Oryza rufipogon* BC2F2 population evaluated in an upland environment. Theor Appl Genet 102:41–52

Nandi S, Subudhi PK, Senadhira D, Manigbas NL, Sen-Mandi S, Huang N (1997) Mapping QTLs for submergence tolerance in rice by AFLP analysis and selective genotyping. Mol Gen Genet 255:1–8

Narayanan NN, Baisakh N, Oliva NP, VeraCruz CM, Gnanamanickam SS, Datta K, Datta SK (2004) Molecular breeding: marker-assisted selection combined with biolistic transformation for blast and bacterial blight resistance in *indica* rice (cv. CO39). Mol Breed 14:61–71

Nas TMS, Sanchez DL, Diaz MGQ, Mendioro MS, Virmani SS (2005) Pyramiding of thermosensitive genetic male sterility (TGMS) genes and identification of a candidate *tms5* gene in rice. Euphytica 145:67–75

Nguyen HT, Babu RC, Blum A (1997) Breeding for drought resistance in rice: physiology and molecular genetics considerations. Crop Sci 37:1426–1434

Nguyen BD, Brar DS, Bui BC, Nguyen TV, Pham LN, Nguyen HT (2003) Identification and mapping of the QTL for aluminum tolerance introgressed from the new source, *Oryza rufipogon* Griff., into indica rice (*Oryza sativa* L.). Theor Appl Genet 106:583–593

Ni JJ, Wu P, Senadhira D, Huang N (1998) Mapping QTLs for phosphorus deficiency tolerance in rice (*Oryza sativa* L.). Theor Appl Genet 97:1361–1369

Paine JA, Shipton CA, Chaggar S, Howells RM, Kennedy MJ, Vernon G, Wright SY, Hinchliffe E, Adams JL, Silverstone AL, Drake R (2005) Improving the nutritional value of Golden Rice through increased pro-vitamin A content. Nat Biotechnol 23:482–487

Pantuwan G, Fukai S, Cooper M, Rajatasereekul S, O'Toole JC (2002) Yield response of rice (*Oryza sativa* L.) genotypes to different types of drought under rainfed lowlands – Part 2. Selection of drought resistant genotypes. Field Crop Res 73:169–180

Peng S, Laza RC, Visperas RM, Sanico AL, Cassman KG, Khush GS (2000) Grain yield of rice cultivars and lines developed in the Philippines since 1966. Crop Sci 40:307–314

Peng S, Laza RC, Visperas RM, Khush GS, Virk PS (2004) Rice: progress in breaking the yield ceiling. Paper presented at proceedings of the 4th international crop science congress. Brisbane, Australia, 26 September–1 October 2004

Price A (2002) QTLs for root growth and drought resistance in rice. In: Molecular techniques in crop improvement. Kluwer Academic Publisher, Dordrecht, pp 563–584

Qiu SQ, Liu KD, Jiang JX, Song X, Xu CG, Li XH, Zhang QF (2005) Delimitation of the rice wide compatibility gene *S5*(n) to a 40-kb DNA fragment. Theor Appl Genet 111:1080–1086

Ren ZH, Gao JP, Li LG, Cai XL, Huang W, Chao DY, Zhu MZ, Wang ZY, Luan S, Lin HX (2005) A rice quantitative trait locus for salt tolerance encodes a sodium transporter. Nat Genet 37:1141–1146

Saito K, Hayano-Saito Y, Maruyama-Funatsuki W, Sato Y, Kato A (2004) Physical mapping and putative candidate gene identification of a quantitative trait locus *Ctb1* for cold tolerance at the booting stage of rice. Theor Appl Genet 109:515–522

Saka N, Tsuji T, Toyama T, Yano M, Izawa T, Sasaki T (2005) Development of cleaved amplified polymorphic sequence (CAPS) markers linked to a green rice leafhopper resistance gene, *Grh3*(t). Plant Breed 125:140–143

Sanchez AC, Brar DS, Huang N, Li Z, Khush GS (2000) Sequence tagged site marker-assisted selection for three bacterial blight resistance genes in rice. Crop Sci 40:792–797

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

Sebastian LS, Ikeda R, Huang N, Imbe T, Coffman WR, McCouch SR (1996) Molecular mapping of resistance to rice tungro spherical virus and green leafhopper. Phytopathology 86:25–30

Septiningsih EM, Prasetiyono J, Lubis E, Tai TH, Tjubaryat T, Moeljopawiro S, McCouch SR (2003) Identification of quantitative trait loci for yield and yield components in an advanced backcross population derived from the *Oryza sativa* variety IR64 and the wild relative *O. rufipogon*. Theor Appl Genet 107:1419–1432

Shimizu A, Guerta CQ, Gregorio GB, Kawasaki S, Ikehashi H (2005) QTLs for nutritional contents of rice seedlings (*Oryza sativa* L.) in solution cultures and its implication to tolerance to iron-toxicity. Plant Soil 275:57–66

Singh S, Sidhu JS, Huang N, Vikal Y, Li Z, Brar DS, Dhaliwal HS, Khush GS (2001) Pyramiding three bacterial blight resistance genes (*xa5*, *xa13* and *Xa21*) using marker-assisted selection into indica rice cultivar PR106. Theor Appl Genet 102:1011–1015

Song WY, Wang GL, Chen LL, Kim HS, Pi LY, Holsten T, Gardner J, Wang B, Zhai WX, Zhu LH, Fauquet C, Ronald P (1995) A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa21*. Science 270:1804–1806

Steele KA, Price AH, Shashidhar HE, Witcombe JR (2006) Marker-assisted selection to introgress rice QTLs controlling root traits into an Indian upland rice variety. Theor Appl Genet 112:208–221

Sun LH, Su CC, Wang CM, Zhai HQ, Wan JM (2005) Mapping of a major resistance gene to the brown planthopper in the rice cultivar Rathu Heenati. Breed Sci 55:391–396

Takai T, Fukuta Y, Shiraiwa T, Horie T (2005) Time-related mapping of quantitative trait loci controlling grain-filling in rice (*Oryza sativa* L.). J Exp Bot 56:2107–2118

Takehisa H, Shimodate T, Fukuta Y, Ueda T, Yano M, Yamaya T, Kameya T, Sato T (2004) Identification of quantitative trait loci for plant growth of rice in paddy field flooded with salt water. Field Crop Res 89:85–95

Takeuchi Y, Hayasaka H, Chiba B, Tanaka I, Shimano T, Yamagishi M, Nagano K, Sasaki T, Yano M (2001) Mapping quantitative trait loci controlling cool-temperature tolerance at booting stage in temperate japonica rice. Breed Sci 51:191–197

Tang DQ, Kasai Y, Miyamoto N, Ukai Y, Nemoto K (2005) Comparison of QTLs for early elongation ability between two floating rice cultivars with a different phylogenetic origin. Breed Sci 55:1–5

Tanksley SD, Nelson JC (1996) Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. Theor Appl Genet 92:191–203

Tanksley SD, Young ND, Paterson AH, Bonierbale MW (1989) RFLP mapping in plant breeding: new tools for an old science. Bio Technol 7:257–264

Thomson MJ, Tai TH, McClung AM, Lai XH, Hinga ME, Lobos KB, Xu Y, Martinez CP, McCouch SR (2003) Mapping quantitative trait loci for yield, yield components and morphological traits in an advanced backcross population between *Oryza rufipogon* and the *Oryza sativa* cultivar Jefferson. Theor Appl Genet 107:479–493

Tian F, Li DJ, Fu Q, Zhu ZF, Fu YC, Wang XK, Sun CQ (2006) Construction of introgression lines carrying wild rice (*Oryza rufipogon* Griff.) segments in cultivated rice (*Oryza sativa* L.) background and characterization of introgressed segments associated with yield-related traits. Theor Appl Genet 112:570–580

Toojinda T, Tragoonrung S, Vanavichit A, Siangliw JL, Pa-In N, Jantaboon J, Siangliw M, Fukai S (2005) Molecular breeding for rainfed lowland rice in the Mekong region. Plant Prod Sci 8:330–333

Umemoto T, Aoki N (2005) Single-nucleotide polymorphisms in rice starch synthase IIa that alter starch gelatinisation and starch association of the enzyme. Funct Plant Biol 32:763–768

Wan JL, Zhai HQ, Wan JM, Ikehashi H (2003) Detection and analysis of QTLs for ferrous iron toxicity tolerance in rice, *Oryza sativa* L. Euphytica 131:201–206

Wan XY, Wan JM, Weng JF, Jiang L, Bi JC, Wang CM, Zhai HQ (2005) Stability of QTLs for rice grain dimension and endosperm chalkiness characteristics across eight environments. Theor Appl Genet 110:1334–1346

Wan JM, Jiang L, Tang JY, Wang CM, Hou MY, Jing W, Zhang L (2006) Genetic dissection of the seed dormancy trait in cultivated rice (*Oryza sativa* L.). Plant Sci 170:786–792

Wang ZY, Zheng FQ, Shen GZ, Gao JP, Snustad DP, Li MG, Zhang JL, Hong MM (1995) The amylose content in rice endosperm is related to the post-transcriptional regulation of the *waxy* gene. Plant J 7:613–622

Wang CM, Yasui H, Yoshimura A, Zhai HQ, Wan JM (2003a) Inheritance and QTL mapping of antibiosis to green leafhopper in rice. Crop Sci 44:389–393

Wang YG, Xing QH, Deng QY, Liang FS, Yuan LP, Weng ML, Wang B (2003b) Fine mapping of the rice thermo-sensitive genic male-sterile gene *tms5*. Theor Appl Genet 107:917–921

Wang CH, Zhang P, Ma ZR, Zhang MY, Sun GC, Ling DH (2004) Development of a genetic marker linked to a new thermo-sensitive male sterile gene in rice (*Oryza sativa* L.). Euphytica 140:217–222

Wang GW, He YQ, Xu CG, Zhang QF (2005) Identification and confirmation of three neutral alleles conferring wide compatibility in inter-subspecific hybrids of rice (*Oryza sativa* L.) using near-isogenic lines. Theor Appl Genet 111:702–710

Wang ZH, Zou YJ, Li XY, Zhang QY, Chen L, Wu H, Su DH, Chen YL, Guo JX, Luo D, Long YM, Zhong Y, Liu YG (2006) Cytoplasmic male sterility of rice with boro II cytoplasm is caused by a cytotoxic peptide and is restored by two related PPR motif genes via distinct modes of mRNA silencing. Plant Cell 18:676–687

Waters DLE, Henry RJ, Reinke RF, Fitzgerald MA (2006) Gelatinization temperature of rice explained by polymorphisms in starch synthase. Plant Biotechnol J 4:115–122

Wisser RJ, Sun Q, Hulbert SH, Kresovich S, Nelson RJ (2005) Identification and characterization of regions of the rice genome associated with broad-spectrum, quantitative disease resistance. Genetics 169:2277–2293

Wissuwa M (2005) Mapping nutritional traits in crop plants. In: Broadley MR, White PJ (eds) Plant nutritional genomics. Blackwell Publishing, Ames, IA pp 220–241

Wissuwa M, Yano M, Ae N (1998) Mapping of QTLs for phosphorus-deficiency tolerance in rice (*Oryza sativa* L.). Theor Appl Genet 97:777–783

Wissuwa M, Wegner J, Ae N, Yano M (2002) Substitution mapping of *Pup1*: a major QTL increasing phosphorus uptake of rice from a phosphorus-deficient soil. Theor Appl Genet 105:890–897

Wu JL, Sinha PK, Variar M, Zheng KL, Leach JE, Courtois B, Leung H (2004) Association between molecular markers and blast resistance in an advanced backcross population of rice. Theor Appl Genet 108:1024–1032

Wu JL, Fan YY, Li DB, Zheng KL, Leung H, Zhuang JY (2005) Genetic control of rice blast resistance in the durably resistant cultivar Gumei 2 against multiple isolates. Theor Appl Genet 111:50–56

Xiao JH, Li JM, Grandillo S, Ahn SN, Yuan LP, Tanksley SD, McCouch SR (1998) Identification of trait-improving quantitative trait loci alleles from a wild rice relative, *Oryza rufipogon*. Genetics 150:899–909

Xu JL, Lafitte HR, Gao YM, Fu BY, Torres R, Li ZK (2005a) QTLs for drought escape and tolerance identified in a set of random introgression lines of rice. Theor Appl Genet 111:1642–1650

Xu X, Kawasaki S, Fujimura T, Wang CT (2005b) A protocol for high-throughput extraction of DNA from rice leaves. Plant Mol Biol Rep 23:291–295

Xu K, Xia X, Fukao T, Canlas P, Maghirang-Rodriguez R, Heuer S, Ismail AI, Bailey-Serres J, Ronald PC, Mackill DJ (2006) *Sub1A* is an ethylene response factor-like gene that confers submergence tolerance to rice. Nature 442:705–708

Yamaguchi T, Lee DY, Miyao A, Hirochika H, An GH, Hirano HY (2006) Functional diversification of the two C-class MADS box genes *OSMADS3* and *OSMADS58* in *Oryza sativa*. Plant Cell 18:15–28

Yao FY, Xu CG, Yu SB, Li JX, Gao YJ, Li XH, Zhang QF (1997) Mapping and genetic analysis of two fertility restorer loci in the wild-abortive cytoplasmic male sterility system of rice (*Oryza sativa* L.). Euphytica 98:183–187

You AQ, Lu XG, Jin HJ, Ren XA, Liu K, Yang GC, Yang HY, Zhu LL, He GC (2006) Identification of quantitative trait loci across recombinant inbred lines and testcross populations for traits of agronomic importance in rice. Genetics 172:1287–1300

Yu J, Hu SN, Wang J, Wong GKS, Li SG, Liu B, Deng YJ et al (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp *indica*). Science 296:79–92

Yue B, Xiong LZ, Xue WY, Xing YZ, Luo LJ, Xu CG (2005) Genetic analysis for drought resistance of rice at reproductive stage in field with different types of soil. Theor Appl Genet 111:1127–1136

Yue B, Xue WY, Xiong LZ, Yu XQ, Luo LJ, Cui KH, Jin DM, Xing YZ, Zhang QF (2006) Genetic basis of drought resistance at reproductive stage in rice: separation of drought tolerance from drought avoidance. Genetics 172:1213–1228

Zhang G, Bharaj TS, Lu Y, Virmani SS, Huang N (1997) Mapping of the *Rf-3* nuclear fertility-restoring gene for WA cytoplasmic male sterility in rice using RAPD and RFLP markers. Theor Appl Genet 94:27–33

Zhu YY, Nomura T, Xu YH, Zhang YY, Peng Y, Mao BZ, Hanada A, Zhou HC, Wang RX, Li PJ, Zhu XD, Mander LN, Kamiya Y, Yamaguchi S, He ZH (2006) *ELONGATED UPPERMOST INTERNODE* encodes a cytochrome P450 monooxygenase that epoxidizes gibberellins in a novel deactivation reaction in rice. Plant Cell 18:442–456

Zou GH, Mei HW, Liu HY, Liu GL, Hu SP, Yu XQ, Li MS, Wu JH, Luo LJ (2005) Grain yield responses to moisture regimes in a rice population: association among traits and genetic markers. Theor Appl Genet 112:106–113

CHAPTER 8

APPLICATION OF GENOMICS FOR MOLECULAR BREEDING IN RICE

NAGENDRA K. SINGH* AND TRILOCHAN MOHAPATRA

National Research Centre on Plant Biotechnology (NRCPB), Indian Agricultural Research Institute (IARI), New Delhi-110 012, India

Abstract: Rice is arguably the most important food crop of the world and due to its small genome size compared to other major cereals, rice was selected as model crop species for decoding of its full genome. The international rice genome sequencing project (IRGSP), a consortium of laboratories from ten different countries, has generated a very high quality map based sequence of the 12 chromosomes of *japonica* rice cultivar 'Nipponbare' and made it available in the public domain. A whole genome draft sequence of *indica* type rice variety '93-11' has also been reported by the Beijing Genomics Institute. Annotation and comparative analysis of these and other partial genomic sequences has provided a wealth of information to the rice geneticists and breeders. Simple Sequence Repeat (SSR) markers are now easily available for any region of the rice genome. SSR markers have also been derived from the expressed sequence tags (ESTs) and unigene sequences, which correspond to the expressed component of the genome and thus have greater potential in comparative genome analysis. Furthermore, millions of single nucleotide polymorphism (SNP) and insertion-deletion (InDel) markers have already been identified in rice. Saturation of the genome with such sequence based SSR and SNP markers is accelerating fine mapping and map-based cloning of genes, and thus, development of gene-based allele-specific markers. Rice improvement programs are expected to benefit greatly from the use of these markers in near future.

1. INTRODUCTION

The genetic information in plants is stored in three sub-cellular compartments namely nucleus, chloroplast and mitochondria. The nuclear genome is much larger and far more complex in both structure and function than the two organelle genomes.

*Author for correspondence: nksingh@nrcpb.org

R.K. Varshney and R. Tuberosa (eds.), Genomics Assisted Crop Improvement: Vol. 2: Genomics Applications in Crops, 169–185.

Beginning from the second decade of the last century, conscious efforts have been made to understand structure and function of the plant nuclear genome employing tools and techniques of Mendelian genetics and cytology. As a result, genes with visible phenotypic effects could be arranged in linear order in different linkage groups based on genetic recombination. The chromosome numbers of most of the important plant species were determined. Chromosome pairing behavior in inter-species hybrids was used to identify possible diploid progenitors of many of the present day allopolyploid crop species. Discovery of DNA as genetic material and elucidation of its double helical structure during 1950s gave a new dimension to the study of genomes. Development of molecular techniques such as DNA cloning and sequencing that followed these discoveries, paved the ways for a detailed analysis of genome structure and function. Automation in laboratory instrumentation facilitating high throughput sample analysis coupled with an array of robust computational methods enabled complete sequencing of several large eukaryotic genomes including that of human, *Drosophila*, *Arabidopsis* and rice. The availability of high quality whole genome sequence provided a thorough understanding of the genome structure and evolution patterns. Novel means of understanding the genome functions could be designed by utilizing the DNA sequence information. These developments led to the birth of a new discipline in biology aptly named as "Genomics".

The widely grown cultivated rice (*Oryza sativa* L.) behaves like a true diploid with gametic chromosome number 12. Rice genome was initially characterized by cytological analysis at mitotic metaphase and chromosome pairing at pachytene stage of meiosis. Kuwada (1910) first reported the chromosome number of cultivated rice as 2n=24, using both mitotic and meiotic cells. Shastry et al. (1960) thoroughly examined pachytene chromosomes of rice and numbered the chromosomes in decreasing order of their length. Morinaga (1937, 1939) first analysed the genomes of different rice species and proposed five different genomes from A to E. Based on several earlier linkage mapping studies using morphological markers, Nagao and Takahashi (1963) proposed 12 possible linkage groups corresponding to the haploid chromosome complement of rice. Cytogenetic stocks such as trisomics and reciprocal translocations were used to establish relationship between these linkage groups and the 12 chromosomes of rice (Iwata 1986). With the construction of molecular marker based high-density linkage maps, use of mapped markers to integrate genetic and physical maps and whole genome sequencing, a wealth of information is now available for the rice researchers, which can be very effectively utilized for various applications in rice genetics and breeding. The main objective of this short review is to highlight the application of genomics for molecular breeding in rice.

2. EARLY PROGRESS IN RICE MOLECULAR GENETICS

The first attempt to characterize the rice genome using molecular markers was made in the year 1988. McCouch et al. (1988) cloned single/low copy genomic sequences and used those as probes in southern hybridization to generate the first

set of restriction fragment length polymorphism (RFLP) markers. Based on segregation of these markers in a small F_2 population of 50 individuals derived from *javanica* x *indica* cross, the molecular map of the rice genome was constructed that contained 135 RFLP markers and covered a total length of 1389 cM. Rice genome had approximately 50% single copy sequences and was less C-methylated than maize and tomato genomes. This work greatly accelerated rice genetics research worldwide. Many genes of agronomic importance for both qualitative and quantitative traits could be mapped and localized into marker intervals using this map. Following this landmark development, several RFLP maps of rice genome were reported. High-density linkage maps were subsequently made using larger sets of RFLP markers by two groups independently; one at the Cornrell University, USA (Causse et al. 1994) and the other at the Rice Genome Research Program, Japan (Kurata et al. 1994; Harushima et al. 1998). The Cornell group extended the first RFLP map by McCouch et al. (1988), while the RGP group used a different set of random genomic clones and cDNA clones prepared from callus and roots as probes in RFLP to develop the map. Kurata et al. (1994) mapped 1383 DNA markers using 186 F_2 plants derived from a cross between a *japonica* variety, Nipponbare, and an *indica* variety, Kasalath. Additional RFLP markers were developed by using cDNA clones from green and etiolated shoots and 934 new markers were added to the map using the same F_2 population after omitting 42 markers for various reasons. This gave a total of 2275 markers mapped to 1174 discrete positions covering 1521.6 cM of the rice genome (Harushima et al. 1998). Subsequently, the Japanese group further enhanced the map density (http://rgp.dna.affrc.go.jp/publicdata/ geneticmap2000/index.html) using 3267 DNA markers and almost all the RFLP markers were converted to Cleaved Amplified Polymorphic Sequence (CAPS) markers. These maps have been of great value in both genetic and physical mapping of genes and map-based sequencing of the rice genome.

RFLP markers are not very convenient to use in high throughput genotyping applications. Plant breeding requires a faster, simpler and more easy-to-use method of analysis. Therefore, while the RFLP maps were being saturated with additional markers, simple sequence repeats (SSRs) markers were being designed for the rice genome (Wu and Tanksley 1993; Zhao and Kochert 1992, 1993). SSRs, also called as microsatellites, are tandemly arranged repeats of 1–6 nucleotide long DNA motifs that frequently exhibit variation in the number of repeats at a locus. As the flanking sequences may be unique, once SSR loci are cloned and sequenced, primers can be designed from the flanking sequences. The resultant sequence tagged microsatellite usually identifies a single locus. Because of high mutation rate of SSRs, these markers are often multi-allelic that often allows resolution on agarose gels. SSRs are also visualized on sequencing gels where single repeat differences can be resolved and, thus, all possible alleles detected. SSRs provide highly informative markers because they are co-dominant, generally highly polymorphic and highly reproducible between laboratories. These are amenable to high-throughput genotyping and thus useful for construction of high-density linkage maps, gene mapping and marker-assisted selection. Screening of a rice genomic library with

poly(dG-dA).(dC-dT) and poly(dG-dT).(dC-dA) probes by Wu and Tanksley (1993) indicated that (GA)n repeats occurred, on average, once every 225 kb and (GT)n repeats once every 480 kb. DNA sequencing of a set of ten randomly selected microsatellites indicated that the numbers of repeats ranged from 12 to 34 and that the patterns of microsatellites in rice were similar to those of humans and other mammals. Multiple alleles, ranging from 5 to 11, were observed at the microsatellite loci in a set of 20 rice accessions. This study was extended further by Panaud et al. (1996) who added 20 new microsatellite. Chen et al. (1997) estimated 5700–10,000 microsatellites in rice genome, with the relative frequency of different repeats decreasing with increasing size of the motif. A map with 120 microsatellite markers that were well distributed throughout the 12 chromosomes of rice was constructed. Cross-transferability of these markers to the wild relatives of the cultivated rice *O. sativa* revealed their utility in monitoring of gene introgression in breeding programmes (McCouch et al. 1997). The SSR map of the rice genome was further extended by adding 188 new markers derived partly from random genomic library (97 SSRs), and partly from known genes (26 SSRs) or EST (65 SSRs) sequences (Temnykh et al. 2000). The GA and CCG motifs were found most frequent among di- and tri-nucleotide SSRs in the genic regions of rice genome. There was uniform distribution of these markers irrespective of their sequence origin. These markers were successfully used to analyze genetic diversity (Yang et al. 1994; Olufowote et al. 1997; Cho et al. 2000), identify elite rice varieties (Singh et al. 2004) and to locate genes and QTLs on rice chromosomes using both intra- and interspecific crosses (Xiao et al. 1998; Bao et al. 2000; Zou et al. 2000; Bres-Patry et al. 2001; Moncada et al. 2001). However, a high-density rice genome map, which is required for fine scale mapping and map based cloning of genes, could not be constructed using SSR markers. This was primarily because of the difficulties in the marker development, as laborious iterations of genomic DNA library screening with SSR probes and sequencing of a large number of SSR positive clones were required.

3. DECODING OF THE RICE GENOME AND ITS IMPACT ON MARKER AND GENE DISCOVERY

Multiple international efforts were aimed at complete decoding of the rice genome. The International Rice Genome Sequencing Project (IRGSP; http://rgp.dna.affrc.go.jp/Seqcollab.html), with partner institutions from ten different countries namely, Japan, United States, China, France, Taiwan, India, Thailand, Korea, Brazil and the United Kingdom was officially launched in February 1998. The objective of this international collaborative effort led by Japan was to accelerate the sequencing of the genome of *japonica* rice cultivar Nipponbare by pooling resources and manpower, and to ensure public access to the high quality sequence data. IRGSP followed a map-based clone-by-clone approach for sequencing and contemplated to finish the genome sequencing by 2008. The multinational company Monsanto also carried out sequencing on contract at the University of Washington, Seattle, USA, following the same approach as IRGSP and announced that it had

sequenced the Nipponbare genome in April 2000 (http://www.rice-research.org). Although incomplete, the sequence was very informative, and Monsanto shared the sequence information with academic researchers and with the IRGSP. In January 2001, another multinational company Syngenta reported phase-I draft sequence of the same japonica cultivar Nipponbare, which was published in 2002 (Goff et al. 2002). It had contracted Myriad Genetics in Salt Lake City, Utah, USA, for sequencing using whole-genome shotgun approach. The fourth effort that surprised the world was that by the Beijing Genomics Institute (BGI), China, where sequencing of *indica* line 93–11 was carried out independently following a similar approach to that of Syngenta. BGI announced its plan to sequence in May 2000, and had the rough phase-I draft by October 2001, which was published in 2002 (Yu et al. 2002) simultaneously as that by Syngenta. While the sequence by Syngenta was not available for public, the *indica* sequence was placed in the public domain by the BGI. The sequence assembly by Syngenta has about 6X genome coverage and that by BGI now has a 6.28X coverage. Based on these assemblies, the rice genome size was predicted to be 433 Mb for *japonica* and 466 Mb for *indica*.

The IRGSP for its sequencing strategy required a physical map of the rice genome based on artificial chromosomes. A yeast artificial chromosome (YAC) physical map covering 63% of the rice genome was first constructed (Saji et al. 2001). However, instability, high chimera frequency, and difficulties in manipulation and purification make YAC clones difficult substrates for genome sequencing. Subsequently, low-copy number bacterial clones, namely bacterial artificial chromosomes (BACs) and P1-derived artificial chromosomes (PACs), were preferred for sequencing of the rice genome. Construction of a BAC library in *indica* rice cv Teqing and fingerprinting of 21,087 BAC clones from this library was reported (Tao et al. 2001). However, very limited genetic anchoring of the BAC contigs was carried out by this group. Subsequently, two deep-coverage BAC libraries from *japonica* rice cv Nipponbare were constructed, fingerprinted with HindIII and assembled into contigs, which were physically anchored onto the rice chromosomes by overgo hybridization, DNA gel blot hybridization, and *in silico* analysis (Chen et al. 2002). This physical map consisted of 65,287 fingerprinted BAC clones (including 2778 singletons) thought to represent 20-fold redundant coverage of the genome. The Rice Genome Research Programme, Japan constructed a transcript map with 6,591 EST markers (Wu et al. 2002), which were then used to screen the PAC and BAC libraries. Employing both Finger Printed Contigs (FPC) and Sequence-Tagged-Connector (STC) approaches, RGP constructed physical maps of rice chromosomes 1, 2, 6, 7, 8 and 9 (Wu et al. 2003). These two BAC and PAC based physical maps, were used to identify clones in the minimum tiling path. The physical gaps in the BAC/PAC tiling path were bridged using a variety of substrates, including PCR fragments, 10-kb plasmid and 40-kb fosmid clones. Individual PAC/BAC clones in the minimum tiling path were shotgun-sequenced and locally assembled. A completed phase-II and high-quality draft of the Nipponbare genome was announced by IRGSP on 18th December 2002. The phase II sequences of the BAC/PAC clones, however contained certain gaps, poor quality

regions, mis-assemblies, and single strand coverage regions. Several 'finishing' protocols were employed to fill the gaps and improve the sequence quality to an accuracy of 99.99%. Mis-assembly was resolved by manual detection, selection of a representative sub-clone, and full sequencing of the sub-clone. Manual editing and re-sequencing of some regions were carried out to improve coverage and the sequence quality. After two years of finishing work, the completion of rice genome sequencing was officially declared in December 2004 and published in August 2005 (IRGSP 2005). A total of 3,401 BAC/PAC clones were sequenced by the IRGSP to average tenfold sequence coverage, assembled, ordered and finished to a sequence quality of less than one error per 10,000 bases. The total assembled nucleotide sequence is 370.7 Mbp (Build 3.0 Pseudomolecules). Taking into account the estimated size of gaps, the entire rice genome is estimated to be 388.82 Mbp. This contrasts the genome size estimated earlier based on draft Phase I sequences by Syngenta and BGI. The high quality sequences obtained by IRGSP covered about 95% of the whole rice genome and 98.9% of the euchromatin region, while the remaining 5% corresponded to mostly centromere and heterochromatin regions.

The sequence of the rice genome has revealed many interesting features. A total of 37,544 non-transposable-element-related protein-coding genes were detected, with gene density of one per every 9.9 kb. This is the largest number of genes identified in plant and animal genomes that have been completely sequenced so far. About 71% of these genes are homologous to *Arabidopsis* genes and 29% exist in clusters. The number of genes that appeared to be unique to rice and the other cereals was 2,859, which might differentiate monocot and dicot lineages. Of the 37,544 predicted protein-coding genes, 17,016 could be supported by full-length cDNAs available in rice (Kikuchi et al. 2003). The density of expressed genes was greater on the distal portions of the chromosome arms except for the highly heterochromatic short arms of chromosomes 4, 9 and 10. About 0.2% of rice genome corresponded to sequences, which are highly homologous to the chloroplast and mitochondrial DNA that suggested repeated transfer of the organellar DNA to the nuclear genome. The transposon content of rice was about 35% with representatives from all known transposon superfamilies. Thus, the high quality sequence generated by the IRGSP provides finer details about the structure and evolution of the rice genome. Besides, this sequence resource is of immense value in many aspects of rice genetics and breeding: i) the predicted gene models would serve as the base for a functional analysis of the rice genome. Each of these genes can be characterised individually through RNAi technology. Gene chips based on these sequences have already been designed for use in whole genome expression profiling to identify and understand the role of genes in complex biological processes such as yield and abiotic stress tolerance. About 50,000 Tos17 retrotransposon insertion lines carrying 5,000,000 insertions have been produced in rice (Miyao et al. 2003). By comparing the sequences flanking these insertion sites with the rice genome sequences 11,487 target loci have been mapped on the 12 chromosomes. It has been observed that 3,243 of the predicted genes carry one Tos17 insertion each. Phenotypic alteration in these mutants can easily be associated with the disrupted

gene, thereby defining its biological function. ii) A large number of sequence based molecular markers such as SSRs and SNPs have been designed that practically saturate the genome. These markers can be used to map and tag genes for both qualitative and quantitative traits at very fine scale into short physical intervals, which in turn is helping isolation of genes by positional candidate approach. iii) Once genes are identified and functionally validated, it would be possible to identify new alleles in the germplasm by association mapping and thus enable designing allele specific markers. Use of such functional markers in crop breeding will make selection most efficient. iv) Rice is a small genome member of the grass family that includes many large genome members such as wheat, barley and maize. Synteny among the genomes of the family members is known. It is therefore possible to extend the rice gene information to other important cereal crops. In the following sections, we further elaborate on the development of sequence-based markers and their use in rice molecular breeding.

4. DEVELOPMENT AND USE OF SSR MARKERS

4.1. SSR Markers from the Genome Sequences of Rice

The first attempt to design new SSR markers *in silico* based on the rice genome sequence was made by Temnykh et al. (2001) using 74,127 BAC-end sequences (http://www.genome.clemson.edu/projects/rice/) with an average length of 500 bp, five complete BAC sequences generated by the US Rice Genome Sequencing Program (http://www.tigr. org/tdb/rice/) and 22 complete PAC sequences developed by the Rice Genome Research Program in Japan (http://rgp.dna.affrc.go.jp/GenomeSeq.html). SSRs were identified in a total of 57.8 Mb DNA sequence and classified as class I (>20bp repeat length) or class II (>12 bp <20 bp). The frequency of the class I SSRs in the BAC end-sequences was one per every 40 Kb, whereas in BAC and PAC sequences, the frequency was much higher (one SSR in every 16 kb). Class II SSRs were estimated to occur every 3.7 kb in BAC ends and every 1.9 kb in fully sequenced BAC and PAC clones. Microsatellites with poly(AT)n repeats represented the most abundant and polymorphic class of SSRs but were frequently associated with the Micropon family of miniature inverted-repeat transposable elements (MITEs) and were difficult to amplify. Temnykh et al. (2001) designed a new set of 200 Class I SSR markers and integrated into the microsatellite map of rice, providing immediate links between the genetic, physical, and sequence-based maps. With this, a total of 500 mapped SSR markers rigorously evaluated for amplification, map position, and allelic diversity became available for the rice researchers. Availability of genomic DNA sequences in the public domain eliminated the efforts that would have been otherwise required for genomic library construction, repeated screening of library, and sequencing of clones and thus accelerated SSR marker development.

Monsanto also analysed its draft sequence for SSR motifs and identified 6,650 di-, tri-, and tetranucleotide repeats of $\geq$ 24 bp flanked by 100 bp of unique sequences on

each side (http://www.rice-research.org). Temnykh et al. (2002) designed primers for 2,041 of these SSRs identified by Mansanto and another 199 SSRs (≥ 20 bp) identified in the IRGSP Phase II BAC/PAC sequences. Of these 2,240 new Class I SSR markers developed *in silico*, 1,825 could be mapped on the 12 rice chromosomes by e-PCR. Genetic mapping using 295 markers revealed that 95% of the e-PCR results were correct. Fifty-six SSR markers (2.8%) hit BAC/PAC clones on two or more different chromosomes and appeared to be multiple copy. Similar to earlier observations based on library screening, the highest proportion of these new markers corresponded to poly(GA) motifs (36%), followed by poly(AT) (15%) and poly(CCG) (8%) motifs. All di-, tri- and tetra-nucleotide motifs were represented among these markers, with the exception of three GC-rich classes, poly(GC), poly(GGCC) or poly(GGGC) motifs. AT-rich SSRs had the longest average repeat tracts, while GC-rich motifs were the shortest. In combination with the pool of 500 previously mapped SSR markers, a total of 2,740 SSR markers were made available for rice, which accounted for approximately one SSR every 157 kb of the rice genome.

Based on the draft sequence of Syngenta, Goff et al. (2002) identified a total of 48,351 SSRs with minimum length of 15 nucleotides at an frequency of one SSR every 8kb. More than 7,000 SSRs were found in the predicted genes, most of these (92%) being trinucleotides, so that length changes should not disturb the open reading frame. Rice genome could be however, saturated with long potentially hypervariable (≥ 20 bp) SSRs, by using the high quality genome sequences (IRGSP 2005). A total of 18,828 Class I di, tri and tetra-nucleotide SSRs, representing 47 distinctive motif families, were identified and annotated in the rice genome. Their physical positions could be defined in relation to widely used RFLP markers (Causse et al. 1998; Harushima et al. 1998) and previously published SSRs (McCouch et al. 2002). There was an average of one hypervariable SSRs in every 19.6 Kb sequence, with the highest density of markers occurring on chromosome 3 (one SSR/17.9 Kb) and the lowest occurring on chromosome 4 (one SSR/24.4 Kb). It is difficult to compare this observation with that of Goff et al. (2002) due to the difference in the quality of the genome sequence, its coverage and minimum length of the Class I markers in these two studies. Large number of SSR markers developed from the finished genome sequence of rice provides a unique opportunity to the rice researchers world over for fine mapping of genes, location of QTLs into short marker intervals and marker assisted selection.

4.2. SSR Markers from the EST/Unigene Sequences of Rice

Another source of SSR markers is the expressed sequence tag (EST) databases, which are continuously growing in size in most crop species (http://www.ncbi.nlm.nih.govEST). These EST databases can be mined for microsatellite motifs that would serve as locus-specific markers. Development of EST-based microsatellite markers thus would involve considerably lower cost and effort. Besides, such markers being derived from the conserved expressed

component of the genome are expected to show greater cross transferability between species and genera (Varshney et al. 2005). The unavailability of genomic and cDNA sequences, thus would not limit the development of robust microsatellite markers and various marker-based applications in related species and genera.

A major disadvantage of the EST-derived microsatellites is the sequence redundancy that yields multiple sets of markers at the same locus. However, more recently the random EST sequences are being assembled into unique gene sequences called unigenes (http://www.ncbi.nlm.nih.govunigenes) that circumvents the problem of redundancy in EST databases. For example, 41,600 EST sequences in barley have been assembled into just 1,240 unigenes. The unigene-based microsatellite markers (UGMS markers) would therefore, have the advantages of unique identity and positions in the transcribed regions of the genome. With the availability of large unigene databases, it is now possible to systematically search for microsatellites in the unigenes. The UGMS markers can be used for accurately assaying functional diversity in the natural populations and the available germplasm collections as well as for comparative mapping and evolutionary studies as anchor markers. Recently, Parida et al. (2006) mined SSRs in 48.8 Mb unigene sequences of rice and carried out a detailed analysis of their nature, frequency and genomic distribution. About 40% of the unigenes contained SSRs (excluding monomeric repeats) at a frequency of one SSR in every 3.6Kb of sequence. The proportion of the trinucleotide repeats was highest (80%) among the repeat motifs present in the unigenes, of which 85% were found within the ORFs, 12% in 5' UTRs and 3% in 3'UTRs. Among the dinucleotide repeats, GA was most common (21.2%) and CG was least (3%) frequent. Interestingly, (GA)n polynucleotides usually occurred in regions with balanced (40–50%) GC content, which favours robust PCR amplification (Temnykh et al. 2001) and therefore could be used for efficient genotyping applications. A total of 13,230 SSRs longer than 12 nucleotides, which included 2,780 Class I types, were identified and primers were designed for their amplification. The Class I SSR markers were physically mapped on the rice genome. The overall map density of these markers was one in every 141 kb and thus considered as a high-density physical map. The size of physical interval however varied from 174 bp to 3.26 Mb with the chromosome 3 having maximum UGMS markers (364) and thus the highest average map density (99 kb) followed by chromosome 1 (355 markers, 121 kb). The least number of markers (128) was placed on chromosome 11 giving the lowest map density (221 kb). Cross transferability of a sizable number of rice class I markers to other cereals namely wheat, barley, maize and sorghum suggested utility of these markers in comparative mapping of cereal genomes and genes.

5. DEVELOPMENT AND USE OF SNP MARKERS

Single nucleotide polymorphisms (SNPs) are the most abundant form of molecular markers present in any genome and rice is no exception. They arise due to spontaneous mutations in the DNA leading to base transitions (purine to purine

and pyrimidine to pyrimidine substitutions) or transversions (purine to pyrimidine and pyrimidine to purine subsitutions). Small insertions and deletions (InDels) of one or two bases are also quite common and often are clubbed together with SNPs. While class I SSR loci are present on an average at every 19.6 kbp in the rice genome (IRGSP 2005), the SNPs are present at every 100 bp or less (http://www.ncbi.nlm.nih.gov/projects/SNP/), and therefore are most useful for fine mapping of genes and QTLs for the agronomic traits in order to develop tightly linked DNA markers for molecular breeding applications. Due to very low frequency of the spontaneous mutations and much greater possibility of base transitions as compared to transversions, SNPs are often bi-allelic in nature. Though, theoretically it is possible to have up to four alleles at any base position and we do see tri- and tetra-allelic SNPs but rarely. In contrast, multiple alleles are common at the individual SSR loci due to differences in the number of repeat units generated by replication slippage and unequal crossing over. A group of SNPs present in a single gene or small stretch of DNA, called haplotype, may have 2^n number of possible combinations, giving them a high discriminating power and multiple alleles at the gene level. Similar to the SSRs, large number of SNPs occur in the intergenic spaces, but these are not likely to have any direct effect on the traits of agronomic importance. The genic SNPs on the other hand may cause single amino acid substitutions, premature termination of the amino acid reading frame or a frame shift in case of InDels, thus directly affecting the function of the genes, some times leading to drastic phenotypic consequences at the trait level. Many genic SNPs are present at the redundant base positions in the amino acid codons, and therefore have no functional significance, but these can still be used as gene based perfect markers for molecular breeding.

Decoding of the genomes of *japonica* rice variety 'Nipponbare' and *indica* rice variety '93-11' and availability of BAC-end sequence data from *indica* rice variety 'Kasalath' has led to the identification of more than 3.9 million rice SNPs and this information is now available in the public domain at the NCBI website (http://www.ncbi.nlm.nih.gov/projects/SNP/). Furthermore, the rice genomics revolution has led to accumulation in the public databases of more than 1.1 million expressed sequence tags (ESTs) from different varieties of rice, a number that is second only to human ESTs, and these provide another important resource for discovering SNPs in the expressed genic sequences of rice. Feltus et al. (2004) aligned draft sequences of rice varieties 'Nipponbare' and '93-11', and after filtering multiple copy and low quality sequences, they identified 408,898 SNPs/InDels. These include only a subset of the available SNPs, in particular excluding large number of SNPs that might occur in repetitive DNA, but with higher likelihood that this subset is useful. Direct sequencing of a random sample of these SNP sites suggested that 79.8% $\pm$ 7.5% of the *in silico* discovered SNPs were real. These SNPs were present in all the twelve rice chromosomes but were not randomly distributed across the genome as 566 rice genomic regions had unusually high (48.6 Mb, 13.6% of the genome) or low (64.7 Mb, 18.1% of genome) polymorphism rates (http://www.plantgenome.uga.edu/snp). The international rice genome sequencing

consortium identified 80,127 polymorphic SNP/InDel sites that distinguish between the *japonica* rice variety 'Nipponbare' and *indica* variety 'Kasalath' for which 15.6 Mb of high quality BAC-end sequence data was generated (IRGSP 2005).

The most challenging task ahead is validation of these large numbers of SNPs through wet lab experiments, and establishing their association with the traits of agronomic importance so that they can be used in practical rice breeding. Most of the SNPs in the public domain databases were identified by *in silico* analysis of the available sequences and have rarely been validated in wet lab conditions. Now we have many novel high throughput techniques for SNP genotyping, and these are being employed increasingly for SNP validation. Restriction fragment length polymorphisms (RFLPs), the earliest type of molecular markers, in fact are due to SNPs leading to loss or gain of restriction sites. The RFLPs represent only a fraction of the total SNP content in a genome and the technique is quite tedious for regular use in rice breeding. Now it has been replaced by analysis using high throughput capillary electrophoresis of PCR amplified DNA fragments, resolution of small size differences between the allelic DNA bases using MALDI-TOF or hybridization of labeled DNA samples to large number of oligonucleotide probes immobilized on glass slides (micro arrays) and nano-size beads (bead arrays). The discovery and development of SNP markers requires techniques like DHPLC, TILLING and most importantly re-sequencing of the allelic DNA fragments, but after validation the subsequent analysis of the SNP loci in different rice varieties and segregating populations can be done using a number of techniques. The most common among these are CAPS, SSCP, SNuPe, DHPLC and MALDI-TOF mass array. Most recently hybridization of sample DNAs to synthetic oligonucleotide probes immobilized on micro or nano size beads is offering unprecedented throughput. These techniques will allow screening of large number of segregating genotypes simultaneously for the presence of favorable alleles of a number of genes in early generation of the breeding program, hence compressing the breeding cycle as well as breeding population for actual replicated field trails.

The present limitations in the molecular breeding of rice are: (i) non-availability of tightly linked or actual gene-based markers for the traits and (ii) high cost of analysis per sample. The first limitation is getting removed rapidly as several genes of agronomic importance are being discovered with the help of rice genome sequence information. The cost of analysis per sample is also coming down with the introduction of marker multiplexing and wider use of the high throughput genotyping techniques. Although the initial cost of the equipments are quite high, establishment of genotyping service facilities will easily take care of this. High throughput SNP mapping projects have already been initiated at IRRI through international collaboration in partnership with private service provider companies like Illumina (http://www.illumina.com/products/prod_snp.ilmn) and Perlegene Sciences Inc. (www.perlegen.com). In due course of time one can see large number of validated SNPs and haplotypes of rice in a way similar to the human SNP database developed by the International HapMap consortium (2006).

6. DISCOVERY OF CANDIDATE GENES AND GENE BASED MARKERS

The most significant impact of the availability of high quality rice genome sequence information has been on rapid discovery of genes of agronomic importance and elucidation of the underlying allelic differences at the level of gene sequence that are directly responsible for the differences in the traits at the morphological level. This is helping development of gene-based perfect markers for molecular breeding applications. These markers will not recombine with the trait because marker itself is responsible for the differences in the trait. A partial list of such genes identified in rice using map-based cloning approach is given in Table 1. This list is likely to grow very fast because intensive research is going on around the world, both in public and private sectors, to identify many more such genes for yield components, quality traits and stress tolerance (both biotic and abiotic) in rice. One difficulty is that, apart from disease resistance genes, most of the important agronomic traits e.g. yield, quality and abiotic stress tolerance, show quantitative inheritance due to involvement of many genes with small individual effects and role of gene into gene (epistasis) and gene into environment interactions. We do not yet have many validated markers for these traits that can explain major part of the phenotypic variation in the segregating populations.

Two different approaches have been used for the identification of candidate genes for these complex quantitative traits. First approach uses forward genetics where two parents with contrasting phenotypes, e.g. salt tolerant and salt susceptible rice varieties, are first crossed to develop a mapping population of about 200 F2, RILs (recombinant inbred lines) or BILs (backcross inbred lines). After identification of the QTL intervals for the salt tolerance using this population, a much larger F2 population of typically more than 1000 individuals developed from intercrossing of

Table 1. Genes for important agronomic traits cloned recently using the rice genome sequence information

S. no.	Trait	Gene	Reference
1.	Bacterial leaf blight resistance	*Xa 21* (NBS-LRR type receptor kinase)	Song *et al.* (1995)
2.	Plant height	*Sd 1* (gibberellin-20-oxidase)	Sasaki *et al.* (2002)
3.	Amylose content	*Sbe 3* (starch branching enzymes)	Liu *et al.* (2004)
4.	Grain number	*OsCKX2* (cytokinin oxydase)	Ashikari *et al.* (2005)
5.	Salt tolerance	*SKC1* (a HKT type transporter)	Ren *et al.* (2005)
6.	Grain aroma	*BAD2* (betaine aldehyde dehydrogenase 2)	Bradburry *et al.* (2005)
7.	Blast resistance	Pik^h (NBS-LRR type protein)	Sharma *et al.* (2005)
8.	Submergence tolerance	*Sub1* (Ethylene response factor-like protein)	Xu *et al.* (2006)
9.	Lodging tolerance	*Lsi 1* (Silicon transporter)	Ma *et al.* (2006)
10.	Seed shattering	*qSH 1* (BEL1-type homeobox)	Konishi *et al.* (2006)

QTL-NILs (differing for a single QTL only), is then used for fine mapping of the individual QTL interval to identify and validate the function of the candidate gene (Ren et al. 2005). Availability of rice genome sequence with complete information on the genes present in specific physical/genetic interval is of great help in this process and many key genes have been cloned during the last two years as compared to only limited examples of map-based cloning before the genomic era (Table 1). The second genomic approach is based on the gene expression profiling of the rice varieties and land races that are known to possess the genes of agronomic importance e.g. 'Pokkali' for salt tolerance (Kawasaki et al. 2001; Sahi et al. 2006) and 'N22' for drought tolerance (Gorantla et al. 2005). Analysing differential expression of genes with and without stress has led to identification of several hundred probable candidate genes that could be responsible for the phenotypic difference at the trait level. Unfortunately, it has not been possible to narrow down the number of candidate genes to a small number due to common stress response pathways and cross talks between alternate response pathways. A combination of the genetic fine mapping and differential expression profiling is likely to speed up gene discovery for complex agronomic traits in rice by drastically bringing down the number of candidate genes from several hundreds to tens, as described for the identification of candidate genes for ovariole number in case of fruit fly (Wayne and McIntyre 2002).

7. PROSPECTS OF HIGH THROUGHPUT MARKER-ASSISTED SELECTION (MAS) IN RICE

Molecular markers are used as gene tags to select for desirable genotypes. It offers several advantages over conventional phenotype based selection: i) Selection can be practiced at any stage of plant growth, ii) Creation of selection environment is not required, iii) Several useful genes can by combined in one genotype irrespective of nature of interaction among them, iv) Heterozygous carriers of useful recessive alleles can be identified easily, and v) by judicious combination of background and foreground selection, it is possible to save time and effort. Marker assisted selection in rice is being practiced for a few traits including resistance against bacterial leaf blight and blast diseases (Sanchez et al. 2000; Singh et al. 2001; Joseph et al. 2004; Liu et al. 2003; Toojinda et al. 2005). Molecular mapping of many other useful genes although has been reported, the identified markers are yet to be used in MAS. Many of the linked markers were RFLPs, which are not amenable to automation and thus a large number of samples cannot be handled in a short time. Moreover, absence of tightly linked markers that are polymorphic enough to be used in several different segregating populations has limited the use of markers in rice breeding. The availability of a large number of sequence based SSR and SNP markers that saturate the rice genome would be of considerable value in fine mapping of target genes. Use of these markers is expected to eliminate the problems due to recombination and lack of polymorphism. Besides, genotyping of a large population using a set of markers can be achieved in a short time since both the

marker classes allow complete automation. For instance, one of the primers in case of SSR markers can be labeled using fluorescent dye and the amplified products can be sized quickly using a automated fragment analyzer. Multiplexing of a set of primer pairs in PCR would significantly increase the speed of analysis (Coburn et al. 2002). Similarly, SNP analysis is completely automated and thus would allow very high throughput genotyping. Since there is no involvement of cost in marker development, routine use of these markers in rice breeding is feasible provided some initial investment is made on laboratory infrastructure. However, MAS is not a complete replacement for evaluation of phenotype in the field. It is recommended that MAS for difficult-to-phenotype traits needs to be combined with phenotypic selection for easy-to-phenotype traits for deriving maximum benefit from the vast genomic resources including molecular markers available in rice. Integration of molecular markers in rice breeding is a necessity for imparting far greater efficiency to the development of superior genotypes.

REFERENCES

Ashikari M, Sakakibara H, Lin S, Yamamoto T, Takashi T, Nishimura A, Angeles ER, Qian Q, Kitano H, Matsuoka M (2005) Cytokinin oxidase regulates rice grain production. Science 309:741–745

Bao JS, Zheng XW, Xia YW, He P, Shu QY, Lu X, Chen Y, Zhu LH (2000) QTL mapping for the paste viscosity characteristics in rice (*Oryza sativa* L.). Theor Appl Genet 100:280–284

Bradbury LMT, Fitzgerald TL, Henry RJ, Jin Q, Waters DLE (2005) The gene for fragrance in rice. Plant Biotech J 3:363–370

Bres-Patry C, Loreux M, Clement G, Bangratz M, Ghesquiere A (2001) Heredity and genetic mapping of domestication-related traits in a temperate *japonica* weedy rice. Theor Appl Genet 102:118–126

Causse MA, Fulton TM, Cho YG, Ahn SN, Chunwongse J, Wu K, Xiao J, Yu Z, Ronald PC, Harrington SE, Second G, McCouch SR, Tanksley SD (1994) Saturated molecular map of the rice genome based on an interspecific backcross population. Genetics 138:1251–1274

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

Chen M, Presting G, Barbazuk WB, Goicoechea JL, Blackmon B, Fang G, Kim H, Frisch D, Yu Y, Sun S et al (2002) An integrated physical and genetic map of the rice genome. Plant Cell 14:537–545

Cho YG, Ishii T, Temnykh S, Chen X, Lipovich L, Park WD, Ayres N, Cartinhour S, McCouch SR (2000) Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). Theor Appl Genet 100:713–722

Coburn JR, Temnykh SV, Paul EM, McCouch SR (2002) Design and application of microsatellite marker panels for semiautomated genotyping of rice (*Oryza sativa* L.). Crop Sci 42:2092–2099

Feltus FA, Wan J, Schulze SR, Estill JC, Jiang N, Paterson AH (2004) Genome alignments of *japonica* and *indica* subspecies. Genome Res 14:1812–1819

Goff SA, Ricke D, Lan TH, Presting G, Wang RL, Dunn M, Glazebrook J et al (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science 296:92–100

Harushima Y, Yano M, Shomura A, Sato M, Shimano T, Kuboki Y, Yamamoto T, Lin SY, Antonio BA, Parco A et al (1998) A high-density rice genetic linkage map with 2275 markers using a single F2 population. Genetics 148:479–494

International HapMap Consortium (2005) A haplotype map of the human genome. Nature 437: 1299–1320

Gorantla M, Babu PR, Lachagari VBR, Feltus FA, Paterson AH, Reddy AR (2005) Functional genomics of drought stress response in rice: Transcript mapping of annotated unigenes of an *indica* rice (*Oryza sativa* L. cv. Nagina 22). Curr Sci 89:496–514

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

Iwata N (1986) The relationship between cytologically identified chromosomes and linkage groups in rice. In: Rice genetics. Proceedings of the international rice genetic symposium, Los Baños, Philippines, Manila (Philippines) 27–31 May 1985; International rice research institute, pp 229–238

Joseph M, Gopalakrishnan S, Sharma RK, Singh VP, Singh AK, Singh NK, Mohapatra T (2004) Combining bacterial blight resistance and Basmati quality characteristics by phenotypic and molecular marker-assisted selection in rice. Mol Breed 13:377–387

Kawasaki S, Borchert C, Deyholos M, Wang H, Brazille S, Kawai K, Galbraith D, Bohnert HJ (2001) Gene expression profiles during the initial phase of salt stress in rice. Plant Cell 13:889–906

Kikuchi S, Satoh K, Nagata T, Kawagashira N, Doi K, Kishimoto N, Yazaki J, Ishikawa M, Yamada H, Ooka H et al (2003) Collection, mapping, and annotation of over 28,000 cDNA clones from *japonica* rice. Science 300:1566–1569

Konishi S, Izawa T, Lin SY, Ebana K, Fukuta Y, Sasaki T,Yano M (2006) An SNP caused loss of seed shattering during rice domestication. Science 312:1392–1396

Kurata N, Nagamura Y, Yamamoto K, Harushima Y, Sue N, Wu J, Antonio BA et al (1994) A 300-kilobase interval genetic map of rice including 883 expressed sequences. Nat Genet 8:365–372

Kuwada Y (1910) A cytological study of *Oryza sativa* L. Bot Mag (Tokyo) 24:267–280

Liu S-P, Li X, Wang C-Y, Li X-H, He Y-Q (2003) Improvement of resistance to rice blast in Zhenzhan 97 by molecular marker-aided selection. Acta Bot Sin 45:1346–1350

Liu X, Gu M, Han Y, Ji Q, Lu J, Gu S, Zhang R, Li X, Chen J, Korban SS, Xu M (2004) Developing gene-tagged molecular markers for functional analysis of starch-synthesizing genes in rice (*Oryza sativa* L.). Euphytica 135:345–353

Ma JF, Tamai K, Yamaji N, Mitani N, Konishi S, Katsuhara M, Ishiguro M, Murata Y, Yano M (2006) A silicon transporter in rice. Nature 440:688–691

McCouch SR, Kochert G, Yu ZH, Wang ZY, Khush GS, Coffman WR, Tanksley SD (1988) Molecular mapping of rice chromosomes. Theor Appl Genet 76:815–829

McCouch SR, Chen XL, Panaud O, Temnykh S, Xu YB, Cho YG, Huang N, Ishii T, Blair M (1997) Microsatellite marker development, mapping and applications in rice genetics and breeding. Plant Mol Biol 35:89–99

Miyao A et al (2003) Target site specificity of the Tos17 retrotransposon shows a preference for insertion within genes and against insertion in retrotransposon rich regions of the genome. Plant Cell 15:1771–1780

Moncada P, Martinez CP, Borrero J, Chatel M, Gauch H Jr, Guimaraes E, Tohme J, McCouch SR (2001) Quantitative trait loci for yield and yield components in an *Oryza sativa* × *Oryza rufipogon* BC2F2 population evaluated in an upland environment. Theor Appl Genet 102:41–52

Morinaga T (1937) On the microsporogenesis of the various interspecific hybrids of *Oryza*. A preliminary note. Jpn J Genet 13:245

Morinaga T (1939) Cytogenetics in rice. (*Oryza sativa* L.). Bot Zool 7:179–183

Nagao S, Takahashi M (1963) Trial construction of twelve linkage groups in Japanese rice (Genetical studies on rice plant, XXVII). J Fac Agric Hokkaido Univ 53:72–130

Olufowote JO, Xu Y, Chen X, Park WD, Beachell HM, Dilday RH, Goto M, McCouch SR (1997) Comparative evaluation of within-cultivar variation of rice (*Oryza sativa* L.) using microsatellite and RFLP markers. Genome 38:1170–1176

Parida S, Kumar KAR, Dalal V, Singh NK, Mohapatra T (2006) Unigene derived microsatellite markers for the cereal genomes. T Appl Genet 112:808–817

Ren ZH, Gao JP, Li LG, Cai XL, Huang W, Chao DY, Zhu MZ, Wang ZY, Luan S, Lin HX (2005) A rice quantitative trait locus for salt tolerance encodes a sodium transporter. Nat Genet 37:1141–1146

Sahi C, Singh A, Kumar K, Blumwald E, Grover A (2006) Salt stress response in rice: genetics, molecular biology, and comparative genomics. Funct Integr Genomics 6:263–284

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

Sanchez AC, Brar DS, Huang N, Li Z, Khush GS (2000) Sequence tagged site marker assisted selection for three bacterial blight resistance genes in rice. Crop Sci 40:792–797

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

Sharma TR, Madhav MS, Singh BK, Shanker P, Jana TK, Dalal V, Pandit A, Singh A, Gaikwad K, Upreti HC, Singh NK (2005) High-resolution mapping, cloning and molecular characterization of the *Pi-kh* gene of rice, which confers resistance to *Magnaporthe grisea*. Mol Gen Genomics 274:569–578

Shastry SVS, Ranga Rao DR, Misra RN (1960) Pachytene analysis in *Oryza*. I. Chromosome morphology in *Oryza sativa*. Indian J Genet Plant Breed 20:5–21

Singh S, Sidhu JS, Huang N, Vikal Y, Li Z, Brar DS, Dhaliwal HS, Khush GS (2001) Pyramiding three bacterial blight resistance genes (*xa5*, *xa13* and *Xa21*) using marker-assisted selection into indica rice cultivar PR106. Theor Appl Genet 102:1011–1015

Singh RK, Sharma RK, Singh AK, Singh VP, Singh NK, Tiwari SP, Mohapatra T (2004) Suitability of mapped sequence tagged microsatellite site markers for establishing distinctness, uniformity and stability in aromatic rice. Euphytica 135:135–143

Song WY, Wang GL, Chen LL, Kim HS, Pi LY, Holsten T, Gardner J, Wang B, Zhai WX, Zhu LH et al (1995) A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa21*. Science 270:1804–1806

Tao Q, Chang YL, Wang JZ, Chen HM, Islam-Faridi MN, Scheuring C, Wang B, Stelly DM, Zhang HB (2001) Bacterial artificial chromosome-based physical map of the rice genome constructed by restriction fingerprint analysis. Genetics 158:1711–1724

Temnykh S, Park WD, Ayres N, Cartinhour S, Hauck N, Lipovich L, Cho YG, Ishii T, McCouch SR (2000) Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa* L.). Theor Appl Genet 100:697–712

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

Toojinda T, Tragoonrung S, Vanavichit A, Siangliw JL, Pa-In N, Jantaboon J, Siangliw M, Fukai S (2005) Molecular breeding for rainfed lowland rice in the Mekong region. Plant Prod Sci 8:330-333

Wayne ML, McIntyre LM (2002) Combining mapping and arraying: an approach to candidate gene identification. Proc Natl Acad Sci USA 99:14903–14906

Wu KS, Tanksley SD (1993) Abundance, polymorphism and genetic mapping of microsatellites in rice. Mol Gen Genet 241:225–235

Wu J, Maehara T, Shimokawa T, Yamamoto S, Harada C, Takazaki Y, Ono N, Mukai Y, Koike K, Yazaki J, Fujii F, Shomura A, Ando T, Kono I, Waki K, Yamamoto K, Yano M, Matsumoto T, Sasaki T (2002) A comprehensive rice transcript map containing 6591 expressed sequence tag sites. Plant Cell 14:525–535

Wu J, Mizuno H, Hayashi-Tsugane M, Ito Y, Chiden Y, Fujisawa M, Katagiri S, Saji S, Yoshiki S, Karasawa W, Yoshihara R, Hayashi A, Kobayashi H, Ito K, Hamada M, Okamoto M, Ikeno M, Ichikawa Y, Katayose Y, Yano M, Matsumoto T, Sasaki T (2003) Physical maps and recombination frequency of six rice chromosomes. Plant J 36:720–730

Xiao J, Li J, Grandillo S, Ahn SN, Yuan L, Tanksley SD, McCouch SR (1998) Identification of trait-improving quantitative trait loci alleles from a wild rice relative, *Oryza rufipogon*. Genetics 150:899–909

Xu K, Xu X, Fukao T, Canlas P, Maghirang-Rodriguez R, Heuer S, Ismail AM, Bailey-Serres J, Ronald PC, Mackill DJ (2006) Sub1A is an ethylene-response-factor-like gene that confers submergence tolerance to rice. Nature 442:705–708

Yang GP, Saghai Maroof MA, Xu CG, Zhang Q, Biyashev M (1994) Comparative analysis of microsatellite DNA polymorphism in landraces and cultivars of rice. Mol Gen Genet 245:187–194

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

Zhao X, Kochert G (1992) Characterization and genetic mapping of a short, highly repeated, interspersed DNA sequence from rice (*Oryza sativa* L.). Mol Gen Genet 231:353–359

Zhao XP, Kochert G (1993) Phylogenetic distribution and genetic mapping of a (GGC)n microsatellite from rice *(Oryza sativa* L.). Plant Mol Biol 21:607–614

Zou JH, Pan XB, Chen ZX, Xu JY, Lu JF, Zhai WX, Zhu LH (2000) Mapping quantitative trait loci controlling sheath blight resistance in rice cultivars (*Oryza sativa* L.) Theor Appl Genet 101: 569–573

CHAPTER 9

MARKER-ASSISTED SELECTION IN SORGHUM

GEBISA EJETA* AND JOSEPH E. KNOLL

Department of Agronomy, Purdue University, 915 W. State Street, West Lafayette, IN 47907, USA

Abstract: Sorghum [*Sorghum bicolor* (L.) Moench] is an important food and feed crop in many parts of the world, and has potential uses in the biofuels industry. Compared to most other cereals, sorghum is more tolerant to many abiotic stresses, including heat, drought, and flooding, making it an ideal crop for growing on marginal lands as demands for food, feed, and energy increase. Though it is generally stress-tolerant, the true potential of sorghum can only be realized through concerted genetic improvement programs. The use of DNA-based markers for the genetic analysis and manipulation of important agronomic and stress-tolerance traits is becoming an increasingly useful tool in sorghum improvement. The known germplasm of sorghum is incredibly diverse, and molecular markers are being used to assess this diversity to help manage large germplasm collections, and to make these collections more useful to breeders. Molecular markers have been used in sorghum to identify quantitative trait loci (QTL) for many complex traits, including pre-flowering and post-flowering drought tolerance, early-season cold tolerance, and resistance to the parasitic weed *Striga*. However, progress in utilizing these QTL had been limited by the lack of a standard genetic map and a common nomenclature for the various linkage groups of sorghum. Fortunately, the genetic map of sorghum has recently become standardized, and has also been linked to physical chromosomes. The use of a common map will facilitate the exchange of marker and QTL information between sorghum research groups. This will allow independent validation of QTL and should expedite efforts to use these QTL for the development of improved sorghum cultivars through marker-assisted selection and trait introgression. Newer, faster marker technologies based on single nucleotide polymorphisms (SNPs), and mapping methods based on linkage disequilibrium (association mapping), will soon become useful tools for future efforts to improve this important crop.

1. INTRODUCTION

Sorghum [*Sorghum bicolor* (L.) Moench] is an important crop in many parts of the world and is grown for food, feed, and industrial purposes. Sorghum ranks fourth among grain crops in the United States, behind maize (*Zea mays* L.), soybean

*Author for correspondence: gejeta@purdue.edu

R.K. Varshney and R. Tuberosa (eds.), Genomics Assisted Crop Improvement: Vol. 2: Genomics Applications in Crops, 187–205.

[*Glycine max* (L.) Merr.], and wheat (*Triticum aestivum* L.) in total production (FAO 2004). It is also a major crop in many parts of Africa and some Asian countries. Compared to most other cereals, sorghum is more tolerant to many abiotic stresses, including heat, drought, and flooding. Combined with its potential use in the emerging biofuels industry, sorghum is an ideal candidate for a more concerted crop improvement program as agriculture is pushed to more marginal lands, and food and energy demands increase. The use of DNA-based markers for the genetic analysis and manipulation of important agronomic traits is becoming an increasingly useful tool in sorghum improvement.

DNA markers have the potential to enhance the operation of a plant breeding program through a number of ways. Genetic fingerprints can be used for the identification of genetic stocks and cultivars, and for the verification of pedigree information (Soller and Beckmann 1983). Comparison of the unique genetic signatures of germplasm accessions has allowed breeders to assess the amount of diversity in their breeding material and to determine relatedness between individuals (Deu et al. 1994; de Oliviera et al. 1996; Yang et al. 1996; Menkir et al. 1997; Abu Assar et al. 2005). As genotyping technologies have improved, very dense genetic maps of sorghum have been created (Menz et al. 2002; Bowers et al. 2003). The ability to associate quantitative phenotypic data with genetic maps has helped to increase our understanding of the inheritance of complex agronomic traits in sorghum (Pereira and Lee 1995; Tuinstra et al. 1996; 1997), which is beginning to lead to marker-assisted plant breeding. However, the application of this technology is still relatively new, and it may take some time before marker-assisted selection (MAS) becomes a routine operation in most sorghum breeding programs. The sorghum genome is expected to be completely sequenced in the very near future and opportunities for integrating genetic maps with the physical map of the crop should increase the effectiveness of knowledge-based breeding of some of the more difficult traits in sorghum. This review summarizes the knowledge base generated and experience gained in the identification and analysis of quantitative trait loci (QTL) and MAS in sorghum both in our own research program as well as in the greater sorghum research community.

2. GENETIC DIVERSITY ASSESSMENT IN SORGHUM GERMPLASM

Analyses of the extent and distribution of genetic variation in a crop are essential for understanding the evolutionary relationships between germplasm accessions and for sampling genetic resources more systematically for both breeding and conservation purposes. Traditionally, genetic resources in sorghum are classified by taxonomists based on a set of morphological characteristics. Cultivated sorghum (*S. bicolor* subsp. *bicolor*) is divided into five major races, namely *bicolor*, *guinea*, *caudatum*, *kafir*, and *durra*, and the ten possible hybrid groups among these races (Harlan and de Wet 1972; Harlan and Stemler 1976; de Wet 1978). There are several ‘wild’ subspecies within *S. bicolor*, and races within these subspecies as well, many of

which harbor potentially valuable traits. However, the morphological characteristics used in the classification of sorghum into different races are conditioned by relatively few genes. On the other hand, important traits which are related to habitat adaptation, and which exhibit enormous variability among sorghum germplasm, are complex and quantitatively inherited. Also, closely related accessions could potentially look very different in the field, and some accessions with similar appearance may actually be of very different origins. Hence, classifying germplasm accessions based solely on a few discrete morphological characters may not provide an accurate indication of the genetic diversity present within the sorghum germplasm.

Several studies have been conducted to characterize the genetic diversity among the various races of sorghum (Deu et al. 1994; de Oliviera et al. 1996), and within smaller collections from specific regions (Yang et al. 1996). Because both natural and human selection efforts should have contributed to the current genetic differences in sorghum, we hypothesized that geographic distribution would have a greater effect on genetic diversity than racial classification. To test this hypothesis, we used molecular markers to analyze genetic diversity among the five major races of cultivated sorghum (Menkir et al. 1997). One-hundred ninety accessions representing each race and each of the major geographical centers of distribution were randomly selected from the world collection of sorghum maintained by the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT). The 190 accessions were divided into 38 sets, in each of which five randomly-selected accessions from all five races were included to ensure parallel comparison of races and to minimize biases while scoring marker products. A total of 82 random amplified polymorphic DNA (RAPD) primers were used for DNA amplification. Only 53 primers produced clearly scorable markers, generating a total of 220 loci. A high level of genetic variation was detected among the sorghum accessions. The results of our study indicated that within-race genetic diversity was high for race *bicolor*, and *guinea*, and low for race *kafir*. Partitioning the genetic variation further revealed that 86% of the total genetic variation occurred among individual accessions and only 14% among races. Examination of the degree of association of accessions with their geographic areas of origin indicated that only 13% of the total genetic variation was attributable to divergence among regions. In spite of the limited differentiation among regions, the extent of genetic diversity within and among regions showed some trends. Though represented by a large number of accessions, Southern African germplasm exhibited the least amount of genetic diversity suggesting a narrow genetic base of accessions from this region. By contrast, sorghums from West Africa exhibited a high level of genetic diversity with a least number of accessions. Genetic diversity in Central and Eastern Africa, as well as the Middle East, was as high as that observed in accessions from West Africa.

Simple sequence repeat (SSR) markers are becoming more popular for diversity analysis studies in sorghum because of their repeatability, simplicity, and ability to be semi-automated. Also, their multiallelic nature makes these markers more informative than most other marker types. Abu Assar et al. (2005) used only 16 SSR

markers to assess the diversity present within a collection of 96 sorghum accessions representing very different geographies and races. Each accession showed a unique genetic signature, demonstrating the utility of SSRs for differentiating accessions. By applying the unweighted pair-grouped method with arithmetic average (UPGMA) clustering to the genotypic data, the accessions tended to cluster based on geographic origin, and within these clusters accessions of similar morphological classification tended to also congregate together.

Germplasm collections at most programs often become very large through time, containing too many accessions for a single breeder to assess in a practical manner. Thus, a system for assembling a smaller number of specific accessions more methodically to establish a so-called 'core collection' has been proposed. The ideal core collection should encompass all of the genetic diversity contained within the larger collection, represented by fewer individual accessions. Such a core collection would serve as a starting point for breeders interested in finding new sources for specific traits. Core collections can be easily evaluated by a breeder for specific characteristics, and if desirable individuals are found they can be utilized in a breeding program, and related accessions from the larger collection can be evaluated if the breeder so desires. However, developing such an informative panel for a core collection can be a challenge. In general, core collections need to be small enough to be of practical use, while containing a significant portion of the genetic diversity of the species represented by the whole collection.

Grenier et al. (2000a) were among the first to describe phenotype-based methods for assembling a core collection from landrace sorghum collections. They described three different methods of subsampling. The first sampling scheme involved stratification of the sorghum collection held at ICRISAT based on photoperiod sensitivity, followed by random sampling. The second scheme was based on quantitative characteristics, and the third was based on geographical origin and other specific traits. All accessions in each subset were genotyped for 15 SSR markers, using semi-automated allele sizing (Grenier et al. 2000b). Similar measures of genetic diversity were obtained for all three sampling strategies. The mean number of alleles per locus was similar for all three sampling methods, as were both the mean and range of diversity indices for all loci. One locus in particular had relatively low diversity, but the level was comparable in all three samples. A similar number of rare alleles were also retained by each sampling strategy, though a few rare alleles were detected in only one sample. Nonetheless, it was concluded that all three sampling methods retained a similar level of genetic diversity, representative of the global set (all three samples combined).

In addition to maximizing genetic diversity in a breeding program or core collection, elimination of redundant accessions is also important for reducing the time and costs required to maintain large collections. Dean et al. (1999) used the same 15 SSR markers to examine the genetic diversity within 19 accessions listed as 'Orange' in the US national sorghum collection. Their analysis revealed two different groups of identical entries, which could potentially be pooled to eliminate redundancy. They also found many nearly identical entries, and concluded that

some of these could also be pooled. The total number of 'Orange' accessions could be reduced to 10, with only a 10% loss of genetic diversity among the entries. In addition, because they genotyped five individual plants from each entry, they found that one accession was not homogeneous, and may have been contaminated with seeds from another source. Thus, the determination of genetic purity of accessions or breeding lines is another important use for molecular markers in sorghum. A large scale genetic diversity project involving molecular characterization of 3000 sorghum accessions at 50 SSR loci is underway at ICRISAT in collaboration with CIRAD and CAAS (R.K. Varshney, personal communication).

3. GENETIC MAPPING IN SORGHUM

The first group of genetic linkage maps of sorghum consisted primarily of RFLP markers derived from maize probes (Hulbert et al. 1990; Whitkus et al. 1992; Binelli et al. 1992; Melake-Berhan et al. 1993). Comparison of these maps with those of maize revealed a high degree of synteny between the two genomes. Whitkus et al. (1992) and Melake-Berhan et al. (1993) also noted that many of the probes which mapped to a single locus in sorghum were duplicated in maize, suggesting possible duplication events in the evolution of maize after its divergence from sorghum. These early maps, however, did not contain enough markers to resolve ten linkage groups, which is the haploid chromosome number for sorghum. Chittenden et al. (1994) published the first 'complete' linkage map of sorghum with ten linkage groups using mostly sorghum-derived RFLP probes, and some from maize. This map was based on an interspecific cross (*S. bicolor* BTx623 × *S. propinquum*), mapped in the F_2 generation. That same year, Pereira et al. (1994) also published a linkage map of sorghum with ten groups. A 'composite' map using the genotypic data from two recombinant inbred (RI) populations was published by Dufour et al. (1997) with linkage group designations following those of Pereira et al. (1994). This map contained 199 markers on 13 linkage groups and was later supplemented in subsequent publications with the addition of more RFLP and AFLP markers (Vos et al. 1995), as well as with morphological markers, reducing the number of linkage groups to 11, with two very small unlinked clusters (Boivin et al. 1999). Xu et al. (1994) also published a map of sorghum using RFLP probes primarily derived from sorghum, and some from maize. This map contained 190 markers on 10 major linkage groups, and four smaller ones. This map was based on the genotypes of 50 F_2 plants from a cross between IS3620C and BTx623. Several later studies improved upon this map by addition of more loci. Using 137 RI lines from this same cross, Peng et al. (1999) generated a linkage map containing 323 mapped loci on 10 linkage groups. The total length of this map was 1,347 cM. Bhattramakki et al. (2000) reported the addition of 147 SSR loci to this map using the same RI population, changing the total map length to 1,406 cM. Though these maps were useful tools for mapping of quantitative trait loci (QTL), the lack of agreement between maps from various research groups, as well as relatively poor map quality, made comparison of results with other studies or research groups very difficult. Clearly, there arose a need among the sorghum research community for a consensus map.

More recently, two very dense genetic linkage maps of sorghum have emerged. Menz et al. (2002) added AFLP markers to the IS3620C × BTx623 map of Bhattramakki et al. (2000) to create a very dense linkage map containing 2,926 loci on 10 linkage groups with a total genetic distance of 1,713 cM. Shortly thereafter, using the interspecific cross (*S. bicolor* BTx623 × *S. propinquum*) of Chittenden et al. (1994), another dense linkage map was generated. This map contained 2,512 loci on 10 linkage groups, and is based entirely on RFLP probes (Bowers et al. 2003). Interestingly, the total genetic distance of this map was much shorter than the map by Menz et al. (2002), at only 1,059.2 cM. This discrepancy may be due to the fact that the interspecific population was mapped in the F_2 generation and the intraspecific population was mapped as RI. However, in a recent paper these two maps were aligned by identifying and mapping markers common to both populations (Feltus et al. 2006). A common nomenclature for the 10 chromosomes of sorghum has also been established, and the 10 linkage groups have been assigned to these chromosomes by the use of fluorescent in-situ hybridization (FISH) of sorghum genomic BAC clones containing mapped marker loci (Kim et al. 2002; 2005). The relationships among the various linkage groups reported in previous studies were also reported by Kim et al. (2005), so that these groups can also be assigned to chromosomes and compared to each other. A common nomenclature for sorghum chromosomes and linkage groups should facilitate the cross-validation of QTL studies by different research groups by making direct comparisons of mapped QTL possible.

4. QTL IDENTIFICATION IN SORGHUM

Molecular markers have been used to identify and characterize QTL associated with many different traits in sorghum, including plant height and maturity (Pereira and Lee 1995), traits associated with domestication (Paterson et al. 1995), disease resistance (Gowda et al. 1995), insect resistance (Nagaraj et al. 2005; Deu et al. 2005), and drought tolerance (Tuinstra et al. 1996; 1997). Identification of QTL often leads to further investigations to identify the underlying gene or genes through fine mapping and map-based cloning. When successfully implemented, such studies provide valuable insight into the genetic mechanisms controlling complex, and often economically important, traits. However, from a practical plant breeding standpoint, QTL are usually identified for the purpose of finding linked molecular markers that can be utilized in trait introgression for crop improvement, and often the specific underlying genes are not identified. For the purposes of this review, examples of QTL identification for tolerance to biotic and abiotic stresses important in sorghum are highlighted.

4.1. Identification of QTL for Striga Resistance

Several parasitic plant species of the genus *Striga* are major pests of sorghum in parts of Africa, often causing complete loss of the crop in severe infestations.

Because efforts to control the pest through chemical or cultural means have been met with limited success and are often not practical in poor areas, developing crops with genetic resistance is currently the best strategy for dealing with Striga infestation. However, field resistance to Striga is a complex quantitative trait that has been difficult to address via conventional plant breeding approaches. We have had a Striga resistance breeding program at Purdue University for several years and have employed a mix of biotechnological approaches to address this problem. We have broken Striga resistance down into its component parts using laboratory assays, and have identified several mechanisms by which sorghum can avoid or resist parasitism by Striga, each of which could be manipulated genetically in the development of resistant cultivars.

The most thoroughly studied mechanism is decreased production of germination stimulants by host roots. Striga seeds require a chemical signal from the roots of a potential host in order to germinate. Several compounds have been identified from the exudates of sorghum roots, and from roots of other host and non-host plant species, that have been shown to stimulate germination of Striga seeds in the laboratory (Cook et al. 1966; 1972; Hauck et al. 1992; Siame et al. 1993). Some sorghum varieties produce abnormally low amounts of these compounds, resulting in decreased Striga germination compared to normal varieties (Vogler et al. 1996). Once germinated, Striga also requires the presence of an additional factor to initiate formation of a haustorium, a rootlike structure that penetrates the host root. A sorghum mutant that exhibits decreased levels of haustorial initiation in Striga seedlings has also been reported (Mohamed 2002), but the essential chemical compound has not been identified. A third mechanism, described as an 'incompatible response' has also been reported, in which Striga seedlings attach to host roots, but fail to develop (Mohamed 2002). This mechanism is also not yet clearly understood. Recently Mohamed et al. (2003) described another Striga resistance mechanism, a hypersensitive response (HR) to haustorium invasion, in the roots of certain sorghum varieties, most notably the cultivars 'Framida' and 'Dobbs', and a wild accession (*Sorghum verticilliflorum* (Steud.) Stapf), designated P47121. Hypersensitive response is a common mechanism of resistance to microbial pathogens among diverse plant species, and is characterized by rapid necrosis of the cells surrounding the site of infection (Agrios 1988). In general, the hypersensitive response is also associated with the production of many different compounds, including reactive oxygen species, phytoalexins and other phenolic compounds, and pathogenesis-related proteins. Following the hypersensitive response in sorghum roots, most attached Striga seedlings fail to develop. It has been shown through classical Mendelian genetics that the hypersensitive response in sorghum roots is conferred by two dominant genes, designated *Hrs1* and *Hrs2* (Hypersensitive response to Striga) (Mohamed 2002). However, the functions of these genes are not known, and the underlying physiology of the hypersensitive response to Striga is currently under investigation.

Field-based selection for Striga resistance must be conducted in areas where Striga is endemic, because the pest must be kept under containment elsewhere. Also,

the inherent variability in field-based selection for Striga resistance may further hinder selection efforts. Though highly informative for characterizing specific Striga resistance mechanisms currently available, laboratory assays particularly for post-attachment mechanisms are not amenable to phenotyping a large number of RI lines or breeding material. Thus, the identification of molecular markers for specific Striga resistance mechanisms will facilitate faster introgression and pyramiding of these traits into sorghum breeding lines.

Using a recombinant inbred population, a genetic linkage map was constructed with over 230 DNA markers, an estimated map size of 1,628 cM, and an average interval of 9.5 cM between adjacent loci. The parents of this population were a Striga-resistant caudatum SRN39 and a highly susceptible Chinese kaoliang 'Shan Qui Red' (SQR). Based on laboratory assay data, the locus for the low germination stimulant (*lgs*) production (Vogler et al. 1996) was mapped at 11.8 cM from an RFLP marker *pio200725*-BamH1, and 13.5 cM from *ssr617g*, an inter-simple sequence repeat (ISSR) marker (Ibrahim 1999). We also used the linkage map to place putative QTL for Striga resistance using phenotypic data from field evaluation of this mapping population against *Striga hermonthica* and *Striga asiatica*. Single marker analyses detected six QTL for resistance to *S. hermonthica* and five QTL for resistance to *S. asiatica*. The QTL detected for resistance to *S. hermonthica* accounted for 37% of the variation in resistance, and QTL detected for resistance to *S. asiatica* accounted for 49% of the variation in resistance. Two of these QTL were on the same linkage group as the *lgs* locus. Interval mapping confirmed most of the QTL detected by single marker analysis (Ibrahim 1999). Based on a series of field evaluations of two independent RI populations, Haussmann et al. (2001) produced evidence that other minor loci play a quantitative role in the production of Striga germination stimulants in sorghum, and that different populations of Striga react differently to germination stimulant levels. Another study by this group placed the *lgs* locus on linkage group I (chromosome 6), and a significant QTL for field resistance to Striga also mapped to this region, helping to confirm the location of this QTL (Haussmann et al. 2004). However, the nearest mapped DNA marker in this population is approximately 30 cM away. Saturation of this region of the genome with more molecular markers is needed, so that closer markers can be identified for use in marker-assisted breeding. Several other QTL for field resistance to Striga were also identified in this and another population (Haussmann et al. 2004).

Dissecting Striga resistance into simpler components based on laboratory assays has enhanced efficiency of breeding crop cultivars with improved resistance to the parasite as well as in gene identification. We have mapped the low haustorium initiation factor (*lhf*) locus in an $F_{2:3}$ population derived from a cross between SQR and a wild sorghum accession P78, in which the low haustorium initiation trait was identified. The *lhf* locus maps to linkage group F (chromosome 9), approximately 19.3 cM from SSR marker *txp359* (Mohamed 2002). This locus would be a good candidate for marker-assisted introgression into breeding lines or locally adapted cultivars if a more tightly linked marker were identified. Although Mendelian segregation ratios indicated two loci conferring HR to Striga, only one locus has

been mapped to date. A gene for HR, derived from the wild accession P47121, resides on linkage group B (chromosome 2), 7.5 cM from *txp96* (Mohamed 2002). Though a closer marker would be desirable for marker-assisted introgression, only 7.5 % recombination should be an acceptable margin for error in a Striga resistance breeding program given the high heritability of this trait.

4.2. Identification of QTL for Drought Tolerance

Tolerance to abiotic stress is generally complex and highly influenced by environmental factors. Because of this, selection for improved abiotic stress tolerance is often difficult, and gains from selection are usually quite low. Drought tolerance is one such trait, and is especially important in sorghum, given its exceptional drought tolerance relative to other crops, and its importance in semi-arid parts of the world. The genetic and physiological mechanisms that condition the expression of drought tolerance in sorghum and other crops are still poorly understood. Controlled by many genes and dependent on the timing and severity of moisture stress, drought tolerance has been difficult to study and characterize. Identification of QTL and associated molecular markers, based on carefully managed replicated tests, should facilitate improvement of drought tolerance in sorghum, and also help to increase our understanding of the underlying physiological and genetic mechanisms controlling this complex trait in sorghum, with possible application in other related crops.

A number of physiological and biochemical traits have been hypothesized to enhance drought tolerance. Yet, only a few of these mechanisms have been demonstrated to be causally related to the expression of tolerance to drought under field conditions. As with Striga resistance, we have made a slow but significant progress via empirical breeding of sorghum for drought tolerance by breaking this complex trait into specific phenological stages of plant development, namely early-season (seedling), midseason (pre-flowering), and late-season (post-flowering) drought responses. Using this approach, we have identified sorghum germplasm that are uniquely pre-flowering or post-flowering drought tolerant and a few that combine tolerance at both stages. We have developed new improved drought tolerant sorghum lines in diverse and elite germplasm backgrounds. Our breeding and selection efforts were based on reliable phenotypic markers associated with morphological and yield-related symptoms that occur at pre-flowering and post-flowering stages of plant development as long as moisture stress treatments are induced in a timely manner. Some of these marker traits are simply inherited, while others appear quantitative, rendering them amenable to QTL marker analysis and marker-assisted introgression. Molecular markers linked to QTL for drought tolerance could be used to increase the efficiency of breeding efforts to select sorghum germplasm with enhanced drought tolerance.

We conducted a number of studies using a RI population derived from a cross between Tx7078, which shows pre-flowering drought tolerance, and B35, which shows post-flowering tolerance. A genetic map was constructed using RAPD and RFLP markers, and the RI lines were carefully evaluated for response to drought

in a series of pre-flowering or post-flowering stress environments, both of which were in turn compared to fully irrigated controls. Pre-flowering drought tolerance was assessed by measuring grain yield under drought stress, yield stability (drought versus irrigated), seed set stability, and plant height stability. Single marker analysis identified six QTL associated specifically with pre-flowering drought tolerance. Despite the fact that drought stress was much more severe in the first growing season, QTL by environment effects were relatively small, suggesting that these QTL are stable across varying levels of drought stress or other confounding climatic factors. In addition, some markers showing significant association with yield or seed set under irrigated conditions also showed significance for drought tolerance (Tuinstra et al. 1996). Assessment of post-flowering drought tolerance in sorghum by measuring grain yield under drought stress, yield stability, seed weight stability, and stay-green score was established in a study conducted under carefully managed drought stress conditions (Wanous et al. 1991). The term 'stay-green' refers to delayed senescence which is associated with post-flowering drought tolerance irrespective of the maturity of the genotype. Twelve QTL for post-flowering drought tolerance were identified by single marker analysis (Tuinstra et al. 1997b). Six QTL for stay-green were identified on five linkage groups. QTL on three of these linkage groups were also positively associated with grain yield under irrigated conditions. This indicates that there may be a physiological link between the expression of stay-green under post-flowering drought and grain yield under non-drought conditions.

Because of the importance given to the stay-green trait as a readily visible and heritable drought response trait, several additional studies have been conducted to characterize and map this trait in various populations. Three of these studies (Crasta et al. 1999; Xu et al. 2000; Subudhi et al. 2000) also used B35 as the source of the stay-green trait. In a recent review comparing these studies, Sanchez et al. (2002) named four major stay-green QTL derived from B35 that were consistently identified: *Stg1* and *Stg2* (chromosome 3), *Stg3* (chromosome 2) and *Stg4* (chromosome 5). Two of these, *Stg1* and *Stg3*, correspond to two QTL identified by Tuinstra et al. (1997b). A QTL for stay-green was also identified by Tao et al. (2000) at the same location as *Stg3* in a RI population derived from a cross between QL41 and QL39. QL41 is a stay-green line derived from B35 so this discovery is not surprising, however the other three QTL were not identified in this study, but several additional stay-green loci were found, some of which were associated with QL39. Using the stay-green line SC56, which is unrelated to B35, Kebede et al. (2001) found stay-green QTL at the same locations as *Stg1* and *Stg4*. They also detected a stay-green QTL on chromosome 1, which appears to correspond to a QTL found by Tuinstra et al. (1997b) and by Crasta et al. (1999). These results indicate possible commonalities between stay-green mechanisms from diverse sources. More recently Haussmann et al. (2002) analyzed the stay-green phenotype at various phenological stages by measuring the percentage of leaf area remaining green at weekly intervals (extrapolated to 15 d, 30 d, and 45 d) after flowering during drought conditions in two RI populations. The stay-green source line E36-1, which is unrelated to B35

and SC56, was common to both populations. Five to nine QTL were identified for each of the three time points in each population, and several of these overlapped, suggesting that some aspects of stay-green may be regulated in a stage-specific manner, while others are not. Three QTL derived from E36-1 were common to both populations. One QTL, which was previously identified in other studies with B35 (Tuinstra et al. 1997b; Crasta et al. 1999; Subudhi et al. 2000) was detected on chromosome 2, though it is different from *Stg3*. Most other stay-green QTL found by Haussmann et al. (2002) appear to be unique to E36-1. From the results of multiple studies it is clear that there are several sources of this trait, and likely several mechanisms by which sorghum can remain green and photosynthetically active during post-flowering drought stress. Some of these mechanisms may be common among the germplasm while others seem to be unique to a particular source. Thus, there appears to be great potential for the use of marker-assisted selection to pyramid multiple stay-green genes from these various sources into drought-tolerant breeding lines and cultivars.

4.3. Identification of QTL for Early-Season Cold Tolerance

Cool temperatures during the early growing season are a major limitation to growing sorghum in certain regions of the United States and other temperate areas. Tolerance to early-season cold temperatures is a major trait that is needed in many of the sorghum production areas of the United States. The trait is essential for early-season stand establishment to take advantage of the extended growing season that would ensue if the crop is planted and established early. Increased grain yields have in fact been correlated with improved seedling vigor in sorghum (Chhina and Phul 1987; Cisse 1995). Improved tolerance to early-season cold temperatures would also help to protect the crop from stand reductions caused by unexpected cold periods in late spring. We have found that several sorghum landraces from China exhibit higher emergence and greater seedling vigor under cool conditions than typical American varieties (Cisse and Ejeta 2003; Singh 1985). Unfortunately, Chinese landrace sorghums also possess other undesirable traits, such as susceptibility to most leaf diseases. Directed introgression of seedling cold-tolerance genes from these landraces into elite, high-yielding lines, without the undesirable traits from the donor parents, could be expedited by marker-assisted selection.

To identify markers for seedling cold tolerance, we conducted QTL analyses using a population of RI lines derived from a cross between ‘Shan Qui Red’ (SQR), a cold-tolerant Chinese line, and SRN39, a cold-sensitive caudatum. The RI lines were evaluated for percent germination under cold and optimal conditions in growth chambers, and for percent emergence and seedling vigor scores (Maiti et al. 1981) in the field under early and later plantings. Analysis of the phenotypic data showed a high degree of heritability for all traits measured (Gunaratna 2002), suggesting that gains from selection for seedling cold tolerance should be significant. However, the correlation between percent germination under controlled conditions and percent field emergence was weak, suggesting that seed germination and emergence are

genetically distinct processes. Composite interval mapping (Zeng 1993; Zeng 1994) identified one QTL associated with germination under optimal conditions on linkage group B (chromosome 2) near marker *txp348*. Several QTL associated with percent emergence and seedling vigor in the field after both early and more optimal plantings were also identified. A region of linkage group A (chromosome 1) near *txp43* was significantly associated with percent emergence and seedling vigor after both early and later plantings, suggesting that the same genetic mechanism may play a critical role in both emergence and subsequent early seedling growth under various conditions. Another QTL for percent seedling emergence after early and later plantings was located on group B near *txp211*, indicating that seedling emergence at different temperatures may be controlled by the same genetic mechanism. However, one QTL for seedling vigor was identified on group D (chromosome 4) near *txp51*, which was specific for early planting only, suggesting a temperature-specific gene for vigor. Interestingly, the QTL for emergence and vigor in the field did not overlap with the QTL for germination, supporting our conclusion that germination and emergence are under the control of different genes (Knoll et al. 2007).

5. MARKER-ASSISTED SELECTION AND INTROGRESSION

When genetic markers are found to be associated with a trait of interest, those markers can be exploited to expedite the development of new cultivars through marker-assisted selection (MAS), or to efficiently improve existing cultivars through marker-assisted introgression of traits. The effectiveness of MAS was first demonstrated by Tanksley and Hewitt (1988) in selection for soluble solids and pH in tomato fruit. More recently, Zhou et al. (2003) used SSR markers to select for a major QTL for scab resistance in two different populations of wheat, and achieved a significant reduction in the percentage of infected spikelets in families carrying the QTL. Yousef and Juvik (2002) backcrossed three QTL markers for seedling emergence into three elite sweet corn inbreds, and then tested the effects of the markers in the resulting BC_2F_1 progeny. They found favorable increases in emergence for two of the markers in all three genetic backgrounds, and the third marker showed favorable increases in two of the populations, thus validating the QTL and the utility of MAS for emergence.

Though marker-assisted selection has been shown to be an effective tool, it is not meant to be a substitute for conventional phenotype-based selection, but rather a means to enhance the efficiency of breeding. Molecular markers can help to increase the efficiency of selection for traits that are otherwise difficult to assess, or in traits such as disease resistance when a uniform and optimal level of inoculation is required, or when the trait of interest is recessive, or is expressed late in the life cycle of the plant (Arús and Moreno-González 1993). Molecular markers also allow for environmentally neutral selection, which means that selection can proceed at off-season sites, and in years with unusual weather. However, the greatest potential of marker-assisted selection appears to be in accelerating the rate of gain from selection for desirable genotypes and in the manipulation of QTL that condition

complex economic traits. Effective use of marker-based selection or marker-assisted introgression should significantly decrease the amount of time required by plant breeders to develop new cultivars. For MAS to be effective, the marker and trait should be as tightly linked as possible to minimize recombinations between the marker and the gene of interest. If a tightly linked marker cannot be found, an alternative approach is to select two flanking markers instead. Selection based on molecular markers is particularly useful in the introgression of specific traits into existing cultivars through repeated backcrossing. In addition to selecting for the markers of interest from the donor parent, a breeder can also select for recovery of recurrent parent alleles elsewhere in the genome to hasten recovery of the recurrent genome (Arús and Moreno-González 1993), especially if there are known markers for specific traits in the recurrent parent. However, phenotypic selection is also essential to recovery of the desired characteristics of the recurrent parent, and should not be overlooked.

5.1. Validation of Putative QTL for Marker-Assisted Selection

Identification of QTL associated with important agronomic traits is essential. However, it is only valuable if confirmation is made of the robustness of the markers through repeated evaluation across populations and environments. We recently completed such a validation experiment for an important agronomic trait in sorghum. To test the utility of linked markers for identifying important QTL for early-season cold tolerance in sorghum in varying genetic backgrounds and environments, we chose three closely linked SSR markers (*txp43*, *txp51*, and *txp211*), representing the three most significant QTL for seedling emergence and/or vigor using two segregating populations. Because 'Shan Qui Red' (SQR) was the original source of the cold-tolerance genes, it was used as a parent in both of these populations: Tx2794 × SQR F_2 and Wheatland ms_3× SQR BC_1F_2. Individual F_3 or BC_1F_3 families were genotyped for the three markers, and five representatives of each genotypic class (27 classes total) were randomly chosen for field testing. The field trials were planted about one month earlier than normal planting time for sorghum at West Lafayette, IN in 2004 and 2005. Percent emergence, seedling vigor scores, and stand biomass were measured. Statistical analyses indicated that the SQR allele of *txp43* had favorable effects on seedling vigor in both populations, and on emergence in the Tx2794 population. A large positive effect of the SQR allele of *txp51* was observed in the Tx2794 population for both vigor and emergence. Slight genotype by environment interaction was observed for *txp51* in the Wheatland population. Marker *txp211* had small but significant effects on seedling vigor and emergence in both populations. Various other interactions between loci were also significant, most notably the interaction of *txp43* and *txp51*, which strongly favored the SQR genotypes at both loci for multiple traits. This study (Knoll and Ejeta 2007) validated QTL markers in different genetic backgrounds, and demonstrated the utility of MAS for a quantitative trait, cold tolerance, evaluated in the field. The results also indicate that these markers should be useful tools for future selection

efforts. In fact, *txp43* and *txp51* from SQR are currently being introgressed into various lines in our breeding program in an effort to develop sorghum hybrids with improved early-season performance.

5.2. Development and Analysis of Near-Isogenic Lines Differing for QTL Markers

Although QTL analysis in our drought tolerance studies identified regions of the sorghum genome that condition the expression of drought tolerance, it provided limited information concerning the expression of individual QTL. Analysis of near-isogenic lines (NILs) that differ at QTL can be an effective approach for the detailed mapping and characterization of individual loci. The usual procedure for developing NILs requires multiple generations of tedious backcrossing and phenotypic or marker-assisted selection at each generation to select individuals carrying the locus or loci of interest. Therefore, the use of NILs in analysis of important agronomic traits has been limited because of the time and effort required to develop them. We developed a simpler procedure for developing NILs for any region of the genome that can be analyzed with molecular markers (Tuinstra et al. 1997a). Using the single-seed descent method of plant breeding, a collection of 98 F_2 individuals were advanced to the F_5 generation. At this point, the expected heterozygosity is 6.25%, meaning most loci in the genome are fixed. At this point, eight to ten individual plants were selfed from each F_5 line, and these were in turn maintained as individual lines. The result, at the $F_{5:8}$ generation was a collection of 98 heterogeneous inbred families (HIFs) each consisting of 8 to 10 NILs. Each individual NIL was then genotyped for markers of interest. An advantage of this method over development of NILs by repeated backcrossing is that this is the only generation that must be genotyped. Using families derived from F_5 lines, approximately 6.25% of the families should be segregating for any given marker. The effects of the segregating marker can then be determined by planting NILs in replicated field trials and comparing the phenotypes of the individual NILs within each family.

Using this procedure we developed sets of NILs for four RAPD markers linked to QTL associated with yield under drought environments and other morphological traits associated with drought tolerance. Three HIFs were identified as segregating for each of the four RAPD markers. Within each family, two NILs were chosen for each marker allele. These NILs were planted in replicated field trials at different locations under pre-flowering drought, post-flowering drought, and fully irrigated conditions. Grain yield was measured in all treatments. Seed set, xylem water potential, and osmotic potential were measured in pre-flowering drought treatments. Seed weight and stay-green score were measured in the post-flowering drought treatments. Seed set and seed weight were measured in the irrigated treatments for estimation of the stability of these traits in their respective drought treatments. In most cases, NILs contrasting for a specific locus differed in phenotype as predicted by QTL analysis. Associations between several markers and grain yield in both drought and non-drought conditions were confirmed (Tuinstra et al. 1998). Some

QTL, however, were only confirmed in one environment, for example the stay-green effect of marker *tM5/75* was confirmed in one environment, but was not statistically significant in another. In some instances, new associations were revealed by analysis of NILs. For example, marker *tH19/50* was found to be associated with stay-green in two environments, though this association was not identified in the original QTL analysis. Further analyses indicated that differences in agronomic performance may be associated with effects of heat tolerance, water status, and expression of stay-green, suggesting that these loci mediate the expression of drought tolerance via different biological mechanisms.

6. FUTURE PROSPECTS FOR MAS IN SORGHUM

The recent publication of dense genetic maps for sorghum, the alignment of those maps, the adoption of a common nomenclature for the linkage groups of sorghum, and assignment of linkage groups to physical chromosomes should facilitate the exchange of genetic marker and QTL information among members of the sorghum community. New molecular marker technologies, especially high-throughput methods, will continue to enhance our ability to generate DNA marker data, though accurate measurement of phenotypic traits will continue to be even more important. A new chip-based technology, named Diversity Arrays Technology (DArT), has been developed which will allow rapid detection of large numbers of single nucleotide polymorphisms (SNPs) in many different plant species without the need for extensive prior sequence information (Jaccoud et al. 2001; Kilian et al. 2003). A DArT array is currently available for sorghum (DArT P/L 2006). Although DArT analysis does not require prior genomic sequence information, the sequence of any specific marker locus can be readily determined by sequencing the corresponding DNA clone from the array. This sequence could then be used to create a marker that is more amenable to routine genotyping. This ability to generate and process large amounts of genotypic data may permit large scale association mapping studies. Association mapping is based on the linkage disequilibrium (LD) within natural or assembled populations, and has been used by human geneticists to associate regions of the human genome with various diseases (Collins et al. 2004). The greatest potential use of this technique for plant geneticists and breeders will be the ability to screen populations or collections of germplasm to identify potential QTL and genetic markers for MAS, without using traditional linkage mapping populations. However, there are some disadvantages of this method compared to mapping in experimental populations. First, there is a higher probability of both Type I and Type II errors (Breseghello and Sorrells 2006), especially if the extent of linkage disequilibrium is less than the density of markers used. Thus, a large number of markers should be analyzed to saturate the map to a sufficient density, often to a resolution of less than one centiMorgan (cM). SNP-genotyping microarrays containing thousands of known loci are currently available to human geneticists (Affymetrix, Inc. 2006), but are not yet available for crop plants, though DArT

(Jaccoud et al. 2001; Kilian et al. 2003) may offer an acceptable alternative. Stratification within the assembled collection of germplasm may also affect estimates of association between traits and genotypes, and should be determined prior to association mapping. Mapping traits based on linkage disequilibrium in germplasm collections will be complimentary to traditional QTL mapping based on recombination within experimental populations. The results of one type of study could be used to confirm the results of the other. Though the future possibilities for MAS in sorghum and other crops appear nearly limitless, it has been suggested that the greatest limitation will not be generating genotypic data, but accurately measuring phenotypes. QTL determination and their potential utility in breeding programs are only as good as the quality of the phenotypic data employed in their generation.

REFERENCES

Abu Assar AH, Uptmoor R, Abdelmula AA, Salih M, Ordon F, Friedt W (2005) Genetic variation in sorghum germplasm from Sudan, ICRISAT, and USA assessed by simple sequence repeats (SSRs). Crop Sci 45:1636–1644

Affymetrix, Inc. (2006) http://www.affymetrix.com/index.affx (Cited 12 September 2006)

Agrios GN (1988) Plant pathology, Academic Press, London

Arús P, Moreno-González J (1993) Marker-assisted selection. In: Hayward MD, Bosemark NO, Romagosa J (eds) Plant breeding: principles and prospects, Chapman & Hall, London, pp 314–331

Bhattramakki D, Dong J, Chhabra AK, Hart GE (2000) An integrated SSR and RFLP linkage map of *Sorghum bicolor* (L.) Moench Genome 43:988–1002

Binelli G, Gianfranceschi L, Pe ME, Taramino G, Busso C, Stenhouse J, Ottaviano E (1992) Similarity of maize and sorghum genomes as revealed by maize RFLP probes. Theor Appl Genet 84:10–16

Boivin K, Deu M, Rami J-F, Trouche G, Hamon P (1990) Towards a saturated sorghum map using RFLP and AFLP markers. Theor Appl Genet 98:320–328

Bowers JE, Abbey C, Anderson S, Chang C, Draye X, Hoppe AH, Jessup R, Lemke C, Lennington J, Li ZK, Lin YR, Liu SC, Luo LJ, Marler BS, Ming RG, Mitchell SE, Qiang D, Reischmann K, Schulze SR, Skinner DN, Wang YW, Kresovich S, Schertz KF, Paterson AH (2003) 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 165:367–386

Breseghello F, Sorrells ME (2006) Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. Genetics 172:1165–1177

Chhina BS, Phul PS (1987) Association of seedling vigour with grain yield and nutritional quality in sorghum. Indian J Agric Sci 57:659–660

Chittenden LM, Schertz KF, Lin Y-R, Wing RA, Paterson AH (1994) A detailed RFLP map of *Sorghum bicolor* × *S. propinquum* suitable for high-density mapping, suggests ancestral duplication of *Sorghum* chromosomes or chromosomal segments. Theor Appl Genet 87:925–933

Cisse N (1995) Heritability estimates, genetic correlation, and identification of RAPD markers linked to seedling vigor and associated agronomic traits in sorghum. PhD. Thesis, Purdue University, West Lafayette, IN

Cisse N, Ejeta G (2003) Genetic variation and relationships among seedling vigor traits in sorghum. Crop Sci 43:824–828

Collins A, Lau W, De la Vega FM (2004) Mapping genes for common diseases: the case for genetic (LD) maps. Hum Hered 58:2–9

Cook CE, Whichard LP, Turner B, Wall ME, Egley GH (1966) Germination of witchweed (*Striga lutea* L.): isolation and properties of a potent stimulant. Science 154:1189–1190

Cook CE, Whichard LP, Wall ME, Egley GH, Coggen P, McPhail AT (1972) Germination stimulants 2. The structure of strigol a potent seed germination stimulant for witchweed (*Striga lutea* L.). J Am Chem Soc 94:6198–6199

Crasta OR, Xu WW, Rosenow DT, Mullet J, Nguyen HT (1999) Mapping of post-flowering drought resistance traits in grain sorghum: association between QTLs influencing premature senescence and maturity. Mol Gen Genet 262:579–588

DArT P/L (2006) Diversity arrays technology Pty. Ltd. http://www.diversityarrays.com/index.html (Cited 12 September 2006)

de Oliveira AC, Richter T, Bennetzen JL (1996) Regional and racial specificities in sorghum germplasm assessed with DNA markers. Genome 39:579–587

de Wet JMJ (1978) Systematics and evolution of *Sorghum* Sect., *Sorghum* (Graminae). Am J Bot 65:477–484

Dean RE, Dahlberg JA, Hopkins MS, Mitchell SE, Kresovich S (1999) Genetic redundancy and diversity among 'Orange' accessions in the U.S. national sorghum collection as assessed with simple sequence repeat (SSR) markers. Crop Sci 39:1215–1221

Deu M, Gonzalez-de-Leon D, Glaszmann J-C, Degremont I, Chantereau J, Lanaud C, Hamon P (1994) RFLP diversity in cultivated sorghum in relation to racial differentiation. Theor Appl Genet 88:838–844

Deu M, Ratnadass A, Hamada MA, Noyer JL, Diabate M, Chantereau J (2005) Quantitative trait loci for head-bug resistance in sorghum. Afr J Biotechnol 4:247–250

Dufour P, Deu M, Grivet L, D'Hont A, Paulet F, Bouet A, Lanaud C, Glaszmann JC, Hamon P (1997) Construction of a composite sorghum genome map and comparison with sugarcane, a related complex polyploid. Theor Appl Genet 94:409–418

FAO (2004) Food and Agriculture Organization of the United Nations, Statistics Division. http://www.fao.org/es/ess/index_en.asp (Cited 12 September 2006)

Feltus FA, Hart GE, Schertz KF, Casa AM, Kresovich S, Abraham S, Klein PE, Brown PJ, Paterson AH (2006) Alignment of genetic maps and QTLs between inter- and intra-specific sorghum populations. Theor Appl Genet 112:1295–1305

Gowda PSB, Xu GW, Frederiksen RA, Magill CW (1995) DNA markers for downy mildew resistance genes in sorghum. Genome 38:823–826

Grenier C, Bramel-Cox PJ, Noirot M, Prasada Rao KE, Hamon P (2000a) Assessment of genetic diversity in three subsets constituted from the ICRISAT sorghum collection using random vs. non-random sampling procedures A. Using morpho-agronomical and passport data. Theor Appl Genet 101:190–196

Grenier C, Deu M, Kresovich S, Bramel-Cox PJ, Hamon P (2000b) Assessment of genetic diversity in three subsets constituted from the ICRISAT sorghum collection using random vs. non-random sampling procedures B. Using molecular markers. Theor Appl Genet 101:197–202

Gunaratna N (2002) Early season cold tolerance in sorghum. MS Thesis, Purdue University, West Lafayette, IN

Harlan JR, de Wet JMJ (1972) A simplified classification of cultivated sorghum. Crop Sci 12:172–176

Harlan JR, Stemler A (1978) The races of sorghum in Africa. In: Harlan JR et al (eds) Origins of African plant domestication, Mouton Publishers, The Hague, pp 465–478

Hauck C, Muller S, Schildknecht H (1992) A germination stimulant for parasitic flowering plants from *Sorghum bicolor*, a genuine host plant. J Plant Physiol 139:474–478

Haussmann BIG, Hess DE, Omanya GO, Reddy BVS, Weiz HG, Geiger HH (2001) Major and minor genes for stimulation of Striga hermonthica seed germination in sorghum, and interaction with different striga populations. Crop Sci 41:1507–1512

Haussmann BIG, Mahalakshmi V, Reddy BVS, Seetharama N, Hash CT, Geiger HH (2002) QTL mapping of stay-green in two sorghum recombinant inbred populations. Theor Appl Genet 106:133–142

Haussmann BIG, Hess DE, Omanya GO, Folkertsma RT, Reddy BVS, Kayentao M, Weiz HG, Geiger HH (2004) Genomic regions influencing resistance to the parasitic weed *Striga hermonthica* in two recombinant inbred populations of sorghum. Theor Appl Genet 109:1005–1016

Hulbert SH, Richter TE, Axtell JD, Bennetzen JL (1990) Genetic mapping and characterization of sorghum and related crops by means of maize DNA probes. Proc Natl Acad Sci USA 87:4251–4255

Ibrahim YH (1999) A sorghum linkage map and predicted response to phenotypic and marker selection for resistance to striga in sorghum. PhD Thesis, Purdue University, West Lafayette, IN

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

Kebede H, Subudhi PK, Rosenow DT, Nguyen HT (2001) Quantitative trait loci influencing drought tolerance in grain sorghum (*Sorghum bicolor* L. Moench). Theor Appl Genet 103:266–276

Kilian A, Huttner E, Wenzl P, Jaccoud D, Carling J, Caig V, Evers M, Heller-Uszynska K, Uszynski G, Cayla C, Patarapuwadol S, Xia L, Yang S, Thomson B (2003) The fast and the cheap: SNP and DArT-based whole genome profiling for crop improvement. In: Tuberosa R, Phillips RL, Gale M (eds) Proceedings of the international congress 'In the wake of the double helix: from the green revolution to the gene revolution', Avenue Media, Bologna, Italy, pp 443–461

Kim J-S, Childs KL, Islam-Faridi MN, Menz MA, Klein RR, Klein PE, Price HJ, Mullet JE, Stelly DM (2002) Integrated karyotyping of sorghum by in situ hybridization of landed BACs. Genome 45:402–412

Kim J-S, Klein PE, Klein RR, Price HJ, Mullet JE, Stelly DM (2005) Chromosome identification and nomenclature of *Sorghum bicolor*. Genetics 169:1169–1173

Knoll JE, Ejeta G (2007) Marker-assisted selection for early-season cold tolerance in sorghum: QTL validation across populations and environments. Manuscript in preparation.

Knoll JE, Gunaratna N, Ejeta G (2007) QTL analysis of early-season cold tolerance in sorghum. Manuscript in preparation.

Maiti RK, Raju PS, Bidinger FR (1981) Evaluation of visual scoring for seedling vigor in sorghum. Seed Sci Technol 9:613–622

Melake-Berhan A, Hulbert SH, Butler LG, Bennetzen JL (1993) Structure and evolution of the genomes of *Sorghum bicolor* and *Zea mays*. Theor Appl Genet 86:598–604

Menkir A, Goldsbrough P, Ejeta G (1997) RAPD based assessment of genetic diversity in cultivated races of sorghum. Crop Sci 37:564–569

Menz MA, Klein RR, Mullet JE, Obert JA, Unruh NC, Klein PE (2002) A high-density genetic map of *Sorghum bicolor* (L.) Moench based on 2926 AFLP, RFLP and SSR markers. Plant Mol Biol 48:483–499

Mohamed A (2002) Identification and characterization of genetic variants in sorghum for specific mechanisms of *Striga* resistance. PhD Thesis, Purdue University, West Lafayette, IN

Mohamed A, Ellicott A, Housley TL, Ejeta G (2003) Hypersensitive response to *Striga* infection in *Sorghum*. Crop Sci 43:1320–1324

Nagaraj N, Reese JC, Tuinstra MR, Smith CM, Amand PS, Kirkham MB, Kofoid KD, Campbell LR, Wilde GE (2005) Molecular mapping of sorghum genes expressing tolerance to damage by greenbug (Homoptera: Aphididae). J Econ Entomol 98:595–602

Paterson AH, Lin Y, Li Z, Schertz KF, Doebley JF, Pinson SRM, Liu S, Stanzel JW, Irvine JE (1995) Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. Science 269:1714–1718

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

Pereira MG, Lee M (1995) Identification of genomic regions affecting plant height in sorghum and maize. Theor Appl Genet 90:380–388

Pereira MG, Lee M, Bramel-Cox P, Woodman W, Doebley J, Whitkus R (1994) Construction of an RFLP map in sorghum and comparative mapping in maize. Genome 37:236–243

Sanchez AC, Subudhi PK, Rosenow DT, Nguyen HT (2002) Mapping QTLs associated with drought resistance in sorghum (*Sorghum bicolor* L. Moench). Plant Mol Biol 48:713–726

Siame BA, Weerasuriya Y, Wood K, Ejeta G, Butler LG (1993) Isolation of strigol, a germination stimulant for *Striga asiatica* from host plants. J Agric Food Chem 41:1486–1491

Singh SP (1985) Sources of cold tolerance in grain sorghum. Can J Plant Sci 65:251–257

Soller M, Beckmann JS (1983) Genetic polymorphism in varietal identification and genetic improvement. Theor Appl Genet 67:25–33

Subudhi PK, Rosenow DT, Nguyen HT (2000) Quantitative trait loci for the stay-green trait in sorghum (*Sorghum bicolor* L. Moench): consistency across genetic backgrounds and environments. Theor Appl Genet 101:733–741

Tanksley SD, Hewitt J (1988) Use of molecular markers in breeding for soluble solids content in tomato – a re-examination. Theor Appl Genet 75:811–823

Tao YZ, Henzell RG, Jordan DR, Butler DG, Kelly AM, McIntyre CL (2000) Identification of genomic regions associated with stay-green in sorghum by testing RILs in multiple environments. Theor Appl Genet 100:1225–1232

Tuinstra MR, Ejeta G, Goldsbrough PB (1997a) Heterogeneous inbred family (HIF) analysis: an approach for developing near-isogenic lines that differ at quantitative trait loci. Theor Appl Genet 95:1005–1011

Tuinstra MR, Grote EM, Goldsbrough PB, Ejeta G (1997b) Genetic analysis of post-flowering drought tolerance and components of grain development in sorghum. Mol Breed 3:439–448

Tuinstra MR, Ejeta G, Goldsbrough PB (1998) Evaluation of near-isogenic sorghum lines contrasting for QTL markers associated with drought tolerance. Crop Sci 38:835–842

Tuinstra MR, Grote EM, Goldsbrough PB, Ejeta G (1996) Identification of quantitative trait loci associated with pre-flowering drought tolerance in sorghum. Crop Sci 36:1337–1344

Vogler RK, Ejeta G, Butler LG (1996) Inheritance of low production of *Striga* germination stimulant in sorghum. Crop Sci 36:1185–1191

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucl Acids Res 23:4407–4414

Wanous MK, Miller FR, Rosenow DT (1991) Evaluation of visual rating scales for green leaf retention in sorghum. Crop Sci 31:1691–1694

Whitkus R, Doebley J, Lee M (1992) Comparative genome mapping of sorghum and maize. Genetics 132:1119–1130

Xu GW, Magill CW, Schertz KF, Hart GE (1994) A RFLP linkage map of *Sorghum bicolor* (L.) Moench. Theor Appl Genet 89:139–145

Xu W, Subudhi PK, Crasta OR, Rosenow DT, Mullet J, Nguyen HT (2000) Molecular mapping of QTLs conferring stay-green in grain sorghum (*Sorghum bicolor* L. Moench). Genome 43:461–469

Yang W, de Oliveira AC, Godwin I, Schertz K, Bennetzen JL (1996) Comparison of DNA marker technologies in characterizing plant genome diversity: variability in Chinese sorghums. Crop Sci 36:1669–1676

Yousef GG, Juvik JA (2002) Enhancement of seedling emergence in sweet corn by marker-assisted backcrossing of beneficial QTL. Crop Sci 42:96–104

Zeng Z (1993) Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. Proc Natl Acad Sci USA 90:10972–10976

Zeng Z (1994) Precision mapping of quantitative trait loci. Genetics 136:1457–1468

Zhou W-C, Kolb FL, Bai G-H, Domier LL, Boze LK, Smith NJ (2003) Validation of a major QTL for scab resistance with SSR markers and use of marker-assisted selection in wheat. Plant Breed 122:40–46

CHAPTER 10

MOLECULAR GENETICS AND BREEDING OF GRAIN LEGUME CROPS FOR THE SEMI-ARID TROPICS

RAJEEV K. VARSHNEY[1,*], DAVID A. HOISINGTON[1], HARI D. UPADHYAYA[1], POORAN M. GAUR[1], SHYAM N. NIGAM[1], KULBHUSHAN SAXENA[1], VINCENT VADEZ[1], NIROJ K. SETHY[2,3], SABHYATA BHATIA[2], RUPAKULA ARUNA[1], M. V. CHANNABYRE GOWDA[4] AND NAGENDRA K. SINGH[5]

[1]*International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru-502 324, India*

[2] *National Institute for Plant Genome Research (NIPGR), Aruna Asaf Ali Marg, New Delhi-110 067, India*

[3]*Present address: Defence Institute of Physiology and Allied Sciences (DIPAS), Defence Research and Development Organization (DRDO), Timarpur, Delhi-110 054, India*

[4]*University of Agricultural Sciences (UAS), Dharwad-500 006, India*

[5]*National Research Centre on Plant Biotechnology (NRCPB), Indian Agricultural Research Institute (IARI), New Delhi-110 012, India*

Abstract: Grain legumes are important crops for providing key components in the diets of resource-poor people of the semi-arid tropic (SAT) regions of the world. Although there are several grain legume crops grown in SAT, the present chapter deals with three important legumes i.e. groundnut or peanut (*Arachis hypogaea*), chickpea (*Cicer arietinum*) and pigeonpea (*Cajanus cajan*). Production of these legume crops are challenged by serious abiotic stresses e.g. drought, salinity as well as several fungal, viral and nematode diseases. To tackle these constraints through molecular breeding, some efforts have been initiated to develop genomic resources e.g. molecular markers, molecular genetic maps, expressed sequence tags (ESTs), macro-/micro- arrays, bacterial artificial chromosomes (BACs), etc. These genomic resources together with recently developed genetic and genomics strategies e.g. functional molecular markers, linkage-disequilibrium (LD) based association mapping, functional and comparative genomics offer the possibility of accelerating molecular breeding for abiotic and biotic stress tolerances in the legume crops. However, low level of polymorphism present in the cultivated genepools of these legume crops, imprecise phenotyping of the germplasm and the higher costs of development and application of genomic tools are critical factors in utilizing genomics in breeding of these legume crops.

*Author for correspondence: r.k.varshney@cgiar.org

R.K. Varshney and R. Tuberosa (eds.), Genomics Assisted Crop Improvement: Vol. 2: Genomics Applications in Crops, 207–241.

1. INTRODUCTION

1.1. Importance of Legume Crops

Grain and forage legumes are grown on some 180 million hectares, or 12% to 15% of the Earth's arable surface (source: FAO Database [http://apps.fao.org/page/collections]). They account for 27% of the world's primary crop production, with grain legumes alone contributing 33% of the dietary protein nitrogen needs of humans (Vance et al., 2000). Grain legumes are key components in the diets of resource-poor people in the developing world; especially those who are vegetarian because of choice or cannot afford to supplement their diets with meat. Grain legumes are a rich source of essential vitamins, minerals, and important amino acids like lysine (Duranti and Gius, 1997; Grusak, 2002). Last but not least, grain legumes can also contribute to the Nitrogen balance of soils where they are grown, making them an indispensable component of the sustainability of the system. Another attractive feature is their ability to fix atmospheric nitrogen in the soil by virtue of their symbiotic association with *Rhizobium* bacteria (Schulze and Kondorsi, 1998; Serraj, 2004), thus reducing the need for N-fertilizers in the cropping systems. Legumes often attract higher market prices than other staple crops, making them an important source of income for farmers.

Legumes belong to the taxonomic family *Fabaceae*, containing over 18,000 species divided into the three sub-families Mimosoideae, Caesalpinoideae and Papilionoideae. Legume species have been cultivated for millennia all over the world because of the nutritional value of their seeds as mentioned above. Among different legumes, soybean (*Glycine max* L.) is the major single contributing species, which is used for multiple applications in the food and feed industries. Others, such as chickpea (*Cicer arietinum* L.), common bean (*Phaseolus vulgaris* L.), groundnut or peanut (*Arachis hypogaea* L.), cowpea (*Vigna unguiculata* L.) and pigeonpea (*Cajanus cajan* L.) contribute significantly to the diets of large numbers of people in Asia, Africa, and South America. The high nutritional value of legumes is achieved by the presence of a wealth of secondary metabolites and in the capacity of legumes to establish a symbiosis with the soil bacteria *Rhizobium*, which supplies nitrogen to the plant in exchange of carbohydrate supply to the microsymbiont (Dixon and Sumner, 2003, Desbrosses et al., 2005). The symbiosis results in the formation of root outgrowth called nodules, which can have different types of shape depending on plants. That symbiosis gets preferentially established under low N conditions, and gets inhibited under excess nitrogen, although certain species are able to obtain most of their nitrogen from the symbiosis in environments that do contain nitrogen. Nodules host the *Rhizobium* bacteria, which differentiate in the nodules into symbiotic bacteroids, and are the site of catalysis of dinitrogen into ammonia by the microbial enzyme nitrogenase. As an energy source to achieve N fixation, the bacteria obtain dicarboxylic acids from the host plant. By a complex amino-acid cycle the reduced nitrogen is provided to the plant (Lodwig et al., 2003) where it is accumulated into proteins. Thus legumes can also help replenish nutrient-depleted soil.

1.2. Legume Crops in the Semi-Arid Tropics

The semi-arid tropics (SAT) covers parts of 55 developing countries where the 75–180 day growing period has a mean daily temperature of more than 20°C. The dry semi-arid tropics have very short growing seasons, separated by very hot and dry periods in which growth without irrigation or stored soil moisture is impossible. Natural soil fertility is often low, in part because soils are highly weathered by the dry-hot and humid-hot cycles, and pest and disease pressure can be intense. Farmers face further substantive risks, even within the growing season, as there are irregular periods of drought and high evaporative demand which can seriously compromise crop productivity. Based on 1996 statistics, the SAT is home to about 1.4 billion people, of which 560 million (40%) are classified as poor, and 70% of the poor reside in rural areas (Ryan and Spencer 2001).

Although a number of crops are grown in SAT areas, among legume crops, chickpea, groundnut common bean, cowpea and pigeonpea provide key components in the diets of resource-poor people in the developing world. We, at ICRISAT, together with our National Agricultural Research System (NARS) partners are engaged on crop improvement in chickpea, groundnut and pigeonpea, therefore in this article we discuss the advances in the area of genetics and genomics applied to breeding in only these three legume crops. In the first instance, a brief introduction of these crops is given in following sections.

1.2.1. Chickpea (Cicer arietinum L.)

Chickpea is the third most important grain legume globally, and second in importance in Asia. It is also an important legume crop in Eastern and Southern Africa. About 90% of the global area and 88% of production is concentrated in Asia. Chickpea has one of the best nutritional compositions of any dry edible legume, and is mainly used for human consumption. The *desi* type (colored seed coat) is usually de-hulled and split to make *dhal* or flour (*besan*), while *kabuli* types (white or cream-colored seed coat) is often cooked as whole grain. The haulms are used for animal feed. Chickpea improves soil fertility through nitrogen fixation (up to 140 kg N/ha). Chickpea area has slightly decreased globally, but has been stable at 9 M ha in Asia for the past 25 years. However, production in Asia has increased by 39% due to a 32% increase in productivity. Even then, the current average yield in Asia (0.8 t/ha) is low, and far below the potential yield (5 t/ha), or research station yields (3.5 t/ha). The global demand for chickpea in 2010 is estimated at 11.1 Mt (up from the current 8.6 Mt). A combination of productivity enhancement through crop improvement enhanced with biotechnological tools, integrated crop management and expansion of area to new niches and production systems are needed to achieve this target.

According to van der Maesen (1987), the cultivated chickpea has been taxonomically placed in the genus *Cicer*, which belongs to the family *Fabaceae* and its monogeneric tribe *Cicereae* Alef. Presently, the genus consists of 43 species divided into 4 sections, namely *Monocicer*, *Chamaecicer*, *Polycicer* and *Acanthocicer*.

This classification is based on their morphological characteristics, lifestyle and geographical distribution (van der Maesen, 1987). Eight of these *Cicer* species share the annual growth habit with chickpea are of particular interest to breeders.

1.2.2. Groundnut (Arachis hypogaea L.)

Groundnut is an important food and cash crop for the resource-poor farmers in Asia and Africa. It is primarily grown for edible oil (48–50%) as well as for direct consumption as food by people. Groundnut haulms are excellent fodder for cattle, and groundnut cake (after oil extraction) is used as animal feed. It contributes significantly to household food security and cash income through the sale of groundnut products. Groundnut productivity in Western and Central Africa (WCA) and Eastern and Southern Africa (ESA) is below the world average yield of 1.4 t/ha. Although groundnut productivity in Asia (1.8 t/ha) exceeds the world average, it is still lower than the yields in developed countries (3 t/ha). The area under groundnut in ESA has increased dramatically from 2.3 to 3.3 M ha during 2000 to 2004. In Asia, the area under groundnut is increasing in China and Vietnam, but is declining in India during 1991–2004. There has been a slight decline in area in WCA. Although global productivity has shown a positive trend, much more needs to be achieved in future.

The genus *Arachis* belongs to the family *Fabaceae*, subfamily Papillionaceae, tribe Aeschynomenae, subtribe Stylosantheae. Cultivated groundnut (*Arachis hypogaea* L.) can be botanically classified into two subspecies, *hypogaea* and *fastigiata* that are distinguished based on branching pattern and distribution of vegetative and reproductive nodes along lateral branches. Each subspecies is again divided into two botanical varieties; subsp. *hypogaea* into var. *hypogaea* (virginia) and var. *hirsuta* and subsp. *fastigiata* into var. *fastigiata* (valencia), var. *vulgaris* (spanish), var. *peruviana* and var. *aequatoriana* (Karpovickas and Gregory, 1994).

1.2.3. Pigeonpea (Cajanus cajan L.)

Pigeonpea is a versatile and multipurpose crop. It is one of the major food legumes in the tropical and sub-tropical regions in Africa, Asia and the Caribbean countries. Its green pods and seeds are consumed as a vegetable, and the dry grains are cooked whole or after dehulling (as *dhal*). The foliage is used as fodder, and the dry sticks are used for fencing, thatching, and as firewood. It fixes atmospheric nitrogen, and the extensive leaf fall adds organic matter to the soil. Dry grain is also used for animal feed. About 90% of the global pigeonpea area (4.4 M ha) is in Asia (about 86% in India). Other major countries where pigeonpea is grown are Myanmar, Nepal, Bangladesh, Pakistan and China. In Sub-Saharan Africa (SSA), pigeonpea is grown in Uganda, Kenya, Malawi, Mozambique, Zimbabwe, Zambia, South Africa, Sudan and Ethiopia; but reliable statistics are not available. Pigeonpea production has shown only a marginal increase during the past two decades (2.2 to 2.9 million t during 1980–98). However, productivity has remained stagnant at 0.7 t/ha, mostly because it is intercropped with cereals or cotton and receives no or little inputs; or

Table 1. Characteristics and genomics data available for some SAT legumes

	Chickpea	Groundnut	Pigeonpea
Species name	*Cicer arietinum* L.	*Arachis hypogaea* L.	*Cajanus cajan* L.
Ploidy level and chromosome number	2n = 2*x* = 16	2n = 4*x* = 40	2n = 2*x* = 22
Genome size[1]	931 Mbp	2891 Mbp	858 Mbp
SSR markers	~700 (Winter et al., 1999; Huettel et al., 1999; Sethy et al., 2003,2006b; Lichtenzveig et al., 2005; Choudhary et al., 2006; Varshney et al., unpublished; Bhatia et al., unpublished results)	~700 (Hopkins et al., 1999; He et al., 2003; Ferguson et al., 2004; Moretzsohn et al., 2004; Palmieri et al., 2005; Mace et al., unpublished; D. Bertioli, Brazil, pers. commun.; S. Knapp, pers. commun.)	~100 (Burns et al., 2001; Odoney et al. 2007)
BAC libraries	3.8 X (Rajesh et al. 2004), 7 X (Lichtenzveig et al. 2005)	6.5–9.0 X (Yuksel and Paterson, 2005)	–
ESTs	~2000 (NCBI, Buhariwalla et al., 2005)	~7538 (NCBI, Luo et al., 2005; S. Knapp, pers. commun.)	More than 884 (NCBI) (Gaikwad et al. unpublished
Gene arrays	768- features microarray (Coram and Pang, 2005a), SAGE Gene Chip (P. Winter, Germany, pers. commun.)	400 unigene array (Luo et al., 2005)	

[1] As per estimate of Royal Botanic Gardens, Kew, UK (http://www.rbgkew.org.uk/cval/)
[2] NCBI = http://www.ncbi.nlm.nih.gov/

gets relegated to marginal and poor soils, often where no other crop can be grown. Additionally, pigeonpea has also generally a poor harvest index.

Pigeonpea belongs to the *Cajaninae* sub-tribe of the economically important leguminous tribe *Phaseoleae* that contains soybean (*Glycine max* L.), common bean (*Phaseolus vulgaris*) L.) and mungbean (*Vigna radiata* L.) (Young et al., 2003). The genus *Cajanus* comprises 32 species most of which are found in India and Australia although one is native to West Africa. Pigeonpea is the only cultivated food crop of the *Cajaninae* sub-tribe and has a diploid genome.

A brief overview on genome size, ploidy level, existing genomics resources in chickpea, groundnut and pigeonpea is given in Table 1.

2. CHALLENGES IN SUSTAINABLE CROP PRODUCTION OF SAT LEGUMES

2.1. Abiotic Stresses

Abiotic stresses severely limit agricultural production. There is a clear consensus that drought is among the most severe stress for legume production in SAT regions

of Asia and Africa while salinity is the second ranked constraint in the production of these legumes in some Asian countries.

2.1.1. Drought

The SAT regions are characterized by short and erratic rainfall (and then long periods with virtually no rain), where crops grown under rainfed conditions suffer from both intermittent and terminal drought stress, and crop grown in residual moisture after the rain suffer terminal drought, thus incurring major yield losses. Water deficit is one of the most severe stresses for sustainable crop production. Worldwide, yield losses each year due to drought are estimated to be around US$500 million (Sharma and Lavanya, 2002).

Water capture by roots and water-use efficiency are probably two important components of the yield architecture, as defined by Passioura (1977) that are important for crops growing under terminal drought conditions. These two traits are the classical component of what is called ‘drought avoidance’, and which means getting more water or using it more efficiently). Drought avoidance is considered to be the major trait of interest to expand production to presently uncropped areas and post-rainy fallows in SAT regions. Although roots have already proved to be beneficial for yield under terminal drought (chickpea, Kashiwagi et al., 2004), there is a need to understand better how root traits contribute to drought avoidance, and a need to explore them in those crops where little information on roots has been acquired (e.g., groundnut). Specifically, there is a need to understand the dynamics of roots, how roots contribute to the overall water budget, and more interestingly how they contribute at the time of grain filling, and how they contribute at the time of flowering. Recent studies at ICRISAT indicate that deeper rooting correlates with a higher harvest index (HI) in chickpea in conditions of more severe drought (Kashiwagi et al., 2004, 2006). This might be related to the root being able to supply water during flowering and allowing less flower drop because of water deficit. Water use efficiency (WUE) or more specifically transpiration efficiency (TE) is another trait that is being addressed in groundnut at ICRISAT by using different biotechnological, physiological and breeding approaches. For TE, there is also a need to understand better the mechanisms that lead to better TE, if we ever want to reach the genes involved.

2.1.2. Salinity

Soil salinity is an important limiting factor for crop yield improvement, which affects 5–7% of arable lands, i.e. approximately 77 million ha worldwide. Most crops are sensitive to salt stress at all stages of plant development, including seed germination, vegetative growth and reproductive growth, although the latter stage is certainly the most sensitive across many crops. Legumes, in general, are sensitive to salinity, and within legumes, chickpea, fababean and pigeonpea are more sensitive than other food legumes. The salinity problem is increasing, in particular in areas where irrigation is a common practice (Ghassemi et al., 1995). Though management options exist to alleviate salt effects, these are often in contradiction with the

immediate economic choices of the concerned farmers; thus crop improvement for salt tolerance appears to be the best and economic alternative.

The problem of salinity is basically two-fold. In one case, soil is saturated with sodium (Na) and soil pH remains within an optimal range for crop growth. This type of salinity refers to coastal or dryland salinity. These are soils that get saturated with sodium because an existing saline ground water table rising (proximity to the sea or salt that has accumulated in the soil profile), bringing the salt to the surface. In a second case, soil is both saturated with Na (exchangeable sodium percentage, ESP, > 6) and pH has reached levels above 8.5–9.0. This type of salinity is also called transient salinity, and is thereafter referred to as *sodicity* or *sodic soils*. In this case, the sodium saturation brings about the same effect as salinity, but the high pH dramatically affects the availability of micronutrients (low availability/solubility of micronutrient salts at these pH levels), the soil structure and porosity (poor drainage, tendency for water logging, little oxygenation because of saturation of the exchange complexes in the soil by sodium). In the past, most studies have focused on *salinity*, and only a few on *sodicity*.

Despite the importance of salinity in crop production worldwide and the abundant knowledge on the effect of salinity on plant growth and development, there has surprisingly been little effort to breed for improved salinity tolerance, with the exceptions of wheat, rice, barley, alfalfa and claims of soybean. Breeding tolerant crop varieties is therefore urgently needed.

2.2. Biotic Stresses

The major biotic factors of SAT legumes are diseases and insect pests. The chickpea diseases of major importance are ascochyta blight (caused by the necrotrophic fungus *Ascochyta rabiei* (Pass.) Lab.), fusarium wilt (caused by *Fusarium oxysporum* f. sp. *ciceris*), Botrytis gray mold and root rots caused by *Sclerotium* and *Pythium.* Majority of these diseases affect all aerial parts of the plant. Among the pigeonpea diseases, sterility mosaic (viral disease), fusarium wilt (caused by the fungus *Fusarium udum* Butler), and phythophthora blight (*Phytophthora drechsleri*) are major diseases causing significant losses of pigeonpea yield. In groundnut, rust, late leaf spot, and early leaf spot are serious diseases worldwide, which cause 50–60% pod yield loss. Rust and late-leaf spot often occur together and the pod yield loss can exceed 70% in the crop. Besides adversely affecting pod yield and its quality, these foliar diseases also affect haulm (fodder) yield and quality. Whereas the level of resistance available in cultivated groundnut to rust is very high, for early- and late-leaf spot, it is low. Wild *Arachis* species harbour many useful resistance genes against various diseases and insect pests. Of the important biotic constraints specific to sub-Saharan Africa (SSA), the groundnut rosette disease (GRD), vectored by aphids, is endemic to the continent and its adjoining islands and epidemics occur often throughout SSA, reducing groundnut production and crippling rural food security.

More than 200 species of insects feed on pigeonpea and chickpea, of which pod borer (*Helicoverpa armigera*), spotted pod borer (*Maruca vitrata*), pod fly (*Melanagromyza obtuse*), pod sucking bugs (*Clavigralla* spp., *Nezara viridula*) and the bruchid (*Callosobruchus* spp.) are most important economically (Singh et al., 1990). *Helicoverpa* causes an estimated loss of US$ 317 million in chickpea and pigeonpea (ICRISAT, 1992), and possibly over US$ 2 billion on other crops worldwide. A conservative estimate is that over US$ 1 billion is spent on insecticides to control this pest. Therefore, in addition to the huge economic losses caused directly by the pest, there are several indirect costs from the deleterious effects of pesticides on the environment and human health (Sharma, 2001). These insect pests feed on various plant parts such as leaves, tender shoots, flower buds, and immature seeds. It has been difficult to breed for *Helicoverpa* resistance in chickpea and pigeonpea because sources with a high level of resistance are not available in the cultivated species of these legumes. Recent studies show potential of utilizing the wild species in insect pest resistance breeding programme as these have shown higher levels of resistance.

3. UTILIZATION OF PLANT GENETIC RESOURCES (PGRS)

Availability and characterization of suitable germplasm is a critical factor for utilizing genetic variation in crop breeding. Fortunately for all the three legume species mentioned in the article, a large number of accessions are present in different genebanks throughout the world (Dwivedi et al., 2006). For instance, ICRISAT, under an agreement with FAO, holds 16,853 cultivated and 117 wild accessions of *Cicer* species, whereas the International Centre for Agricultural Research in Dryland Areas (ICARDA), Syria, under the same FAO agreement, maintains 8,342 cultivated and 255 wild accessions. Other institutions holding chickpea germplasm are the National Bureau of Plant Genetics Resource (NBPGR), India (14,566 accessions); Centre for Legume Improvement in Mediterranean Area (CLIMA) (4,351 accessions) and AusPGRIS (7922 accessions) in Australia; United States Department of Agriculture (USDA), USA (4,662 accessions); and the Seed and Plant Improvement Institute, Iran (4,925 accessions). The European *Cicer* database contains 3,700 cultivated accessions from 11 countries (Pereira et al. 2001). For groundnut, ICRISAT holds, under the same agreement with FAO, 14,126 accessions of cultivated peanut and 293 accessions of wild *Arachis* species from 93 countries. Other institutions holding large numbers of peanut accessions are the National Research Centre for Groundnut (NRCG), India (7,935 accessions) and the USDA Southern Regional Plant Introduction Station, USA (6,233 accessions). In the United States, wild *Arachis* species are maintained at North Carolina State University, Raleigh (250 accessions) and at the Texas Agricultural Experiment Station (TAMU), Texas (300 accessions). For pigeonpea, ICRISAT holds under the agreement with FAO 12,398 pigeonpea accessions of cultivated and 314 accessions of wild species from 74 countries. Other institutions holding substantial amounts of pigeonpea germplasm

include the NBPGR (5,454 accessions) in India and the USDA, Southern Regional Plant Introduction Station (4,116) in USA.

3.1. Core and Mini-Core Collections

Despite the availability of a large number of germplasm, only limited numbers of accessions have been used in breeding programme not only in SAT legumes but other crop species as well (Dwivedi et al., 2006). One of the main reasons for this fact may be the large sizes as well as non-availability of information on germplasm collections. Core collections present a manageable and cost-effective entry point into germplasm collections for identifying parental genotypes with new sources of disease and pest resistance or abiotic stress tolerance. Evaluation of core collections is usually the most efficient and reliable means of carrying out an initial search of the germplasm collections. For instance, early evaluation of limited number of germplasm accessions led to premature conclusion that no variability for salinity tolerance existed in chickpea (Saxena, 1984). However, recent screening of large number of germplasm accessions, including the chickpea mini-core collection, revealed very large variation, readily usable for breeding purposes (Vadez et al., 2006). Evaluation of larger amounts of germplasm through multi-location trials is both very expensive and time consuming; large-scale generation of accurate and precise evaluation data from such trials is generally not possible, thus dramatically reducing the probability of identifying desirable material. Core collections usually consist 10% of the entire germplasm collection that represents the collections variability (Brown, 1989). These representative subsample collections are developed from the entire collection, using all available information on accessions including the origin and geographical distribution plus characterization and evaluation data. Ten percent of most crop germplasm collections are a much more feasible amount of material for intensive and precise evaluation.

Most core collections have been designed from global or regional collections held within international agricultural research centers or national program gene banks, while a few have also been developed for wild accessions (Tohme et al., 1996). After evaluating a total of 16,991 chickpea accessions for 13 traits and 14,310 groundnut and 12,153 accessions of pigeonpea for 14 traits each, the core collections of chickpea, groundnut and pigeonpea with 1,956 (Upadhyaya et al., 2001a), 1,704 (Upadhyaya et al., 2003) and 1,290 accessions (Reddy et al., 2005), respectively have been developed at ICRISAT. In addition, the core collection of 505 genotypes of chickpea was developed after analysis of 3,315 genotypes (Hannan et al., 1994). Similarly for groundnut, an USDA core collection with 831 genotypes after evaluating 7,432 accessions for 24 traits (Holbrook et al., 1993) and an Asian core collection based on evaluating 4,738 genotypes for 15 traits (Upadhyaya et al., 2002) are available. Although these core collections have been useful for identifying diverse sources for traits of interests and broadening the genetic base of cultivars for a crop (Upadhyaya et al., 2001b, 2006a; Krishnamurthy et al. 2003;

Serraj et al., 2004), even a core collection can be too large so a further reduction is also valuable providing it is not associated with losing too much of the spectrum of diversity. Upadhyaya and Ortiz (2001) developed a strategy for sub-sampling a core collection to develop a mini-core collection, based on selecting 10% of the core accessions representing the variability of larger collection of species. In this process, the core collection is evaluated for various morphological, agronomic, and quality traits to select a 10% subset from this core subset (i.e., 1% of the entire collection) that captures a large proportion (i.e. more than 80% of the entire collection) of the useful variation. Selection of core and mini-core collections is based on standard clustering procedures used to separate groups of similar accessions combined with various statistical tests to identify the best representatives. The mini-core collection developed at ICRISAT for chickpea consisted of 211 accessions (Upadhyaya and Oritz, 2001), while the groundnut (Upadhyaya et al., 2002) and pigeonpea (Upadhyaya et al., 2006b) mini-core consists of 184 accessions and 146 accessions, respectively. Both core or mini-core germplasm collections have been used for identifying a range of germplasm with beneficial traits for use in breeding programs (see Dwivedi et al., 2006 for references). Increasing concern of trade and food processors for consistent and better quality and physical specifications, however, suggest further characterization of core or mini-core collections for quality and market traits.

3.2. Molecular Characterization of PGRs

The core or mini core collections have been developed based on morphological or agronomic traits; little information is available on molecular genetic diversity present in the germplasm collection. Molecular characterization of germplasm is a particularly useful tool for assisting genebank curators to better manage genetic resources, helping them to identify redundant germplasm and to provide scientists with the most diverse germplasm for applications in research and breeding (Bretting and Widrlechner, 1995; Virk et al., 1995; Brown and Kresovich, 1996; van Treuren et al., 2001; Upadhyaya et al., 2006b). Accessions with the most distinct DNA profiles are likely to contain the greatest number of novel alleles (Tanksley and McCouch, 1997). As a part of the Generation Challenge Programme (GCP) of the CGIAR, molecular characterization of global composite collections of the SAT legumes is in progress at ICRISAT. For example, genotyping of about 3000 chickpea accessions (Upadhyaya et al., 2006a) with 50 SSR markers and 1000 groundnut accessions with 20 SSR markers, in collaboration with ICARDA (Syria) and EMBRAPA (Brazil) respectively has already been completed. Molecular characterization of 1000 pigeonpea accessions at 20 SSR loci is in progress. These studies provide estimates on genetic diversity and the population structure of the germplasm that can be used to define the most diverse collection, called ‘reference collection’ for using in association mapping studies (see later).

4. MOLECULAR BREEDING FOR SAT LEGUMES

Legume breeders have made major contributions to combat the problem of both abiotic and biotic stresses in the past but the pace and extent of improvements must be dramatically increased to attend to parallel demands. Recent advances in the area of biotechnology have offered the tools in the form of molecular markers to assist the breeding practices (Jain et al., 2002). Molecular markers are powerful diagnostic tools that detect DNA polymorphism both at the level of specific loci and at the whole genome level (reviewed by Azhaguvel et al., 2006). As compared to morphological traits/markers, molecular markers have several advantages as they are phenotypically neutral and are not influenced by pleiotropic and epistatic interactions, and their expression is not dependent on plant age/part (Jones et al., 1997). In fact the use of molecular markers in improving the breeding efficiency in plant breeding was suggested as early as in 1989 (Tanskley et al., 1989; Melchinger, 1990). In this regard, once linkage between a gene for the agronomic trait of interest and marker locus is established, then DNA diagnostic tests can be used to guide plant breeding (Morgante and Salamini, 2003; Gupta and Varshney, 2004). The selection of useful lines for breeding with the help of linked molecular markers is called marker-assisted selection (MAS). Use of MAS is especially advantageous for traits with low heritability where traditional selection is difficult, expensive or lack accuracy or precision.

A variety of molecular markers exist, such as RFLPs (*R*estriction *F*ragment *L*ength *P*olymorphism*s*, Botstein et al., 1980), RAPDs (*R*andom *A*mplification of *P*olymorphic *D*NA*s*, Williams et al., 1990), AFLP*s* (*A*mplified *F*ragment *L*ength *P*olymorphsim*s*, Vos et al., 1995) and microsatellites or SSRs (*S*imple *S*equence *R*epeats, Tautz, 1989). Among the different classes of molecular markers, SSR markers are often chosen as the preferred markers for a variety of applications in breeding because of their multiallelic nature, codominant inheritance, relative abundance and extensive genome coverage (Gupta and Varshney, 2000). More recently, markers such as SNPs (*S*ingle *N*ucleotide *P*olymorphism*s*, Rafalski, 2002) and DArT (*D*iversity *Ar*ray *T*echnology, Killian et al., 2005) have been added to list of preferred marker systems for breeding.

MAS in breeding has revolutionized the improvement of temperate field crops (Koebner, 2004; Varshney et al., 2006) and will have similar impacts on breeding of tropical legume crops, particularly for traits where phenotyping is only possible late in the season, and where screening of traits is difficult or prohibitively expensive. Breeding for enhanced drought and salinity tolerance is notoriously difficult due to the genetic complexity of these traits, the high genotype-by-environment interaction and the difficulties of carrying out precise phenotypic evaluation under field conditions. Part of the problem comes from the difficulty to assess the relative contribution of different traits on the yield under terminal drought. Thus, these are traits where MAS could greatly enhance the effectiveness and impact of plant breeding programs.

4.1. Molecular Tools for SAT Legume Genomics

Molecular markers and molecular genetic linkage maps are the prerequisites for undertaking molecular breeding activities. However, the progress towards development of a reasonable number of molecular markers and molecular genetic maps for cultivated species has been very slow in almost all the three legume crops discussed in this chapter. One of the main reasons for this fact may have been the low level of genetic diversity present in the cultivated gene pools of these species, at least with the detection tools that are currently available. Nevertheless, because of the development of more sophisticated molecular tools, some progress has been made in the area of molecular mapping in these legume species.

4.1.1. Chickpea

The beginnings of the linkage map development in chickpea were based on morphological and isozyme loci. However, their small numbers and the fact that expression of these markers is often influenced by the environment, makes them unsuitable for routine use. Further, there is an extremely low level of polymorphism among genotypes of cultivated chickpea (*C. arietinum* L.). Therefore, interspecific crosses (*C. arietinum* × *C. reticulatum, C. arietinum* × *C. echinospermum*) were exploited for developing genetic linkage maps (Gaur and Slinkard, 1990a, 1990b). The earlier maps were sparse and represented less than 30 loci mapped in a very small portion (about 250 cM) of the chickpea genome (Gaur and Slinkard 1990a, 1990b; Kazan et al. 1993). Integration of molecular markers into genetic linkage maps in chickpea was started with the work of Simon and Muehlbauer (1997). Due to the lack of more recently available molecular markers, Simon and Muehlbauer (1997) employed RFLP and RAPD markers that showed limited polymorphism in the cultivated species (Udupa et al., 1993; Banerjee et al., 1999).

Subsequent development of SSR or microsatellite markers revolutionized genetic analysis and opened new possibilities for the study of complex traits in plant species especially crops like chickpea having a narrow genetic background. As a result, several hundred SSR markers have been developed in chickpea (Huettel et al., 1999; Winter et al., 1999; Sethy et al., 2003, 2006a, 2006b; Lichtenzveig et al., 2005; Choudhary et al., 2006). The majority of these markers have been mapped using interspecific mapping populations (Winter et al., 1999, 2000; Tekeoglu et al., 2002; Pfaff and Kahl, 2003). A genetic map constructed from an interspecific cross, however, may not represent the true recombination distance map order of the cultivated genome due to uneven recombination of homeologous chromosomes and distorted genetic segregation ratios (Flandez-Galvez et al., 2003a). Therefore, in the framework of targeting traits of breeding importance, molecular genetic linkage maps, with SSR markers, have been developed using intraspecific mapping populations from the cultivated gene pool (Cho et al., 2002, Flandez-Galvez et al., 2003a). The genetic linkage maps developed to date with DNA based molecular markers in chickpea are summarized in Table 2.

Table 2. Important genetics maps available for some SAT legume crops

Mapping population	Features of genetic map	Genome coverage	Reference
Chickpea			
F_2, interspecific (*C. arietinum* × *C. reticulatum*) and F_2, interspecific (*C. arietinum* × *C. echinospermum)*	7 linkage groups with 3 morphological and 26 isozymes	200 cM	Gaur and Slinkard, 1990a, 1990b
F_2, interspecific (*C. arietinum* × *C. reticulatum*)	8 linkage groups with 5 morphological and 23 isozymes	257 cM	Kazan et al., 1993
F_2, interspecific (*C. arietinum* × *C. reticulatum*) and F_2, interspecific (*C. arietinum* × *C. echinospermum*)	10 linkage groups with 9 morphological, 27 isozyme, 10 RFLP and 45 RAPD loci	527 cM	Simon and Muehlbauer, 1997
RIL, interspecific (*C. arietinum* 'ICC 4958' × *C. reticulatum* 'PI489777')	11 linkage groups with 120 STMS loci	613 cM	Winter et al., 1999
RIL, interspecific (*C. arietinum* 'ICC 4958' × *C. reticulatum* 'PI489777')	16 linkage groups with 118 SSR, 96 DAF, 70 AFLP, 37 ISSR, 17 RAPD, 8 isozyme, 3 cDNA, 2 SCAR and 3 morphological marker	2,078 cM	Winter et al., 2000
RIL, interspecific (*C. arietinum* 'FLIP 84–92C' × *C. reticulatum* 'PI489777')	9 linkage groups with 89 RAPD, 17 ISSR, 9 isozyme, and 1 morphological marker	982 cM	Santra et al., 2000
RIL, interspecific (*C. arietinum* 'ICC 4958' × *C. reticulatum* 'PI489777')	8 linkage groups, integration of 55 SSR and 1 RGA	1,175 cM	Tekeoglu et al., 2002
RIL, intraspecific (*C. arietinum* 'ICCV 2' x *C. arietinum* JG62')	14 linkages groups with 68 SSR, 34 RAPD, 4 ISSR, and 5 morphological markers	297 cM	Cho et al., 2002
RIL, intraspecific (*C. arietinum* 'ILC 1272' × *C. arietinum* 'ILC3279')	8 linkage groups with 52 SSR, 3 *Ascochyta blight* resistance loci	–	Udupa and Baum, 2003
RIL, interspecific (*C. arietinum* 'ICC 4958' × *C. reticulatum* 'PI489777')	incorporated 47 DR gene specific markers to Winter et al. (2000) 2500 cM , total 296 markers, 12 linkage groups	2,500 cM	Pfaff and Kahl, 2003
F_2, intraspecific (*C. arietinum* 'ICC 12004' x *C. arietinum* 'Lasseter')	8 linkage groups with 54 SSR, 3 ISSR, 12 RGA loci	535 cM	Flandez- Galvez et al., 2003a

(*Continued*)

Table 2. (*Continued*)

Mapping population	Features of genetic map	Genome coverage	Reference
F_2 interspecific (*C. arietinum* 'Lasseter' × *C. echinospermum* 'PI527930')	8 linkage groups with 14 SSR , 54 RAPD, 9 ISSR, 6 RGA loci	570 cM	Collard et al., 2003
RIL, intraspecific (*C. reticulatum* 'PI 359075' × *C. arietinum* 'FLIP 84–92C')	11 linkages groups with 53 SSRs	–	Cho et al., 2004
RIL, intraspecific (*C. arietinum*- two populations, CA2139 × JG62, CA2156 × JG62)	10 linkages groups with 118 RAPD, 13 SSR, 3 ISSR, and 4 morphological markers	–	Cobos et al., 2005
RIL, interspecific (*C. arietinum* 'Hadas' × *C. reticulatum* 'Cr205')	9 linkages groups with 91 SSR, 2 CytP450 markers	345 cM	Abbo et al., 2005
Groundnut			
F_2, interspecific (2*x*) (*A. stenosperma* × *A. cardenassi*)	11 linkage groups with 117 RFLP loci	1,063 cM	Halward et al., 1993
BC interspecific (2*x*) (*A. stenosperma* × (*A. stenosperma* × *A. cardenassi*)	11 linkage groups with 167 RAPD loci	800 cM	Garcia et al., 1995
BC interspecific (4*x*) (*A. batizocoi* × (*A. cardenasii* × *A. diogoi*)	23 linkage groups with 370 RFLP loci.	2,210 cM	Burrow et al., 2001
F(2:3), intraspecific (*A. hypogaea*) (4*x*) ICG 12991 (Spanish) × ICGV SM 93541	5 linkage groups with 12 AFLP loci	139.4 cM of the genome	Herselman et al., 2004
F_2 interspecific (A Genome, 2*x*) (*A. duranensis* × *A. stenosperma*)	11 linkage group with 204 SSR loci	1,231 cM	Moretzsohn et al., 2005
F_2 interspecific (B Genome, 2*x*) (*A. ipaensis* × *A. magna*)	11 linkage group with 94 SSR loci	754.8 cM	Gobbi et al., 2006; D. Bertioli, Brazil (pers. communication)
Pigeonpea			
F_2, interspecific (*C. cajan* × *C. scarabaeoides*)	~200 DArT loci	–	A. Killian, Australia (pers. communication)

Two independent interspecific-derived populations have been extensively employed for genetic linkage map development in chickpea: (i) *C. arietinum* 'ICC 4958' × *C. reticulatum* 'PI489777' at the University of Frankfurt, Germany, (ii) *C. arietinum* 'FLIP 84–92C' × *C. reticulatum* 'PI599072' at Washington State University, Pullman, USA. Among the different types of molecular markers developed for chickpea, SSR markers have proved very useful in linkage mapping and formed the basis for the map initially developed by Winter et al. (1999) that spanned a distance of 613 cM and consisted of 120 SSR markers. This map was greatly extended by Winter et al. (2000) and subsequently by Pfaff and Kahl (2003) with his addition of 47 defense response (DR) genes. The extended map covers a distance of 2500 cM arranged in 12 linkage groups and represents the most extensive linkage map in chickpea. Relatively smaller maps derived from intraspecific (within *C. arietinum*) crosses, have been developed and are being extended (Cho et al., 2002,2004; Flandez-Galvez et al. 2003a; Cobos et al., 2005).

In summary, a reasonable number of SSR markers representing the entire chickpea genome are available at present. The repository of SSR markers for chickpea is being extended by serious efforts by developing new microsatellite markers at NIPGR (Sethy et al., 2003; Chaudhary et al., 2006) and ICRISAT, Patancheru. For instance, a set of about 200 SSRs has been developed at NIPGR (Bhatia et al. unpublished). Similarly sequencing of a microsatellite enriched library of a chickpea (*C. arietinum*) genotype ICC 4958 at ICRISAT, in collaboration with University of Frankfurt, provides another set of about 200 SSRs that can be used to develop markers (Varshney et al., unpublished data). Therefore immediate priority should be accorded to saturation of the existing 'reference' intraspecific as well as interspecific genetic maps with the presently available >500 new (unmapped) SSR markers (Lichtenzveig et al., 2005; Sethy et al., 2006a,b; Choudhary et al. 2006; Bhatia et al., unpublished results; Varshney et al., unpublished results).

4.1.2. Groundnut

The paucity of DNA polymorphism in cultivated groundnut posed a considerable obstacle to genetic mapping in groundnut. For instance, earlier studies using RAPD and RFLP approaches detected limited DNA variation in *Arachis* species (Kochert et al., 1991; Halward et al., 1992; Paik-Ro et al., 1992). The use of a synthetic amphidiploid TxAG-6 (Simpson et al., 1993) made possible the generation of the first molecular map representing the entire tetraploid genome of groundnut. The discovery of a high level of polymorphism between the cultivar Flourunner and the parents of TxAG-6 by RAPD analysis (Burrow et al., 1996) was followed by RFLP analysis showing 83% polymorphism on a per band basis (Burrow et al., 2001). By using 78 BC_1F_1 lines generated from the cross (TxAG-6 x Florunner), mapping of 220 cDNA probes integrated 370 RFLP loci into 23 linkage groups. The total length of the first tetraploid map was 2210 cM, which was slightly greater than twice the length (1063 cM) of the map previously reported from a cross between two A-genome diploid species (Halward et al., 1993). The common markers mapped

in both crosses showed a high degree of collinearity between the diploid and tetraploid chromosomes (Burrow et al., 2001). These studies have been summarized in the database PeanutMap (http://peanutgenetics.tamu.edu/cmap; Jesubatham and Burrow, 2006).

In terms of mapping the diploid genomes of *Arachis*, the first genetic map was constructed by Halward et al. (1993) based on the 87 F_2 lines derived from a cross of *A. stenosperma* x *A. cardenasii* and contained 117 RFLP loci on 11 linkage groups with a genome coverage of 1400 cM. RFLP analysis is time consuming and labor intensive. RAPD and AFLP were used to detect DNA polymorphism in several studies in different germplasm collections (He and Prakash, 1997; Subramanian et al., 2000; Dwivedi et al., 2001; Raina et al., 2001; Milla et al., 2005), but represent dominant markers with low information content. As a result of extensive efforts of several laboratories, a large number of microsatellite markers have been generated in groundnut (Hopkins et al., 1999; He et al., 2003; Ferguson et al., 2004; Moretzsohn et al., 2004; Mace et al., unpublished; D. Bertioli, Brazil, pers. commun.; S. Knapp, USA, pers. commun.). The availability of more than 500 SSR markers in groundnut provides the opportunity to integrate these markers into a genetic linkage map. However, these markers have been integrate only in the AA- genome map (Moretzsohn et al., 2005) by using an F_2 population obtained from a cross between two diploid species with AA genome (*A. durasenis* and *A. stenosperma*). The genetic map had 80 SSR loci on 11 linkage groups covering 1231 cM. Similar efforts to prepare a genetic map for BB genome are underway in Brazil. As of now, the genotyping of a F_2 population derived from cross between *A. ipaensis* (KG30076) and *A. magna* (KG30097) has resulted in development of 11 linkage groups with 94 markers (Gobbi et al. 2006). As a part of Generation Challenge Programme (GCP) of CGIAR, preparation of the first genetic map for tetraploid cultivated groundnut species is in progress at ICRISAT. However, the lower level of polymorphism between the parental genotypes of existing mapping populations (e.g. TAG24 × ICGV 86031 developed at ICRISAT, GPBD4 × TAG24 developed at UAS Dharwad) poses a serious problem. Nevertheless, we expect to prepare the partial/genome wide map with about 100 SSR loci (Varshney et al., unpublished results). The progress in the area of genome mapping of *Arachis* species is summarized in Table 2.

4.1.3. Pigeonpea

In case of pigeonpea, molecular markers (RFLPs) were used as early as 1994 to study genetic diversity in wild species using nuclear DNA probes (Nadimpalli et al., 1994). Subsequently, Ratnaparkhe et al. (1995) attempted to study DNA polymorphism in cultivars and wild species. The level of polymorphism among the wild species was extremely high, while little polymorphism was detected within *C. cajan* accessions. In order to characterize a few putative cytoplasmic male sterlity lines, maize mitochandrial DNA (mt DNA) specific probes were used in RFLP analysis (Sivaramakrishanan et al., 1997). Recently, AFLP analysis was carried out with a few cultivars and two wild species (*Cajanus volubilis, Rhynchosia bracteata*)

using two *Eco*RI and 14 *Mse*I primers (Punguluri et al., 2006). The two wild species shared only 7% bands with the pigeonpea cultivars, whereas 87% common bands were seen among cultivars. The cluster analysis revealed low polymorphism among pigeonpea cultivars and very high polymorphism between cultivated pigeonpea and its wild relatives. Similar results were obtained in a very recent analysis using DArT markers (Yang et al., 2006).

In terms of development of SSR markers, about 10 SSR markers are available in public domain (Burns et al., 2001). To develop a resource of microsatellite markers for pigeonpea, primer pairs were generated for 39 microsatellite loci at ICRISAT. These markers (19 polymorphic loci) yielded an average of 4.9 alleles per locus while the observed heterozygosity ranged from 0.17–0.80 with a mean of 0.42 per locus (Odeney et al., 2007). However, to the best of our knowledge, there is no report on any genetic mapping in pigeonpea. In collaboration with ICRISAT, some efforts are underway to develop the first generation map for pigeonpea based on DArT markers at DArT Pty. Ltd. (A. Killian, pers. commun.).

4.2. Trait Mapping and Marker-Assisted Selection

Marker-assisted selection (MAS) offers great promise for improving the efficiency of conventional plant breeding. Molecular markers are especially advantageous for traits where conventional phenotypic selection is difficult, expensive, or lacks accuracy or precision. Molecular mapping and identification of molecular markers associated with genes and QTLs for traits are prerequisites for the MAS. As mentioned above, though not excellent, some progress has been made in the area of development of molecular markers or construction of genetic maps in chickpea and groundnut. As a result, molecular markers linked to a few abiotic or biotic stress tolerance/resistance as well as agronomic traits have been identified recently.

4.2.1. Chickpea

Genetic mapping in chickpea has focussed on tagging agronomically relevant genes such as ascochyta blight resistance (Tekeoglu et al., 2002; Udupa and Baum, 2003; Collard et al., 2003; Flandez-Galvez et al., 2003b; Millan et al., 2003; Cho et al., 2004; Iruela et al., 2006), fusarium wilt resistance (Benko-Iseppon et al., 2003; Sharma et al., 2004) and yield-influencing characters such as double podding and other morphological characters (Cho et al., 2002; Rajesh et al., 2002; Abbo et al., 2005; Cobos et al., 2005). Progress in the area of mapping of ascochyta blight resistance has been summarized recently by Millan et al. (2006). Since apparently all major blight resistance QTLs are tagged with SSR markers, pyramiding of resistance genes via MAS should now be feasible and awaits its proof-of-principle. The genetic control of this disease bred into cold tolerant germplasm would be a major breakthrough for yield increases in Mediterranean-type environments in many parts of the world.

In order to address the issue of drought tolerance through molecular markers, more than 1500 chickpea germplasm and released varieties were screened for

drought tolerance at ICRISAT. The most promising drought tolerant variety was ICC 4958 that had 30% more root volume than the popular variety Annigeri (Saxena et al., 1993); therefore, root traits were considered important parameters to improve the drought tolerance (Kashiwagi et al., 2006). Selection for root traits is very difficult, since it involves laborious methods such as digging and measuring root length and density. Molecular tagging of major genes for root traits may enable MAS for these traits and could greatly improve the precision and efficiency of breeding. In this direction, a set of 257 recombinant inbred lines (RILs) was developed from the cross Annigeri × ICC 4958 at ICRISAT and glasshouse-evaluated to identify molecular markers for root traits. After screening the parental genotypes with over 250 STMS and 100 EST markers and the mapping population with 57 poymorphic markers, a QTL flanked by STMS markers TAA170 and TR55 on LG 4A was identified that accounted for maximal phenotypic variation in root length (33%), root weight (33%) and shoot weight (54%) (Chandra et al., 2004). Genotyping of two other mapping populations (ICC 4958 × ICC 1882 and ICC 8261 × ICC 283), which have larger genetic variation than Annigeri × ICC 4958 with SSR markers is in progress at ICRISAT.

For improving cold tolerance, AFLP markers have been linked to the trait using bulked segregant analysis of F_2 progeny of a cross between the chilling sensitive cultivar Amethyst and the chilling tolerant ICCV 88516 (Clarke and Siddique, 2003). Candidate AFLP markers were converted into SCAR markers (Paran and Michelmore, 1993) to overcome the limitations of the dominant AFLP marker system. The most promising primers were based on a 560 bp fragment containing a simple sequence repeat (3 bp repeat microsatellite) with nine repeats in the susceptible parent and ten repeats in the tolerant parent. The three-base difference was visualised on a vertical acrylamide gel, and was very useful in the selection of chilling tolerant progeny derived from crosses between ICCV 88516 and Amethyst. Unfortunately, there has been no success in applying these SCAR markers to other breeding materials.

In the case of flowering, a major gene (*efl-1*) for time of flowering was reported by Kumar & van Rheenen (2000), and another one (*ppd*) by Or et al. (1999). The latter gene controls time to flowering through photoperiod response (Hovav et al., 2003). Cho et al. (2002) mapped a QTL for days to 50% flowering to LG 3. Another QTL was also located on this linkage group in an interspecific RIL population and explained 28% of the total phenotypic variation (Cobos et al., 2005).

In addition to the above mentioned traits, molecular mapping for other traits is in progress in many laboratories. For instance, SSR-based genotyping and phenotyping of one mapping population (ICCV 2 × JG 62) is in progress at NIPGR and ICRISAT to identify the molecular markers associated with salinity tolerance.

4.2.2. Groundnut

There are very few genetic maps available based on cultivated groundnut genotypes. The available maps, based on interspecific crosses, will be useful in locating specific

genes of interest in the interspecific crosses and also providing valuable information about genome organization and evolution. However, these markers will be of less value in elite cultivated germplasm, in which very little polymorphism exist.

Although marker-trait association has been little used within *A. hypogaea*, even with the limitations afforded by present technologies, it has much potential for introgressing genes from closely related *Arachis* species into the cultivated genome. For instance, Garcia et al. (1995) showed introgression of genes from *A. cardenasii* into *A. hypogaea* in 10 of 11 linkage groups on the diploid RFLP map developed by Halward et al. (1993). Subsequently, Garcia et al. (1996) used RAPD and SCAR technologies to map two dominant genes conferring resistance to the nematode by using the mapping population derived from the cross *A. hypogaea* x *A. cardenasii.* Burrow et al. (1996) identified RAPD markers linked to nematode resistance in another interspecific cross involving the species *A. hypogaea*, *A. batizocoi*, *A. cardenasii* and *A. diogoi*. Such linkage of RAPD markers with components of early leaf spot and corn rootworm resistance was shown in another interspecific cross (Stalker and Mozingo, 2001). By using the BSA approach with an F_2 population derived from the cross (ICG 12991 × ICGVSM 93541) and phenotyping the F_3 population, twenty putative AFLP markers were identified of which12 mapped to five linkage groups. Interestingly, mapping of a single recessive gene on linkage group 1 (3.9 cM from a marker originating from the susceptible parent) explained 76% of the phenotype variation for aphid resistance. AFLP markers were used to establish marker-trait association for tomato spotted wilt virus resistance in groundnut (Milla 2003). Marker-trait association studies for several other traits, e.g, water use efficiency (WUE), rust and late leaf spot (LLS) are underway at ICRISAT and UAS Dharwad.

4.2.3. Pigeonpea

Higher level of heterogeneity and very low level of genetic variation in cultivated pigeonpea has hampered development of genetic maps and marker-trait association analysis. Recently, the use of RAPD markers through BSA approach showed association of two RAPD loci with fusarium wilt resistance (Kotresh et al., 2006). It is anticipated that development of higher number of polymorphic SSR markers and DArT arrays (A. Killian, pers. commun.) in pigeonpea will facilitate trait mapping in the near future.

5. NOVEL GENETIC AND GENOMICS APPROACHES

New technologies promise to resolve constraints that have been limiting the impact of linkage based molecular mapping. Such modern genomics approaches have been used in some cereal and other plant species, and legume improvement can be benefited by exploring such approaches.

5.1. Association Mapping and Advanced Backcross QTL (AB-QTL) Analysis

In general, a low level of polymorphism has been a major constraint in developing genetic maps in the legume crops mentioned in this chapter. Further, species like pigeonpea, which is of regional importance in Asia and Africa, has not been explored at the international level. Non-availability of resistance sources in cultivated genepools of these species for several fungal and viral diseases, e.g., pod borer in chickpea and pigeonpea, sterility mosaic in pigeonpea, aflatoxin in groundnut, and the difficulties of crossing cultivated species with wild species are other barriers that hampered the development of appropriate mapping populations in these legume species. Novel approaches, based on classical genetics, like linkage disequilibrium (LD) based association mapping (Hirschhorn and Daly, 2005), advanced back-cross QTL (AB-QTL) analysis (Tanksley and Nelson, 1996) offers the possibility to overcome at least a few barriers. For instance, an appropriate natural population, genebank or breeding material may be used in LD-based association analysis. In this regard, emergence of novel marker systems such as SNPs and DArTs and developments in this direction for the mentioned legume species would make it possible to undertake candidate gene sequencing (using SNP assays) as well as whole genome scanning (using DArTs) based approaches for association analyses. In contrast to the numerous linkage disequilibrium (LD) studies in human and other mammals, there are very few publications on this topic in agriculturally important crops including legumes (Virk et al., 1996; Beer et al., 1997; Pakniyat et al., 1997; Forster et al., 1997; Igartua et al., 1999; Remington et al., 2001; Thornsberry et al. 2001; Turpeinen et al. 2001; Hansen et al. 2001; Sun et al. 2001, 2003; Skot et al., 2002; Ivandic et al., 2002, 2003; Amirul Islam et al., 2004; Zhu et al., 2003; Simko et al., 2004). Traditionally the plant community has been reticent to use LD mapping believing that it can lead to spurious and non-functional associations due to mutation, genetic drift, population structure, breeding systems and selection pressure (Hill and Weir, 1994; Pritchard et al., 2000). However, most of these limitations are being overcome in recent mammalian studies by following precautions that minimize circumstantial correlations and maximize the accuracy of association statistics (Yu et al., 2006; Yu and Buckler, 2006; Ersoz et al., 2007). Unfortunately the real value of LD mapping in legume species remains to be demonstrated as most of the reports to date are based on small population sizes or a limited number of markers and generally lack validation.

Advanced-backcross QTL analysis (AB-QTL), proposed by Tanksley and Nelson (1996), involves transferring the QTLs of agronomically important traits from a wild species to a crop variety. In this approach, a wild species is backcrossed to a superior cultivar with selection for domestication traits. Selection is imposed to retain individuals that exhibit domestication traits such as non-shattering. The segregating BC_2F_2 or BC_2F_3 population is then evaluated for traits of interest and genotyped with polymorphic molecular markers. These data are then used for QTL analysis, potentially resulting in identification of QTLs, while transferring these QTLs into adapted genetic backgrounds. The AB-QTL approach has been

evaluated in many crop species to determine if genomic regions (QTLs) derived from wild or unadapted germplasm have the potential to improve yield (for a review, see Varshney et al., 2005). However, the wild species chromosome segments masked the magnitude of some of favourable effects that were identified for certain introgressed alleles (Septiningsih et al., 2003). Thus, yield promoting QTL did not have a substantial contribution to the phenotype and the best lines were inferior to commercial cultivars in some studies. In tomato, however, the pyramiding of independent yield promoting chromosome segments resulted in new varieties with increased productivity under normal and stress conditions (Wang D. et al., 2004). One disadvantage is that the value of the wild accession for contributing useful QTL alleles is unknown prior to a major investment in mapping. Nevertheless, the approach holds a great potential to harness the potential of wild species for crop improvement in case of legume species where only low level of genetic variation and source of resistance/tolerance to biotic/abiotic stresses exist in the cultivated gene pool.

5.2. Transcriptomics and Functional Genomics

Functional genomics has revolutionized biological research and is predicted to have a similar impact on plant breeding through the evolution of marker-assisted to genomics-assisted breeding (Varshney et al., 2005). The salient challenge of applied genetics and functional genomics is the identification of the genes underlying a trait of interest so that they can be exploited in crop improvement programmes. Among legume species, much work in terms of development of functional genomics resources such as ESTs, genome sequencing, array development has been done either in model species like lotus (*Lotus japonicus* L.) and medicago (*Medicago truncatula* L.) or major species like soybean. In contrast, only a limited number of ESTs have been generated so far in legume species of SAT (Table 1). These ESTs can be used to develop the molecular markers as shown in chickpea (Buhariwalla et al., 2005) and groundnut (Luo et al., 2005) as well as to develop cDNA arrays. At NIPGR, the chickpea ESTs are being developed from seeds (both developing and maturing) and symbiotic root nodules in association with *Mesorhizobium ciceri*. So far about 1000 seed specific unigenes have been identified (unpublished results). The most striking feature of these ESTs is that, majority of them are putative or unknown proteins. The use of suppression subtractive hybridization (SSH) to prepare the subtracted cDNA library of 7-day old symbiotic root nodules lead to the identification of three putative genes regulated during symbiotic relationship with *M. ciceri*. Further validation with Northern analysis has lead to the identification three putative genes up-regulated during symbiotic association in a temporal manner.

The macro- and micro-arrays based on EST/gene sequence information have been successfully utilized in many plant species for understanding the basic physiology, developmental processes, environmental stress responses, and for identification and genotyping of mutations. Recently in chickpea, a small array with 768 features has been developed (Coram and Pang, 2005a) that has been used to identify genes

responsible for ascochyta blight resistance (Coram and Pang, 2005b, 2006), drought and salinity tolerance (E. Pang, pers. commun.). The candidate genes identified by EST sequencing (and gene prediction) and functional genomics approaches can be further verified through real time PCR analysis (Luo et al., 2005) and genetical genomics/ expression genetics approaches (Jansen and Nap, 2001; Varshney et al., 2005) after conducting gene expression analysis in quantitative fashion using segregating mapping populations. By analyzing the expression levels of genes or clusters of genes within a segregating population, it is possible to map the inheritance of that expression pattern. The QTLs identified using expression data in a mapping population are called e(xpression)QTLs. The eQTLs can be classified as *cis* or *trans* acting based on location of transcript compared to that of the eQTL influencing expression of that transcript (de Konig and Haley, 2005). Because of this feature, eQTL analysis makes it possible to identify factors influencing the level of mRNA expression. The regulatory factor (second order effect) is of specific interest because more than one QTL can be putatively connected to a *trans*-acting factor (Schadt et al., 2003). Thus, the mapping of eQTLs allows multifactorial dissection of the expression profile of a given mRNA or cDNA, protein or metabolite into its underlying genetic components as well as localization of these components on the genetic map (Jansen and Nap, 2001). In recent years, in many plant species, the genetical genomics approach has demonstrated its power (see Kirst and Yu, 2007).

Another powerful approach of gene discovery is 'Serial Analysis of Gene Expression (SAGE)' (Velculescu et al., 1995) that utilizes the advantage of high-throughput sequencing technology to obtain a quantitative profile of gene expression which measures not the expression level of a gene, but quantifies a 'tag' which represents the transcriptome product of a gene. A tag for the purpose of SAGE, is a nucleotide sequence of a defined length, directly adjacent to the 3'-most restriction site for a particular restriction enzyme. The data product of the SAGE technique is a list of tags, with their corresponding count values, and thus is a digital representation of cellular gene expression. Based on the length of tags, several modified forms of SAGE, e.g., MicroSAGE, MiniSAGE, LongSAGE and SuperSAGE, have been developed (Sharma et al., 2007). In fact, by using SuperSAGE methodology, over 220,000 SuperTags describing the differential transcription profiles of chickpea roots and nodules have already been sequenced at University of Frankfurt (G. Kahl, pers. commun.). Targeted gene-expression chips are being developed by adding SuperTag oligonucleotides derived from the most informative genes expressed differentially under stress- and non-stress conditions and from large-versus small root systems to a gene expression chip (P. Winter, pers. commun.).

In groundnut, recent activities in the area of functional genomics have produced a gene chip with 400 unigenes after cluster analysis of 1825 ESTs and used for identifying the genes associated with disease resistance and drought tolerance (Luo et al., 2003, 2005). Further to validate the microarray and EST data by EST-discovery, real-time PCR analysis was conducted for 10 specific genes (Luo et al.,

2005). The use of suppression subtractive hybridization (SSH) to prepare the subtracted cDNA libraries and identify the genes regulated during interaction with the fungus *Cercosporidium personatum* (causing the disease late leaf spot) is in progress in Brazil (Nobile et al., 2006). To understand the molecular mechanisms of drought tolerance, the use of differential expression of mRNA transcripts and proteins are underway at Florida A & M University (Katam et al., 2006). With the development of more functional genomics resources in SAT legumes, it is anticipated that the use of functional genomics and expression genetics approaches may help the community to dissect the complex traits and devise strategies for crop improvement.

5.3. Comparative Genomics

In recent years, the availability of ESTs and genome sequence data for model legumes i.e. medicago (*M. truncatula*), and lotus (*L. japonicus*) and major crop legumes like soybean has opened the possibilities of transfer of information from model to crop legumes and vice-versa (Gepts et al., 2005, Young et al., 2005). Identification of putative orthologs from related genomes will facilitate comparative genomics and comparative genetic mapping. Using 274 unique low copy gene specific markers from *M. truncatula* and *G. max*, Choi et al. (2004, 2006) have demonstrated that gene-specific markers are transferable across *Papilionoid* legume species may find utility in phylogenetic relationship assessment at different, but overlapping, taxonomic levels. Moreover, majority of these markers (85.3%) are also linked to the legume genetic maps. Similarly, Gutierrez et al. (2005) have studied the conservation of 209 EST-SSR markers from the model legume *M. truncatula* in three major European crop legumes i.e. faba bean (*Vicia faba*), pea (*Pisum sativum*) and chickpea and have reported 36%–40% transferability range for this class of markers. Recently, extensive efforts have been made to develop bioinformatics tools and pipelines after exploiting the genomics resources of model species as well as other legume species and as a result about 450 cross species markers have been developed (Fredslund et al., 2005, 2006a, 2006b). For many markers, the map position in lotus and/or medicago is known and in other legume species such as groundnut, soybean, chickpea, these markers are being mapped. These studies will provide more anchor points to relate different legume genomes, Moreover, the identification of the cross-genera transferable legume SSR markers will cut down the cost and labor associated with development of SSR markers in the orphan legumes and will help in comparative mapping and map-based cloning of orthologous genes. Since the EST-SSR markers reveal very less polymorphism in legumes (Gutierrez et al. 2005), the alternative source is the genome specific genomic SSRs. By virtue of their long polymorphic microsatellite repeat stretches and the variable microsatellite flanking region, the genomic microsatellites are a promising source of cross-transferable markers in self-pollinating legume species (Sethy et al., manuscript in preparation). The levels and patterns of conservation of

Cicer genomic SSR markers across model, crop and fodder legumes have demonstrated that the genomic SSRs find a mean average transferability of nearly 25% across *M. truncatula*, *L. japonicus*, soybean, pea, lentil, pigeonpea, blackgram, mungbean and *Trifolium alexandrinum* (Figure 1) and often conserved in the model plant *A. thaliana*. Moreover, the *Cicer* markers have been demonstrated to be

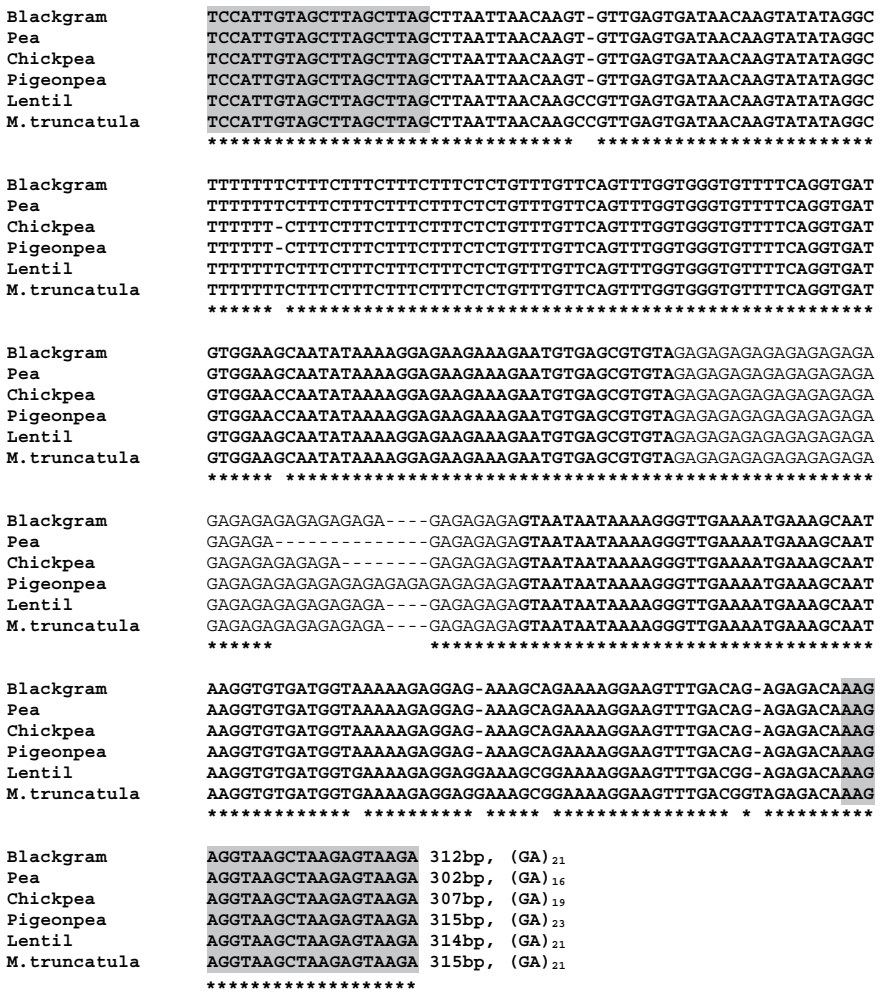


Figure 1. Multiple sequence alignment of the size variant alleles of the legume accessions at the chickpea STMS marker NIPGR19 locus. Accessions of *M. truncatula* (SA27783), blackgram (IC342955), lentil (IC383669), pea (RFP16) and pigeonpea (IC347150) along with chickpea (Pusa362) are analyzed. The *asterisks* indicate similar sequences and '–' indicate alignment gaps. The repeat region is indicated in boldface and shadowed boxes indicate conserved primer binding sites. Allele sizes and repeat motifs are mentioned at the end of the sequence

polymorphic even within *M. truncatula*, soybean and blackgram opening the possibility of comparative mapping and generation of a consensus legume genetic map (Sethy et al., unpublished results).

6. TOWARDS A BRIGHT FUTURE OF MOLECULAR BREEDING IN SAT LEGUMES

Traditional cropping systems across the world have depended on the rotation of cereal and legume crops. However, with increasing intensification of agriculture during the twentieth century, there has been a substantial emphasis on cereals as the pre-eminent food commodity in national production and international trade. In turn, this has been reflected by a continuous and cumulative increase in funding for research and breeding of cereal crops (Goff and Salmeron, 2004) that has resulted in the state-of-the-art in legumes falling further and further behind. Nevertheless, progress in the genomics of two legume species, medicago and lotus, as model genomes offers the potential for real technological leap-frogging amongst legume crops.

Although during the past few years, significant progress has been made in the area of genomics of SAT legume crops as a large number of molecular (SSR) markers and ESTs have been developed, there is still a need to develop more SSR, SNP or DArT markers and dense genetic maps for the species mentioned in this chapter. Further the generation of some BAC and BIBAC libraries in case of chickpea and groundnut offers the possibility to develop genome wide or local physical maps to isolate genes for resistance/tolerance to biotic/abiotic stresses as well as agronomic traits (Yuksel et al., 2005). Thus molecular breeding through existing tools in combination with continuous incremental changes such as improvements in genetics and biometrics, plus revolutionary changes including automation of breeding trials and computerization of phenotyping will be very useful for legume improvement (Dwivedi et al., 2006). In addition to linkage based trait mapping, several other approaches such as LD-based association mapping, AB-QTL analysis, transcriptomics and functional genomics can be used to identify the molecular markers or candidate genes for traits of interest in breeding. Beyond its increased power of selection, marker or genomics-assisted breeding offers additional advantages in the economics of scale both in terms of cost and time as very different traits can be manipulated using the same technology. The proof-of-function of candidate genes can be obtained by using TILLING (Targeting Induced Local Lesions In Genomes, see Till et al., 2007) population, while the EcoTILLING approach may be used for allele mining to improve the traits. Allele mining for candidate genes should provide superior alleles and haplotypes for the traits (Varshney et al., 2005).

Recent studies show strong correlation between the degree of synteny and phylogenetic distance in legumes (Young et al., 2003; Wang M.L. et al., 2004; Choi et al., 2004). Therefore, advances in the area of genomics of medicago and lotus may be used to transfer information on genes involved in nitrogen fixation and other physiological processes of agronomic importance in SAT legume crops by utilizing

the comparative genomics approach combined with bioinformatics. However, the extent to which genetic knowledge from model systems will readily translate into economic impact in related crops remains to be empirically demonstrated (Thro et al., 2004; Koebner and Varshney, 2006). Genomics research in the legume crops together with model systems will soon routinely define the location of genomic regions controlling a target trait as well as identify underlying candidate genes and their sequences through mapping, mutation analysis and transcriptomics. Based on this new knowledge it will be possible to develop highly precise DNA markers for selection or introgression of desired traits. While the newly developed genetic and genomics tools will certainly enhance the prediction of phenotype, they will not entirely replace the conventional breeding process.

ACKNOWLEDGEMENTS

Authors would like to thank the past and present colleagues at ICRISAT as well as our partners for their dedicated work in the area of legume improvement. Thanks are due to Priti Swamy, Yogendra Khedikar and Rachit Saxena, research scholars with RKV at ICRISAT for their help in collecting some literature and formatting the references.

REFERENCES

Abbo S, Molina C, Jungmann R, Grusak MA, Berkovitch Z, Reifen R, Kahl G, Winter P, Reifen R (2005) QTL governing carotenoid concentration and weight in seeds of chickpea (*Cicer arietinum* L.). Theor Appl Genet 111:185–195

Amirul Islam FM, Beebe S, Muñoz M, Tohme J, Redden RJ, Basford KE (2004) Using molecular markers to assess the effect of introgression on quantitative attributes of common bean in the Andean gene pool. Theor Appl Genet 108:243–252

Azhaguvel P, Vidya Saraswathi D, Sharma A, Varshney RK (2006) Methodological advancement in molecular markers to delimit the gene(s) for crop improvement. In: Teixera da Silva J. (ed) Floriculture, ornamental and plant biotechnology: Advances and topical issues, Global Science Books, London, UK, pp 460–499

Banerjee H, Pai RA, Sharma RP (1999) Restriction fragment length polymorphism and random amplified polymorphic DNA analysis of chickpea accessions. Biol Plant 42:197–199

Beer SC, Siripoonwiwat W, O'Donoughue LS, Souza E, Matthews D, Sorrells ME (1997) Association between molecular markers and quantitative traits in an oat germplasm pool: can we infer linkages. Jour Agr Genom 3 (http://wheat.pw.usda.gov/jag/papers97/paper197/jqtl1997–01.html)

Benko-Iseppon AM, Winter P, Huettel B, Staginnus C, Muehlbauer FJ, Kahl G (2003) Molecular markers closely linked to *Fusarium* resistance genes in chickpea show significant alignments to pathogenesis-related genes located on *Arabidopsis* chromosomes 1 and 5. Theor Appl Genet 107:379–386

Botstein D, White DL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314–331

Bretting PK, Widrlechner MP (1995) Genetic markers and plant genetic resource management. Plant Breed Rev 31:11–86

Brown AHD (1989) Core collections: a practical approach to genetic resources management. Genome 31:818–824

Brown SM, Kresovich S (1996) Molecular characterization for plant genetic resources conservation. In: Paterson AH (ed) Genome mapping of plants. Academic Press, San Diego, pp 85–93

Buhariwalla HK, Jayashree B, Eshwar K, Crouch JH (2005) ESTs from chickpea roots with putative roles in drought tolerance. BMC Plant Biol 5:16

Burns MJ, Edwards KJ, Newbury HJ, Ford-Lloyd BV, Baggott CD (2001) Development of simple sequence repeat (SSR) markers for the assessment of gene flow and genetic diversity in pigeonpea (*Cajanus cajan*). Mol Ecol Notes 1:283–285

Burow MD, Starr JL, Simpson CE, Paterson AH (1996) Identification of RAPD markers in peanut (*Arachis hypogaea* L.) associated with root-knot nematode resistance derived from *A. cardenasii*. Mol Breed 2:307–319

Burrow MD, Simpson CE, Paterson AH, Starr JL (1996) Identification of peanut (*Arachis hypogaea* L.) RAPD markers diagnostic of root-knot nematode (*Meloidogyne arenaria* (Neal) Chitwood) resistance. Mol Breed 2:369–379

Burrow MD, Simpson CE, Starr JL, Paterson AH (2001) Transmission genetics of chromatin from a synthetic amphidiploid to cultivated peanut (*Arachis hypogaea* L.): broadening the gene pool of a monophyletic polyploid species. Genetics 159:823–837

Chandra S, Buhariwalla HK, Kashiwagi J, Harikrishna S, Rupa Sridevi K, Krishnamurthy L, Serraj R, Crouch JC (2004) Identifying QTL-linked markers in marker-deficient crops. In new directions for a diverse plant: Proceedings of the 4th international crop science congress brisbane, Australia, 26 September–1 October 2004

Cho S, Kumar J, Shultz JF, Anupama K, Tefera F, Muehlbauer FJ (2002) Mapping genes for double podding and other morphological traits in chickpea. Euphytica 125:285–292

Cho S, Chen W, Muehlbauer FJ (2004) Pathotype-specific genetic factors in chickpea (*Cicer arietinum* L.) for quantitative resistance to ascochyta blight. Theor Appl Genet 109:733–739

Choi HK, Mun JH, Kim DJ, Zhu H, Baek JM, Mudge J, Roe B, Ellis N, Doyle J, Kiss GB, Young ND, Cook DR (2004) Estimating genome conservation between crop and model legume species. Proc Nat Acad Sci 101:15289–15294

Choi HK, Luckow MA, Doyle J, Cook DR (2006) Development of nuclear gene-derived molecular markers linked to legume genetic maps. Mol Gen Genome 276:56–70

Choudhary S, Sethy NK, Shokeen B, Bhatia S (2006) Development of sequence tagged microsatellite site markers for chickpea (*Cicer arietinum* L.). Mol Ecol Notes 6:3–95

Clarke H, Siddique KHM (2003) Chilling tolerance in chickpea – novel methods for crop improvement. In: Sharma RN, Yasin M, Swami SL, Khan MA, William AJ (ed) International Chickpea Conference, Indira Gandhi Agricultural University, Raipur, India, pp 5–12

Cobos MJ, Fernandez MJ, Rubio J, Kharrat M, Moreno MT, Gil J, Millan T (2005) A linkage map of chickpea (*Cicer arietinum* L.) based on populations from Kabuli × Desi crosses: location of genes for resistance to *fusarium wilt* race 0. Theor Appl Genet 110:1347–1353

Collard BCY, Pang ECK, Ades PK, Taylor PWJ (2003) Preliminary investigation of QTLs associated with seedling resistance to ascochyta blight from *Cicer echinospermum*, a wild relative of chickpea. Theor Appl Genet 107:719–729

Coram TE, Pang ECK (2005a) Isolation and analysis of candidate ascochyta blight defence genes in chickpea. Part I. Generation and analysis of an expressed sequence tag (EST) library. Physiol Mol Plant Pathol 66:192–200

Coram TE, Pang ECK (2005b) Isolation and analysis of candidate ascochyta blight defence genes in chickpea. Part II. Microarray expression analysis of putative defence-related ESTs. Physiol Mol Plant Path 66:201–210

Coram TE, Pang ECK (2006) Expression profiling of chickpea genes differentially regulated during a resistance response to *Ascochyta rabiei*. J Plant Biotechnol 4:647–666

De Koning DJ, Haley CS (2005) Genetical genomics in humans and model organisms. Trends Genet 21:377–381

Desbrosses GG, Kopka J, Udvardi MK (2005) *Lotus japonicus* metabolic profiling. Development of gas chromatography-mass spectrometry resources for the study of plant–microbe interactions. Plant Physiol 137:1302–1318

Dixon RA, Sumner LW (2003) Legume natural products: understanding and manipulating complex pathways for human and animal health. Plant Physiol 131:878–885

Duranti M, Gius C (1997) Legume seeds: protein content and nutritional value. Field Crops Res 53:31–45

Dwivedi SL, Gurtu S, Chandra S, Yuejin W, Nigam SN (2001) Assessment of genetic diversity among selected groundnut germplasm. I: RAPD analysis. Plant Breed 120:345–349

Dwivedi SL, Blair MW, Upadhyaya HD, Serraj R, Balaji J, Buhariwalla HK, Ortiz R, Crouch JH (2006) Using genomics to exploit grain biodiversity in crop improvement. Plant Breed Rev 26: 171–310

Ersoz ES, Yu J, Buckler ES (2007) Applications of linkage disequillibrium and association mapping in crop plants. In: Varshney RK, Tuberosa R (eds) Genomics assisted crop improvement Vol 1: Genomics approaches and platforms, Springer, The Netherlands, pp 97–120

Ferguson ME, Burrow M, Schulze S, Bramel PJ, Paterson A, Kresovich S, Mitchell S (2004) Microsatellite identification and characterization in peanut (*A. hypogaea* L.). Theor Appl Genet 108:1064–1070

Flandez-Galvez H, Ford R, Pang ECK, Taylor PWJ (2003a) An intraspecific linkage map of the chickpea (*Cicer arietinum* L.) genome based on sequence tagged based microsatellite site and resistance gene analog markers. Theor Appl Genet 106:1447–1453

Flandez-Galvez H, Ford R, Taylor PWJ (2003b) Mapping QTLs governing resistance to ascochyta blight in chickpea. Theor Appl Genet 107:1257–1265

Forster BP, Russell JR, Ellis RP, Handley LL, Robinson D, Hackett CA, Nevo E, Waugh R, Gordon DC, Keith R, Powell W (1997) Locating genotypes and genes for abiotic stress tolerance in barley: a strategy using maps, markers and the wild species. New Physiol 137:141–147

Fredslund J, Schauser L, Madsen LH, Sandal N, Stougaard J (2005) PriFi: using a multiple alignment of related sequences to find primers for amplification of homologs. Nucl Acids Res 33: W516–W520

Fredslund J, Madsen LH, Hougaard BK, Nielsen AM, Bertioli D, Sandal N, Stougaard J, Schauser L (2006a) A general pipeline for the development of anchor markers for comparative genomics in plants. BMC Genome 7:207

Fredslund J, Madsen LH, Hougaard BK, Sandal N, Stougaard J, Bertioli D, Schauser L (2006b) GeMprospector – Online design of cross-species genetic marker candidates in legumes, grasses. Nucl Acids Res 34:W630–W640

Garcia GM, Stalker HT, Kochert G (1995) Introgression analysis of an interspecific hybrid population in peanuts (*Arachis hypogaea* L.) using RFLP and RAPD markers. Genome 38:166–176

Garcia GM, Stalker HT, Shroeder E, Kochert GA (1996) Identification of RAPD, SCAR and RFLP markers tightly linked to nematode resistance genes introgressed from *Arachis cardenasii* to *A. hypogaea*. Genome 39:836–845

Gaur PM, Slinkard AE (1990a) Genetic control and linkage relations of additional isozymes markers in chickpea. Theor Appl Genet 80:648–653

Gaur PM, Slinkard AE (1990b) Inheritance and linkage of isozyme coding genes in chickpea. J Hered 81:455–459

Gepts P, Beavis WD, Brummer EC, Shoemaker RC, Stalker HT, Weeden NF, Young ND (2005) Legumes as a model plant family. Genomics for food and feed report of the cross-legume advances through genomics conference. Plant Physiol 137:1228–1235

Ghassemi F, Jackman AJ, Nix HA (1995) Salinization of land and water resources. CAB international, Wallingford, UK, p 526

Gobbi A, Teixeira C, Moretzsohn M, Guimaraes P, Leal-Bertioli S, Bertioli D, Lopes CR, Gimenes M (2006) Development of a linkage map to species of B genome related to the peanut (*Arachis hypogaea* – AABB). Plant and animal genomes XIV conference, San Diego, CA, USA, P 679 (http://www.intl-pag.org/14/abstracts/PAG14_P679.html)

Goff SA, Salmeron JM (2004) Back to the future of cereals. Sci Am 291:42–48

Grusak, MA (2002) Enhancing mineral content in plant food products. J Am Coll Nutr 21:178S–183S

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

Gupta PK, Varshney RK (2004) Cereal genomics. Kluwer Academic Publishers, Dordrecht, The Netherlands

Gutierrez MV, Vaz Patto MC, Huguet T, Cubero JI, Moreno MT, Torres AM (2005) Cross-species amplification of *Medicago truncatula* microsatellites across three major pulse crops. Theor Appl Genet 110:1210–1217

Halward TM, Stalker HT, Kochert G (1993) Development of an RFLP linkage map in diploid peanut species. Theor Appl Genet 87:379–384

Halward TM, Stalker HT, LaRue E, Kochert G (1992) Use of single-primer DNA amplifications in genetic studies of peanut (*Arachis hypogaea* L.). Plant Mol Biol 18:315–325

Hannan RM, Kaiser WJ, Muehlbauer FJ (1994) Development and utilization of the USDA chickpea germplasm core collection. Agron Abstr, ASA, Madison, WI, p 217

Hansen M, Kraft T, Ganestam S, Sall T, Nilsson NO (2001) Linkage disequilibrium mapping of the bolting gene in sea beet using AFLP markers. Genet Res 77:61–66

He G, Prakash CS (1997) Identification of polymorphic DNA markers in cultivated peanut (*Arachis hypogaea* L.). Euphytica 97:143–149

He G, Meng R, Newman M, Gao G, Pittman RN, Prakash CS (2003) Microsatellites as DNA markers in cultivated peanut (*A. hypogaea* L.). BMC Plant Biol 3:3

Herselman L, Thwaites R, Kimmins FM, Courtois B, van der Merwe PJA, Seal SE (2004) Identification and mapping of AFLP markers linked to peanut (*Arachis hypogaea* L.) resistance to the aphid vector of groundnut rosette disease. Theor Appl Genet 109:1426–1433

Hill WG, Weir BS (1994) Maximum-likelihood estimation of gene location by linkage disequilibrium. Am J. Hum Genet 54:705–714

Hirschhorn JN, Daly MJ (2005) Genome-wide association studies for common diseases and complex traits. Nat Rev Genet 6:95–108

Holbrook CC, Anderson WF, Pittman RN (1993) Selection of a core collection from the US germplasm collection of peanut. Crop Sci 33: 859–861

Hopkins MS, Casa AM, Wang T, Mitchell SE, Dean RE, Kochert GD, Kresovich S (1999) Discovery and characterization of polymorphic simple sequence repeats (SSRs) in peanut. Crop Sci 39: 1243–1247

Hovav R, Upadhyaya KC, Beharav A, Abbo S (2003) Major flowering time gene and polygene effects on chickpea seed weight. Plant Breed 122:539–541

Huettel B, Winter P, Weising K, Choumane W, Weigand F, Kahl G (1999) Sequence-tagged microsatellite site markers for chickpea (*Cicer arietinum* L.). Genome 42:210–217

Igartua E, Casas AM, Ciudad F, Montoya JL, Romagosa I (1999) RFLP markers associated with major genes controlling heading date evaluated in a barley germplasm pool. Heredity 83:551–559

Iruela M, Rubio J, Barro F, Cubero JI, Millan T, Gil J (2006) Detection of two quantitative trait loci for resistance to ascochyta blight in an intra-specific cross of chickpea (*Cicer arietinum* L.): development of SCAR markers associated with resistance. Theor Appl Genet 112:278–287

Ivandic V, Thomas WTB, Nevo E, Zhang Z, Forster BP (2003) Associations of simple sequence repeats with quantitative trait variation including biotic and abiotic stress tolerance in *Hordeum spontaneum*. Plant Breed 122:300–304

Ivandic V, Hackett CA, Nevo E, Keith R, Thomas WTB, Forster BP (2002) Analysis of simple sequence repeats (SSRs) in wild barley from the fertile crescent: associations with ecology, geography, and flowering time. Plant Mol Biol 48:511–527

Jain SM, Brar DS, Ahloowalia BS (2002) Molecular techniques in crop improvement. Kluwer Academic Publishers, Dordrecht, The Netherlands

Jansen RC, Nap JP (2001) Genetical genomics: the added value from segregation. Trends Genet 17: 388–391

Jesubatham AM, Burrow MD (2006) Peanut Map: an online genome database for comparative molecular maps of peanut. BMC Bioinformatics 7:375

Jones N, Ougham H, Thomas H (1997) Markers and mapping: we are all geneticists now. New Phytol 137:165–177

Kashiwagi J, Krishnamurthy L, Upadhyaya HD, Krishna H, Chandra S, Vadez V, Serraj R (2004) Genetic variability of drought-avoidance root traits in the mini-core germplasm collection of chickpea (*Cicer arietinum* L.). Euphytica 146:213–222

Kashiwagi J, Krishnamurthy L, Crouch JH, Serraj R (2006) Variability of root length density and its contributions to seed yield in chickpea (*Cicer arietinum* L.) under terminal drought stress. Fields Crop Res 95:171–181

Katam R, Vasanthaiah HKN, Basha SM (2006) Differential expression of mRNA transcripts and proteins in leaf tissue of peanut genotypes to water stress. In: Plant and animal genome XIV conference, San Diego, CA, USA, P446 (www.intl-pag.org/14/abstracts/PAG14_P446.html)

Kazan K, Muehlbauer FJ, Weeden NE, Ladizinsky G (1993) Inheritance and linkage relationships of morphological and isozyme loci in chickpea (*Cicer arietinum* L.). Theor Appl Genet 86: 417–426

Killian A, Huttner E, Wenzl P, Jaccoud D, Carling J, Caig V, Evers M, Heller-Uszynska K, Uszynski G, Cayla C, Patarapuwadol S, Xia L, Yang S, Thomson B (2005) The fast and the cheap: SNP and DArT – based whole genome profiling for crop improvement. In: Proceedings of the international congress, 'In the wake of double helix: from the green revolution to the gene revolution', 27–31 May 2003, Bologna, Italy, pp 443–461

Kirst M, Yu Q (2007) Genetical genomics: successes and prospects in plants. In: Varshney RK, Tuberosa R (eds) Genomics assisted crop improvement Vol 1: genomic approaches and platforms, springer, The Netherlands, pp 245–265

Kochert G, Halward T, Branch WD, Simpson CE (1991) RFLP variability in peanut (*Arachis hypogaea* L.) cultivars and wild species. Theor Appl Genet 81:565–570

Koebner RMD (2004) Marker assisted selection in the cereals: the dream and the reality. In: Gupta PK, Varshney RK (eds) Cereal genomics. Kluwer Academic Publishers, Netherlands, pp 199–252

Koebner RMD, Varshney RK (2006) The development and application of genomic models for large crop plant genomes. In: Varshney RK, Koebner RMD (eds) Model plants and crop improvement CRC Press, FL pp 1–10

Kotresh H, Fakrudin B, Punnuri S, Rajkumar B, Thudi M, Paramesh H, Lohithswa H, Kuruvinashetti M (2006) Identification of two RAPD markers genetically linked to a recessive allele of a *Fusarium* wilt resistance gene in pigeonpea (*Cajanus cajan* (L.) Millsp.). Euphytica 149:113–120

Krapovikas A, Gregory WC (1994) Taxonomy of genus *Arachis* (Legumonisae). Bonplandia 8:1–186

Krishnamurthy L, Kashiwagi J, Upadhyaya HD, Serraj R (2003) Genetic diversity of drought avoidance root traits in the mini core germplasm collection of chickpea. Intl Chickpea Pigeonpea Newsl 10: 21–24

Kumar J, Van Rheenen HA (2000) A major gene for time of flowering in chickpea. J Hered 91:67–68

Lichtenzveig J, Scheuring C, Dodge J, Abbo S, Zhang HB (2005) Construction of BAC and BIBAC libraries and their applications for generation of SSR markers for genome analysis of chickpea (*Cicer arietinum* L.). Theor Appl Genet 110:492–510

Lodwig EM, Hosie AH, Bourdes A, Findlay K, Allaway D, Karunakaran R, Downie JA, Poole PS (2003) Amino-acid cycling drives nitrogen fixation in the legume–Rhizobium symbiosis. Nature 422:672–674

Luo M, Dang P, Guo BZ, He G, Holbrook CC, Bausher MG, Lee RD (2005) Generation of expressed sequence tags (ESTs) for gene discovery and marker development in cultivate peanut. Crop Sci 45:346–353

Luo M, Dang P, Guo BZ, Holbrook CC, Bausher M (2003) Application of EST technology in functional genomics of *Arachis hypogaea* L. Phytopathology 94:S55

Melchinger AE (1990) Use of molecular markers in breeding for oligogenic disease resistance. Plant Breed 104:1–19

Milla SR (2003) Relationships and utiligation of Arachis germplasm in peanut improvement. PhD Thesis, North Carolina State University, USA, pp 1–150

Milla SR, Isleib TG, Stalker HT (2005) Taxonomic relationships among *Arachis* sect. *Arachis* species as revealed by AFLP markers. Genome 48:1–11

Millan T, Rubio J, Iruela M (2003) Markers associated with Ascochyta blight resistance in chickpea and their potential in marker-assisted selection. Field Crops Res 84:373–384

Millan T, Clarke HJ, Siddique KHM, Buhariwalla HK, Gaur PM, Kumar J, Gil J, Kahl G, Winter P (2006) Chickpea molecular breeding: new tools and concepts. Euphytica 147:81–103

Moretzsohn MC, Hopkins MS, Mitchell SE, Resovich SK, Valls JFM, Ferreira ME (2004) Genetic diversity of peanut (*Arachis hypogaea* L.) and its wild relatives based on the analysis of hypervariable regions of the genome. BMC Plant Biol 4:11

Moretzsohn MC, Leoi L, Proite K, Guimara PM, Leal-Bertioli SCM, Gimenes MA, Martins WS, Valls JFM, Grattapaglia D, Bertioli DAJ (2005) Microsatellite-based, gene-rich linkage map for the AA genome of *Arachis* (Fabaceae). Theor Appl Genet 111:1060–1071

Morgante M, Salamini F (2003) From plant genomics to breeding practice. Curr Opin Biotech 14: 214–219

Nadimpalli RG, Jarret RL, Phatak SC, Kochert G (1994) Phylogenetic relationships of the pigeon pea (*Cajanus cajan* L.) based on nuclear restriction fragment length polymorphism. Genome 36:216–223

Nobile PM, Lopes CR, Barata T, Barsalobres C, Guimaraes P, Leal-Bertioli S, Gimenes M (2006) Identification and characterization of peanut (*Arachis hypogaea* L.) ESTs regulated during interaction with *Cercosporidium personatum* (Berk and Curt) Deighton In: XIV international plant and animal genome conference, San Diego, CA, USA, P680 (wwwintl-pagorg/14/abstracts/PAG14_P680html)

Odeny DA, Jayashree B, Ferguson M, Hoisington D, Crouch J, Gebhardt C (2007) Development, characterization and utilization of microsatellite markers in pigeonpea. Plant Breed 126:130–136

Or E, Hovav R, Abbo S (1999) A major gene for flowering time in chickpea. Crop Sci 39:315–322

Paik-Ro OG, Smith RL, Knauft DA (1992) Restriction fragment length polymorphism evaluation of six peanut species within the Arachis section. Theor Appl Genet 84:201–208

Pakniyat H, Powell W, Baired E, Handley LL, Robinson D, Scrimgeour CM, Novo E, Hackett CA, Caligari PDS, Foster BP (1997) AFLP variation in wild barley (*Hordeum spontaneum* C Koch) with reference to salt tolerance and associated ecogeography. Genome 40:332–341

Palmieri DA, Bechara MD, Curi RA, Gimenes MA, Lopes CR (2005) Novel polymorphic microsatellite markers in section *Caulorrhizae* (*Arachis*, Fabaceae). Mol Ecol Notes 5:77–79

Paran I, Mitchelmore RW (1993) Development of reliable PCR-based markers linked to downy mildew resistance gene in lettuce. Theor Appl Genet 85:985–993

Passioura J (1977) Grain yield, harvest index and water use of wheat. J Aust Inst Agric Sci 43:117–120

Pereira MG, de Sousa MMT, Duarte IM (2001) Status of European *Cicer* database In: Magioni L, Ambrose M, Schachl R, Duc G, Lipman E (compilers) Report of a working group on grain legumes. Third Meeting, 5–7 July 2001, Krakow, Poland IPGRI, Rome, Italy, pp 45–46

Pfaff T, Kahl G (2003) Mapping of gene-specific markers on the genetic map of chickpea (*Cicer arietinum* L). Mol Gen Genomics 269:243–251

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

Punguluri SK, Janaiah K, Govil JN, Kumar PA, Sharma PC (2006) AFLP fingerprinting in pigeonpea (*Cajanus cajan* (L.) Millsp) and its wild relatives. Genet Resour Crop Evol 53:523–531

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

Raina SN, Rani V, Kojima T, OgiharaY, Singh KP (2001) RAPD and ISSR fingerprints as useful genetic markers for analysis of genetic diversity, varietal identification, and phylogenetic relationships in peanut (*Arachis hypogaea* L.) cultivars and wild species. Genome 44:763–772

Rajesh PN, Tullu A, Gil J, Gupta VS, Ranjekar PK, Muehlbauer FJ (2002) Identification of an STMS marker for the double-podding gene in chickpea. Theor Appl Genet 105:604–607

Rajesh PN, Coyne C, Meksem K, Sharma KD, Gupta V, Muehlbauer FJ (2004) Construction of a *Hind* III bacterial artificial chromosome library and its use in identification of clones associated with disease resistance in chickpea. Theor Appl Genet 108:663–669

Ratnaparkhe MB, Gupta VS, Ven Murthy MR, Ranjekar PK (1995) Genetic fingerprinting of pigeonpea [*Cajanus cajan* (L.) Millsp] and its wild relatives using RAPD markers. Theor Appl Genet 91:893–898

Reddy LJ, Upadhyaya HD, Gowda CLL, Singh S (2005) Development of core collection in pigeonpea (*Cajanus cajan* (L.) Millasp). Genet Resour Crop Evol 52:1049–1056

Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc Natl Acad Sci USA 25: 11479–11484

Ryan JG, Spencer DC (2001) Future challenges and opportunities for agricultural R&D in the semi-arid tropics, International crops research institute for the semi-arid tropics, Patancheru 502 324, Andhra Pradesh, India, p 83

Santra DK, Tekeoglu M, Ratnaparkhe MB, Gupta VS, Ranjekar PK, Muehlbauer FJ (2000) Identification and mapping of QTLs conferring resistance to Ascochyta blight in chickpea. Crop Sci 40: 1606–1612

Saxena NP (1984) Chickpea in the physiology of tropical field crops. In: Goldworthy PR, Fisher NM (eds) Wiley, New York, pp 419–452

Saxena KB, Singh L, Ariyanagam RP (1993) Role of partially cleistogamy in maintaining genetic purity of pigeonpea. Euphytica 66:225–229

Schadt EE, Monks SA, Drake TA, Lusis AJ, Che N, Colinayo V, Ruff T G, Milligan SB, Lamb JR, Cavet G, Linsley PS, Mao M, Stoughton RB, Friend SH (2003) Genetics of gene expression surveyed in maize, mouse and man. Nature 422:297–301

Schultze M, Kondorosi A (1998) Regulation of symbiotic root nodule development. Annu Rev Genet 32:33–57

Septiningsih EM, Trijatmiko KR, Moeljopawiro S, McCouch SR (2003) Identification of quantitative trait loci for grain quality in an advanced backcross population derived from the *Oryza sativa* variety IR64 and the wild relative *O rufipogon*. Theor Appl Genet 107:1433–1441

Serraj R (2004) Symbiotic nitrogen fixation:challenges and future prospects for application in tropical agroecosystems. Oxford & IBH, New Delhi, India

Serraj R, Krishnamurthy L, Upadhyaya HD (2004) Screening chickpea mini-core germplasm for tolerance to salinity. Intl Chickpea Pigeonpea Newsl 11:29–32

Sethy NK, Shokeen B, Bhatia S (2003) Isolation and characterization of sequence-tagged microsatellite sites markers in chickpea (*Cicer aeritinum* L.). Mol Ecol Notes 3:428–430

Sethy NK, Choudhary S, Shokeen B, Bhatia S (2006a) Identification of microsatellite markers from *Cicer reticulatum*:molecular variation and phylogenetic analysis. Theor Appl Genet 112:347–357

Sethy NK, Shokeen B, Edwards KJ, Bhatia S (2006b) Development of microsatellite markers and analysis of intraspecific genetic variability in chickpea (*Cicer arietinum* L.). Theor Appl Genet 112:1416–1428

Sharma HC (2001) Cytoplasmic male-sterility and source of pollen influence the expression of resistance to sorghum midge, *Stenodiplosis sorghicola.* Euphytica 122:391–395

Sharma KK, Lavanya M (2002) Recent developments in transgenics for abiotic stress in legumes of the semi-arid tropics In: Ivanaga M (ed) Genetic engineering of crop plants for abiotic stress, JIRCAS Working Report No 23:61–73 Tsukuba, Japan

Sharma KD, Winter P, Kahl G, Muehlbauer FJ (2004) Molecular mapping of *Fusarium oxysporum* f.sp. *ciceris* race 3 resistance gene in chickpea. Theor Appl Genet 108:1243–1248

Sharma PC, Matsumura H, Terauchi R (2007) Use of serial analysis of gene expression (SAGE) for transcript analysis. In: Varshney RK, Tuberosa R (eds) Genomics assisted crop improvement Vol I: genomics approaches and platforms, Springer, the Netherlands, pp 227–244

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

Simon CJ, Muehlbauer FJ (1997) Construction of a chickpea linkage map and its comparison with maps of pea and lentil. J Hered 88:115–119

Simpson CE, Nelson SC, Starr JL, Woodard KE, Smith OD (1993) Registration of TxAg-6 and TxAg-7 peanut germplasm lines. Crop Sci 33:1418

Singh KB, Kumar J, Haware MP, Lateef SS (1990) Disease and pest resistance breeding: which way to go in the nineties. In Chickpea in the nineties: Proceeding of the second international workshop on Chickpea improvement, 4–8 December 1989, International crops research institute for the semi-arid tropics, Patancheru 502 324, Andhra Pradesh, India, pp 223–238

Sivaramakrishnan S, Seetha K, Nageshwar Rao A, Singh L (1997) RFLP analysis of cytoplasmic male-sterile lines of pigeonpea [*Cajanus cajan* (L.) Millsp] developed by interspecific crosses. Euphytica 93:307–312

Skot L, Hamilton NRS, Mizen S, Chorlton KH, Thomas ID (2002) Molecular genecology of temperature response in *Lolium perenne*: 2 Association of AFLP markers with ecogeography. Mol Ecol 11: 1865–1876

Stalker HT, Mozingo LG (2001) Molecular markers of *Arachis* and marker assisted selection. Peanut Sci 28:117–123

Subramanian V, Gurtu S, Nageswara Rao RC, Nigam SN (2000) Identification of DNA polymorphism in cultivated groundnut using random amplified polymorphic DNA (RAPD) assay. Genome 43:656–660

Sun GL, William M, Liu J, Kasha KJ, Pauls KP (2001) Microsatellite and RAPD polymorphisms in Ontario corn hybrids are related to the commercial sources and maturity ratings. Mol Breed 7:13–24

Sun G, Bond M, Nass H, Martin R, Dong Z (2003) RAPD polymorphisms in spring wheat cultivars and lines with different level of *Fusarium* resistance. Theor Appl Genet 106:1059–1067

Tanksley SD, McCouch SR (1997) Seed banks and molecular maps: unlocking genetic potential from the wild. Science 277:1063–1066

Tanksley SD, Nelson JC (1996) Advance backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTL from unadapted germplasm to the elite breeding lines. Theor Appl Genet 92:191–203

Tanksley SD, Young ND, Paterson AH, Bonierbale MW (1989) RFLP mapping in plant breeding: new tools for an old science. Biotechnology 7:257–264

Tautz D (1989) Hypervariablity of simple sequences as a general source of polymorphic DNA markers. Nucl Acids Res 17:6463–6471

Tekeoglu M, Rajesh PN, Muehlbauer FJ (2002) Integration of sequence tagged microsatellite sites to the chickpea genetic map. Theor Appl Genet 105:847–854

Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler IV ES (2001) *Dwarf8* polymorphisms associate with variation in flowering time. Nature Genet 28:286–289

Thro AM, Parrott W, Udall JA, Beavis WD (2004) Genomics and plant breeding: the experience of the initiative for future agricultural and food systems. Crop Sci 44:1893–1919

Till BJ, Comai L, Henikoff S (2007) TILLING and EcoTILLING for crop improvement. In: Varshney RK, Tuberosa R (eds) Genomics assisted crop improvement Vol I: Genomics Approaches and Platforms, Springer, The Netherlands, pp 333–350

Tohme J, Gonzalez OD, Beebe S, Duque MC (1996) AFLP analysis of gene pools of a wild bean core collection. Crop Sci 36:1375–1384

Turpeinen T, Tenhola T, Manninen O, Nevo E, Nissila E (2001) Microsatellite diversity associated with ecological factors in *Hordeum spontaneum* populations in Israel. Mol Ecol 10:1577–1591

Udupa SM, Baum M (2003) Genetic dissection of pathotype-specific resistance to ascochyta blight disease in chickpea (*Cicer arietinum* L.) using microsatellite markers. Theor Appl Genet 106: 1196–1202

Udupa SM, Sharma A, Sharma AP, Pai RA (1993) Narrow genetic variability in *Cicer arietinum* L. as revealed by RFLP analysis. J. Plant Biochem Biotechnol 2:83–86

Upadhyaya HD, Ortiz R (2001) A mini core subset for capturing diversity and promoting utilization of chickpea genetic resources. Theor Appl Genet 102:1292–1298

Upadhyaya HD, Bramel PJ, Singh S (2001a) Development of a chickpea core subset using geographic distribution and quantitative traits. Crop Sci 41:206–210

Upadhyaya HD, Nigam SN, Singh S (2001b) Evaluation of groundnut core collection to identify sources of tolerance to low temperature at germination. Indian J Plant Genet Resour 14:165–167

Upadhyaya HD, Bramel PJ, Ortiz R, Singh S (2002) Developing a mini core of peanut for utilization of genetic resources. Crop Sci 42:2150–2156

Upadhyaya HD, Ortiz R, Bramel PJ, Singh S (2003) Development of a groundnut core collection using taxonomical, geographical and morphological descriptors. Genet Resour Crop Evol 50:139–148

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

Upadhyaya HD, Reddy LJ, Gowda CLL, Reddy KN, Singh S (2006b) Development of a mini core subset for enhanced and diversified utilization of pigeonpea germplasm resources. Crop Sci 46:2127–2132

Vadez V, Krishnamurthy L, Gaur PM, Upadhyaya HD, Hoisington DA, Varshney RK, Turner NC, Siddique KHM (2006) Tapping the large genetic variability for salinity tolerance in chickpea. Proceeding of the Australian society of agronomy meeting, 10–14 September 2006 (http://wwwregionalorgau/au/asa/2006/ concurrent/environment/4561_vadezhtm)

Van der Maesen, LJG (1987) Origin, history, and taxonomy of chickpea. In: Saxena MC, Singh KB (eds) The Chickpea, CABI/ICARDA, Wallingford, UK, pp 11–43

van Treuren R, van Soest LJM, van Hintum ThJL (2001) Marker-assisted rationalization of genetic resource collections: a case study in flax using AFLPs. Theor Appl Genet 103:144–152

Vance CP, Graham PH, Allan DL (2000) Biological nitrogen fixation phosphorus: a critical future need In: Pedrosa FO, Hungria M, Yates MG, Newton WE (eds) Nitrogen fixation: from molecules to crop productivity. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp 506–514

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

Varshney RK, Hoisington D, Tyagi AK (2006) Advances in cereal genomics and applications in crop breeding. Trends Biotechnol 24:490–499

Velculescu VE, Hang LZ, Vogelstein B, Kinzler KW (1995) Serial analysis of gene expression. Science 270:484–487

Virk PS, Newbury HJ, Jackson MT, Ford-Lloyd BV (1995) The identification of duplicate accessions within a rice germplasm collection using RAPD analysis. Theor Appl Genet 90:1049–1055

Virk PS, Ford-Lloyd BV, Jackson MT, Pooni HS, Clemeno TP, Newbury HJ (1996) Predicting quantitative variation within rice germplasm using molecular markers. Heredity 76:296–304

Vos P, Hogers R, Bleeker M, Reijans M, Lee T van de, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M (1995) AFLP: a new technique for DNA fingerprinting. Nucl Acids Res 23:4407–4414

Wang D, Graef GL, Procopiuk AM, Diers BW (2004a) Identification of putative QTL that underlie yield in interspecific soybean backcross populations. Theor Appl Genet 108:458–467

Wang ML, Gillaspie AG, Newman ML, Dean RE, Pittman RN, Morris JB, Pederson GA (2004b) Transfer of simple sequence repeat (SSR) markers across the legume family for germplasm characterization and evaluation. Plant Genet Resour 2:107–119

Williams JGK, Kubelik ARK, Livak JL, Rafalski JA, Tingey SV (1990) DNA polymorphisms amplified by random primers are useful as genetic markers. Nucl Acids Res 18:6531–6535

Winter P, Pfaff T, Udupa SM, Huttel B, Sharma PC, Sahi S, Arreguin-Espinoza R, Weigand F, Muehlbauer FJ, Kahl G (1999) Characterization and mapping of sequence-tagged microsatellite sites in the chickpea (*Cicer arietinum* L.) genome. Mol Gen Genet 262:90–101

Winter P, Benko-Iseppon AM, Huttel B, Ratnaparkhe M, Tullu A, Sonnante G, Pfaff T, Tekeoglu M, Santra D, Sant VJ, Rajesh PN, Kahl G, Muehlbauer FJ (2000) A linkage map of chickpea (*Cicer arietinum* L.) genome based on recombinant inbred lines from a *C. arietinum* × *C. reticulatum* cross: localization of resistance genes for Fusarium wilt races 4 and 5. Theor Appl Genet 101: 1155–1168

Yang S, Pang W, Ash G, Harper J, Carling J, Wenzl P, Huttner E, Zong X, Kilian A (2006) Low level of genetic diversity in cultivated Pigeonpea compared to its wild relatives is revealed by diversity arrays technology. Theor Appl Genet 113:585–595

Young ND, Mudge J, Ellis THN (2003) Legume genomes: more than peas in a nod. Curr Opin Plant Biol 6:199–204

Young ND, Cannon SB, Sato S, Kim D, Cook DR, Town CD, Roe BA, Tabata S (2005) Sequencing the genespaces of *Medicago truncatula* and *Lotus japonicus*. Plant Physiol 137:1174–1181

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

Yu J, Pressoir G, Briggs WH, Vroh BI, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203–208

Yüksel B, Paterson AH (2005) Construction and characterization of a peanut *Hin*dIII BAC library. Theor Appl Genet 111:630–639

Yüksel B, Bowers JE, Estill J, Goff L, Lemke C, Paterson AH (2005) Exploratory integration of peanut genetic and physical maps and possible contributions from *Arabidopsis*. Theor Appl Genet 111: 87–94

Zhu YL, Song QJ, Hyten DL, Van Tassell CP, Matukumalli LK, Grimm DR, Hyatt SM, Fickus EW, Young ND, Cregan PB (2003) Single-nucleotide polymorphisms in soybean. Genetics 163:1123–1134

CHAPTER 11

GENOMICS APPROACHES TO SOYBEAN IMPROVEMENT

TRI D. VUONG, XIAOLEI WU, MD S. PATHAN, BABU VALLIYODAN AND HENRY T. NGUYEN*

National Center for Soybean Biotechnology and Division of Plant Sciences, University of Missouri, Columbia, MO 65211, USA

Abstract: Soybean, *Glycine max* (L.) Merr., has become the major source of edible vegetable oils and high protein feeds for livestocks in the world. A native of Eastern Asia, soybean was introduced into the USA and South America where it has become the most economically important agricultural crop and export commodity. In recent years, as demand for soybean increased due to the values of seed oil and protein, as well as industrial and nutriceutical uses, it has received more attention by scientists aiming to the development and employment of genomic technology for soybean improvement. Several DNA marker systems, such as restriction fragment length polymorphism (RFLP), simple sequence repeat (SSR), and single nucleotide polymorphism (SNP), were integrated into the soybean genetic linkage map, which has been successfully utilized for mapping quantitative trait loci (QTL) linked to desirable traits and marker-assisted breeding of disease resistance and seed composition. The availability of a large number of expressed sequence tags (EST) and BAC sequences facilitated the discovery of new SNP and SSR markers in soybean toward the construction of high resolution genetic maps. Integrated genetic and physical maps will provide an invaluable resource for gene identification and positional cloning of important quantitative trait loci in soybean. Functional genomics has emerged as a new and rapidly evolved discipline to identify and understand gene functions via an integrated approach which includes transcriptomics, proteomics, metabolomics, translational genomics, and bioinformatics. The completion of whole soybean genome sequencing is anticipated in a few years. The availability of the soybean genome sequences in combination with the integrated genetic and physical maps will be invaluable resources providing soybean researchers powerful and efficient genomic tools to identify and characterize genes or QTLs for agronomic traits of soybean. As a result, it facilitates marker-assisted breeding and soybean improvement.

*Author for correspondence: nguyenhenry@missouri.edu

R.K. Varshney and R. Tuberosa (eds.), Genomics Assisted Crop Improvement: Vol. 2: Genomics Applications in Crops, 243–279.

1. INTRODUCTION

Soybean [*Glycine max* (L.) Merr.] is an important leguminous crop producing the most valuable source of protein and vegetable oil closest to the optimum dietary essential amino acids profiles for human and animal nutrition (Lusas 2004). Domestication of the crop likely occurred between the 11th and 15th century BC, or perhaps earlier, from northeast China, and subsequently reached China's southern region and neighboring countries such as Korea, Japan, Thailand, the Philippines, etc. (Hymowitz 2004). It was not until the 18th century when soybean was introduced into Europe. In the western hemisphere, soybean was first brought into North America in 1765 and then into Central and South America during the mid 1900s (Hymowitz 2004). Since then, soybean has become a major economic crop for soybean producing areas.

Worldwide, the expansion in soybean production is likely due to the increased value of protein and oil in soybean seeds. Among over 19,000 soybean accessions of USDA-ARS Soybean Germplasm Collection (http://www.ars-grin.gov/cgi-bin/npgs/html/stats/genus.pl?Glycine), it was reported that soybean has a wide range of genetic variation in protein and oil contents. The range of protein content is from 34.1 to 56.8% of dry seed mass, with a mean of 42.1%, while the range of oil content is 8.3 to 27.9%, with a mean of 19.5%. Between these two seed composition traits, there exists a widely known, strong negative correlation (Hurburgh et al. 1990), indicating that efforts to improve protein content often results in decreased oil content, and vice versa. It is extremely rare to find germplasm in which the content for both protein and oil is relatively high (Wilson 2004). Thus, soybean workers who have extended great efforts to improve these important seed traits normally accept a sacrifice of one trait while attempting to improve the other.

The soybean production in the world has been steadily increasing for the last ten years, rising from 127 million metric tons in 1995 to over 209 million metric tons in 2005 (FAO, 2006, http://faostat.fao.org). Among the countries growing soybean, the United States led production at 82.8 millions tons in 2005, equaling approximately 40% of the world's total soybean production. The other two major soybean producing countries exist in South America. Brazil and Argentina produced 50 and 38 million tons, respectively, and accounted for approximately 42% of the world's total production in 2005 (www.soystats.com). It is noted that although the soybean is a native of China and has been widely utilized as valuable protein source in people's diets in many other countries of Asia, soybean production in this area was only 11% of the world total. A 2005 report indicated that China produced 17 million tons (8% of the world total) with an average seed yield of 1.78 t/ha and India produced 6.6 million tons (3% of the world total) with an average yield of 0.96 t/ha (FAO, 2006, http://faostat.fao.org). In the European Union and Eastern Europe, the two largest soybean producers are Italy and Russian Federation. Each produced less than a half million tons, or 0.5% of the world total.

Since domestication, soybean seeds have been an essential and dominant protein source for human nutrition in Asian countries. Several nutritious foods were

developed from soybean. Among these, miso, soybean sauce, soybean milk, tempeh, and tofu were the most traditional soybean-based foods (Hymowitz 2004). In recent years, efforts have been made to validate nutritional and pharmaceutical values of soybean protein. In nutrition studies, Birt et al. (2004) reported that consumption of soybean-based foods reduces cancer, blood serum cholesterol, osteoporosis and heart disease. Thus, soybean products have been recommended as a source of supplemental protein for the human diet in the United States (http://www.soyfoodsillinois.uiuc.edu). Besides human consumption, soybean is also an important protein source for animal feed. In 2004, soybean provided the source for 67% of the world's protein meal (Lucas 2004). In the USA, protein meal production has steadily increased over the last ten years (www.soystats.com) due to increasing demand of soybean meals for the livestock and poultry industries.

Among oilseed crop plants such as peanut, rapeseed, and sunflower, soybean provided 57% of the world's total oilseed production (www.soystats.com). Preference for soybean oil may largely be due to the fact that soybean seed has a different composition than other legumes having relatively high levels of protein and oil and lower overall carbohydrate content. The oil portion of soybean is composed primarily of five fatty acids. Palmitic and stearic acids are saturated fatty acids and comprise 15% of the oil. In addition, soybean oil is rich in unsaturated fatty acids such as oleic, linoleic, and linolenic, which make up 85% of the oil in seed. Soybean is also a good source of the minerals, vitamin B, folic acid, and isoflavones which are attributed to slowing cancer development, heart disease, and osteoporosis (Wilson 2004).

In addition to valuable protein and oil source for human foods and animal feed, soybean also has numerous industrial uses and applications. Since the 1920s many US laboratories and universities have investigated potential industrial applications of soybean meals and oil. A large spectrum of soybean-based industrial uses have been adopted, for instance, building and construction materials, cleaners, concrete, plastics, paper, engine oils, printing inks, paints, lubricants, hydraulic fluids, pesticides, cosmetics, pharmaceuticals (Lusas 2004). In particular, since the late 1970s fatty acids in soybean oil were employed for biodiesel development, in which over 113 millions liters were consumed in the US in 2004 (www.soystats.com).

2. MOLECULAR MARKER SYSTEMS AND GENOMIC MAP

2.1. Taxonomy and Cytogenetics

Soybean was first described by Linnaeus (1737) as *Phaseolus max* based on his specimen and *Dolichos soja* based upon descriptions of other writers. Several years later, he realized that these two designations were actually the same plant. Since then, the correct nomenclature of soybean has been a subject of long debates among plant taxonomists (Hymowitz and Singh 1987). After thoroughly re-examining specimens of the soybean, Merrill (1917) proposed *Glycine max* as a new designation for soybean and this nomenclature has to date been widely

accepted. The genus *Glycine* is divided into two subgenera *Glycine* (wild perennials) and *Soja* (annuals).

Perennial subgenus *Glycine* has 22 species with chromosome number (2n) varying from 40 to 80 and geographical distribution mainly in Australia and neighboring islands, such as Indonesia, Papua New Guinea (Hymowitz 2004). Of these, species *G. tomentella* with four cytotypes (2n = 38, 40, 78, 80) and widely distributed in Asia and Australia has been intensively employed in soybean breeding programs due to valuable gene resources and cross-compatibility (Singh and Hymowitz 1999).

Annual subgenus *Soja* has two species, wild annual soybean *Glycine soja* and cultivated soybean *Glycine max*. Both species have the same chromosome number (2n = 40) and genome designation (GG) (Hymowitz 2004). Ohashi et al. (1984) proposed the scientific name of the wild annual soybean to be changed to *G. max* (L.) Merr. subsp. *soja* (Sieb. and Zucc.) Ohashi. Yet, the proposal in new nomenclature has not been popularly adapted. The US National Plant Germplasm System (USDA-GRIN, http://www.ars.grin.gov) and Hymowitz lab still remain the original nomenclature, *G. soja*, for this wild annual soybean (Hymowitz 2004).

The cultivated soybean, *G. max*, is a true domesticate and morphologically extremely variable in leaf shape, flower color, seed coat color, seed size and weight. This diversity of the cultivated soybean is perhaps due to the development of land races taking place for thousand years in East Asia. These land races containing specific traits have been grown by individual farmer families for generations for various uses including food, animal feed, and medicinal herbs. Today, these land races provide invaluable sources of generic diversity within soybean germplasm collections (Hymowitz 2004).

In addition to these two annual soybean species, a third form known as *G. gracilis* has been proposed (Skvortzow 1927) and a subject of long debates. Many studies showed that this new proposed species was a variant of *G. max* (Hermann 1962; Shoemaker et al.1986) based upon the fact that there was no cross-ability barrier among *G. soja*, *G. max* and *G. gracilis*; hybrid seeds germinated normally; and F1 plants were totally pollen- and seed-fertile (Singh and Hymowitz 1989). Subsequently, Wu et al. (2001) used SSR markers to evaluated genetic diversity among *Glycine* species and suggested that *G. garcilis* should be a separate species from *G. max*. However, neither the International Legume Database (ILDIS, http://www.ildis.org) nor USDA-GRIN recognizes *G. garcilis* as a separate species.

Both cultivated soybean and its annual relative, *G. soja*, have a diploid chromosome number of 2n = 40. However, Hymowitz et al. (1998) reported most genera in the *Phaseolae* tribe have a genome complement of 2n = 22, suggesting that *Glycine* was probably derived from a diploid ancestor with n = 11. It was believed that through thousands of years of evolution aneuploid loss occurred, yielding n = 10, followed by polyploidization events caused by additions, deletions, mutations, and rearrangements to generate a current diploid number of 2n = 40 (Lackey 1980). In spite of polyploidization most regions of soybean genome act like diploid. By using primary trisomics approach, in which one additional chromosome of genome, 2x+1=41, was generated for each of 20 chromosomes, soybean

cytogeneticists were able to locate genes onto a specific chromosome and associate molecular linkage groups (MLGs) with specific chromosomes. With a subset of 20 trisomics (Xu et al. 2000) and SSR markers, Zou et al. (2003) assigned 11 LGs to respective soybean chromosomes. However, this approach has encountered difficulty in making crosses, generating a large numbers of hybrids, counting and identifying chromosomes. It is anticipated that with bacterial artificial chromosome (BAC) clones and fluorescent in situ hybridization (FISH) approach (Hans and Jackson 2006), existing MLGs can soon be assigned to 20 corresponding chromosomes of soybean.

The soybean genome consists of ~1100 Mbp and is relatively larger than the genome of Arabidopsis and rice, but much smaller than the genome of maize and barley (Arumuganathan and Earle 1991). It is believed that due to the polyploidization event(s) occurring in an ancient ancestor, the soybean genome has a high percentage of internal duplication regions distributed among the chromosomes (Pagel et al. 2004). When using RFLP probes in a mapping study, Shoemaker et al. (1996) reported more than half of the RFLP probes detected multiple hybridizing fragments, indicating the ancient polyploid nature and duplication segments in the soybean genome. With the construction of BAC clones and the FISH technique, Hans and Jackson (2006) were able to anchor BAC clones to soybean linkage groups and identify duplicated regions of the soybean genome. However, the polyploid nature and extent of segmental duplications in the soybean genome result in a number of drawbacks for efforts to conduct whole genome sequencing and assembly due to repetitive DNA sequences.

2.2. Molecular Marker Systems

Development of new soybean varieties through conventional breeding is time and labor intensive and also takes large amounts of space in the greenhouse and field for evaluation. These constraints limit the number of plants which can be evaluated in the field each season and it takes about five to eight years to develop a variety. Breeders select traits of interest based on phenotypic performances and most of the traits are influenced by the environment. Dekkers and Hospital (2002) mentioned that selection based on genotype will greatly increase breeding efficiency. Plant breeders are interested in new technologies that are cost effective and improve efficiency of crop breeding. With the advancement of molecular marker technology, selection is based on DNA markers which are tightly linked with a trait(s)/gene(s) of interest rather than phenotype alone. Molecular markers have complemented the traditional breeding and have increased efficiency and reliability in breeding programs.

During the last decade, development of several molecular marker systems facilitated identification of genetic variation present in genomic DNA sequences of various crop plants, including soybean. Among these systems, restriction fragment length polymorphism (RFLP) (Botstein et al. 1980), random amplified polymorphic

DNA (RAPD) (Williams et al. 1990; Welsh and McClelland 1990), DNA amplification fingerprinting (DAF) (Caetano-Anolles et al. 1991), amplified fragment length polymorphism (AFLP) (Vos et al. 1995); and simple sequence repeat (SSR) or microsatellites (Litt and Luty 1989) are most commonly employed in soybean molecular genetic analysis. Recently, significant efforts are directed forward SNP marker development.

2.2.1. Restriction fragment length polymorphism

As the first generation of molecular markers to be used in animal and plant molecular genetic mapping, RFLPs are caused by changes in DNA sequence of genome, such as mutations, insertions, or deletions of DNA fragments. These changes result in the gain, loss, or movement of some restriction sites, which is the basis for generating RFLPs. These DNA markers are bi-allelic codominant; thus, a unique locus and chromosomal position can be precisely identified with a specific probe that could be either genomic DNA fragment, cDNA, or expressed sequence tags (ESTs). RFLP markers are robust and accurate markers for genomic genotyping; this genotyping technique, however, is a labor-intensive and expensive procedure. Therefore, many improvements have been made for the RFLP analysis to be a cost-efficient, high-throughput approach (Nguyen and Wu 2005).

In soybean, RFLPs were first utilized as DNA probes to differentiate soybean cultivars (Apuya et al. 1988, Keim et al. 1989). Later, Keim et al. (1990) published the first RFLP-based map of the soybean genome using an F_2 mapping population derived from a cross of *G. max* (A81-356022) and *G. soja* (PI468916). The mapping work was subsequently expanded by Shoemaker and Olson (1993) with the addition of 355 RFLP loci. In a separate investigation, Rafalski and Tingey (1993) mapped more than 600 RFLP loci in a different cross of cultivated soybean and wild soybean. It was reported the level of RFLP polymorphism was relatively low, 33%, despite the wide cross of cultivated and wild soybean employed (Shoemaker et al. 2004). Moreover, the duplicated nature of the soybean genome is an additional factor confounding genotype analyses in soybean. Despite complications, a number of studies were conducted using RFLP markers to characterize genetic variation in soybean germplasm and map quantitative trait loci (QTL) for agronomic traits (Keim et al. 1990; Concibido et al. 1996; Lee et al. 1996; Mian et al. 1998). To date, 709 RFLP loci were mapped to 20 MLGs of soybean genome (Song et al. 2004).

2.2.2. Random amplified polymorphic DNA

The advent of the polymerase chain reaction (PCR) technology facilitated the subsequent development of many PCR-based DNA marker systems for genetic mapping and DNA fingerprinting. Among these, RAPD marker was the pioneer of the second generation of PCR-based molecular markers. The assay requires no prior knowledge of DNA sequence for detection of genetic variation and amplified fragments can be simply separated in agarose gel electrophoresis. For PCR amplification, a single arbitrary primer is added to a reaction. The primer randomly anneals to homologous sequences in the genome. Amplicons throughout the genome are

targeted and amplified if the primer also anneals to sequences on complementary strands not far from the 3' end of the other primers (Welsh and McCelland 1990). Because the primers are relatively short in sequence (10 nucleotides), they possibly hybridize to genomic sites at which they fortuitously match or almost match; as a result, multiple products with relatively complex pattern are produced under conditions of low stringency with a single random primer. However, the RAPD assay has disadvantages in the procedure. For instance, there is a problem of reproducibility within and between laboratories over time, making the marker information difficult to share and repeat (Nguyen and Wu 2005). With these drawbacks, RAPD has not been widely employed for genetic mapping and genomic fingerprinting. In soybean, there have been attempts to use RAPD assay for construction of genetic map. Ferreira et al. (2000) incorporated 106 RAPD markers into a framework of 250 existing RFLP loci using recombinant inbred lines from a cross of cultivated soybeans.

2.2.3. Amplified fragment length polymorphism

Combining the advantages of RFLP analysis and PCR technology, the AFLP assay was developed (Vos et al. 1995) and widely used for genetic map and genomic fingerprinting. Like RAPD markers, the AFLP assay requires no prior knowledge of genomic DNA sequence. Restriction sites of genomic DNA are digested with two restriction endonucleases to generate restriction fragments. The DNA adaptors are ligated to the ends of the restriction fragments and PCR amplifications are performed to amplify selected subpopulations of the pool of fragments. Selectivity results from the addition of two or three arbitrary nucleotides to the 3' ends of the PCR primers. Between these primers, the forward primer is labeled with radioactive ^{33}P. The amplified restriction fragments can be separated on denaturing polyacrylamide gels and visualized by means of autoradiography (Vos et al. 1995) or silver staining (Chalhoub et al. 1997). When PCR products are labeled with different fluorescent dyes, the amplified restriction fragments can be visualized with automated DNA sequencer (Myburg et al. 2001).

In soybean, AFLP markers were used to construct a genetic map using a subset of 42 RILs from a cross of cultivated soybean (Keim et al. 1997). This map has a total of 650 AFLP loci mapped to several linkage groups. One of the drawbacks of AFLP markers is the clustering of high density markers resulted from the use of *Eco*RI/*Mse*I restriction enzymes (Young et al. 1999). However, marker clustering appeared to be eliminated when AFLP markers were generated with *Pst*I/*Mse*I restriction enzymes (Keim et al. 1997). Because the large amount of marker data can be generated with a single AFLP assay without the need for prior knowledge of DNA sequence, AFLPs have proven useful for saturating specific genomic regions using bulked segregant analysis (Michelmore et al. 1991) or for comparison of near-isogenic lines for a trait of interest (Muehlbauer et al. 1991).

2.2.4. Simple sequence repeat

Simple sequence repeat or microsatellite markers (Litt and Luty 1989), the most popular PCR-based DNA markers, were developed based upon genetic variation

in the number of repeat units (motifs), which are composed of 1∼6-bp short DNA sequences, such as dinucleotide repeats (AT)*n* or (CT)*n*, and trinucleotide repeats (ATT)*n* (Li et al. 2002). It has been shown that the SSR markers were abundant, highly polymorphic, codominant and distinguish multiple alleles within a plant species (Cregan et al. 1999a; Cho et al. 2000; Eujayl et al. 2002; Thiel et al. 2003). These characteristics make SSR technology an ideal marker system for DNA fingerprinting, genetic mapping, studies of genetic diversity, population genetics, and marker-assisted selection (MAS). However, the SSR technology has major drawbacks of high cost of development, prior knowledge of the flanking sequence for designing locus-specific PCR markers. Moreover, the difficulty, which frequently challenges researchers, is to distinguish alleles that differs only one or a few repeat units in size. In order to overcome these technical difficulties, alternative approaches of allelic variation detection are recommended, for instance, radioisotope is used to label one of primers and PCR products are separated in denaturing polyacrylamide gels (Shan et al. 1999). Currently, with the development of fluorescence-based genotyping system, the PCR products that are labeled with different fluorescent dyes can be multiplexed based upon allele size range and dyes for semi-automated detection using capillary electrophoresis DNA sequencer systems.

In soybean, Akkaya et al. (1992) and Morgante and Olivieri (1993) reported the discovery of several SSR markers in a set of 38 *G. max* and five *G. soja*. These authors also demonstrated attractive features of the molecular marker system, such as high levels of polymorphisms, codominant inheritance, and locus specificity, which have been also found in many other plant genomes (Broun and Tanksley 1996; Sanchez de le Hoz et al. 1996; Lu et al. 2005). Subsequently, several reports (Cregan et al. 1994; Maughan et al. 1995; Morgante et al. 1994; Rongwen et al. 1995) detected very high levels of allelic variation, for instance one locus with 26 alleles among groups of 91 cultivated soybean and five wild soybean genotypes. Since then, the SSR markers have provided an excellent complement to RFLP markers for use in soybean molecular biology, genetics, and plant breeding research (Shoemaker et al. 2004). A large set of more than 600 SSR markers were first developed and mapped in an integrated molecular linkage map of soybean (Cregan et al. 1999a). The development of additional SSR markers was subsequently continued and up to date there are more than 1000 SSR loci (Song et al. 2004) were mapped to 20 homologous linkage groups (LGs) presumed to be associated with the 20 pairs of soybean chromosomes (USDA-ARS SoyBase website: http://soybase.agron.iastate.edu). With the ongoing efforts in soybean BAC end and whole genome shotgun sequencing, additional SSR markers could be easily identified and mapped.

2.2.5. Single nucleotide polymorphism

Recently, a novel molecular marker system was developed and successfully employed in human and mice genomic projects (Collins et al. 1998; Kwok et al. 1996; Lindblad-Toh et al. 2000). This marker system based upon changes in single

Transitional SNP

Wild type: 5′-TGTTATCCAT**C**ACCTTTTCCCTCA-3′
Mutant line: 5′-TGTTATCCAT**T**ACCTTTTCCCTCA-3′

Transversional SNP

Wild type: 5′-AGATTTAATT**A**TGGTAAAACTAGGA-3′
Mutant line: 5′-AGATTTAATT**C**TGGTAAAATTAGGA-3′

Indel

Wild type: 5′-GGGCATTCT**TATA**TTCATTCAAATT-3′
Mutant line: 5′-GGGCATTCT....TTCATTCAAATT-3′

Figure 1. SNPs and Insertion/Delection (Indels) are prevalent sequence variations between individuals of all animal and plant species. These variations are consequences of either transitional or transversional mutation events

DNA base between homologous DNA fragments or small insertions and deletions (indels), which was collectively called single nucleotide polymorphisms (SNPs) (Figure 1). The single nucleotide changes were due to point mutations, one of the many types of mutations that naturally occur in genomes with relatively low mutation rates (Kahl et al. 2005). With high level of polymorphism and even distribution across genomes, SNPs are valuable gene-based molecular markers suitable for high-throughput, automated genotyping. The great potential use of SNPs for genetic mapping of complex traits, pharmacogenetics, and medical diagnostics has received much attention in medical sciences (Gu et al. 1998; McCarthy and Hilfiker 2000).

Following remarkable progress made in the human and animal genome projects, SNPs markers have been widely investigated and developed in many important plant genomes, such as *Arabidopsis* (Kawabe et al. 2000; Kuittinen and Aguade 2000), barley (Kanazin et al. 2002; Rostoks et al. 2005), rice (Zhao et al. 2002; Monna et al. 2006), maize (Tenaillon et al. 2001; Vroh Bi et al. 2006), sunflower (Lai et al. 2005), and soybean (Zhu et al. 2003). In soybean, SNPs marker development was first reported by Scallon et al. (1987) when comparing 3543 bp of the *Gy4* glycinin locus plus flanking DNA in two genotypes. Since then, several additional SNPs were identified by independent research groups (Xue et al. 1992; Zhu et al. 1995). Recently, when comparing sequences of 143 DNA fragments derived from coding and noncoding regions, which was based upon sequence analysis of 25 diverse cultivated soybean genotypes, Zhu et al. (2003) showed that nucleotide diversity, indicated by θ value, was 0.00053 and 0.00111 in coding and noncoding DNA sequence, respectively. It indicated that SNP frequency in soybean genome was much lower than the similar estimates of *Arabidopsis* ($\theta = 0.0017$) (Olsen et al. 2002) and maize ($\theta = 0096$) (Tenaillon et al. 2001). It was speculated that low level of linkage disequilibrium coupled with limited haplotype diversity in this investigation suggested that cultivated soybean genome is a mosaic of a

limited number of haplotypes (Zhu et al. 2003). The authors also suggested that SNP discovery focused on noncoding perigenic regions, where greater sequence polymorphism is present, will permit successful SNP diversity in soybean. The results of a subsequent study of SNP discovery in soybean supported the above opinions (Van et al. 2005).

For decades, random sequencing of gene transcripts has been recognized as simple and efficient method of identification of many the expresses genes in an organism (Putnay et al. 1983). Hence, these sequences, known as Expressed Sequence Tags (ESTs), have become valuable and useful source of DNA sequences in which SNPs can be efficiently discovered. In soybean, the Soybean EST Project database contained more than 391,034 publicly available ESTs from 84 libraries (www.ncbi.nlm.nih.gov/) as of August 2006. Employing this DNA sequence resource, Van et al. (2005) identified 97 SNPs present in coding sequence and noncoding regions in 15 soybean genotypes from Korea and the United States and concluded that a diverse set of soybean genotypes can be successfully utilized for SNP discovery.

Many efforts have been made to identify and map SNPs associated with traits of interest in soybean. Using a recombinant inbred line (RIL) population of two Korean soybean genotypes, Kim et al. (2004) identified three SNPs closely linked to lipoxygenase-2 gene (*Lx2*) in soybean seed, which is responsible for generation of grassy-beany and bitter flavors. Jeong and Saghai Maroof (2004) used allelic-specific PCR approach and detected two SNP markers tightly linked to two mosaic virus resistance genes, *Rsv1* and *Rsv3*, in soybean. Using a sequence of the *GmNARK* gene (http://www.ncbi.nlm.nih.gov/), which is responsible for supernodulation mutation in soybean, Kim et al. (2005) identified two SNPs for this gene and successfully developed PCR-based markers specific for these SNPs. Lately, Wu et al. (2005) and Hyten and Cregan (2006) developed several SNPs for *rhg1* and *Rhg4* loci responsible for SCN resistance in soybean. In addition to the existing SSR/RFLP-based soybean linkage map, 1,183 gene-based SNPs were mapped to 20 soybean MLGs (Cregan et al. 2006). The availability of the new integrated SSR/RFLP/SNP genetic map would provide soybean researchers valuable sources of markers for a diversity of applications, such as QTL mapping, MAS, positional cloning and association analysis.

2.3. Soybean Genetic and Physical Map

2.3.1. Genetic map

As mentioned above**,** soybean is thought to be the product of a diploid ancestor ($n = 11$), which underwent aneuploid loss ($n = 10$), polyploidization ($n = 20$) and diploidization. Thus, it is expected that any given gene will be present four times in the genome, which is a big challenge for the genomic and biological research in dissecting genes or genomic regions conferring to important agronomic traits. A good soybean genetic map, therefore, would be an important tool for molecular genetics analysis. Through over fifteen years' efforts, the soybean composite genetic

map is well developed. The map encompasses 20 linkage groups and contained 1,849 markers, including 1,015 SSRs, 709 RFLPs, 73 RAPDs, 24 classical traits, six AFLPs, ten isozymes, and 12 others (Song et al. 2004). With the development of thousands of SNP markers, a high density of genetic map will be coming soon.

2.3.2. Physical map

A physical map for the soybean cultivar 'Forrest' was constructed from 78,001 BAC and BIBAC representing 9.6 haploid genomes (Wu et al. 2004a). This map is under further development in attempts to reduce the number of contigs, create the minimum tile path, and add additional DNA markers (e.g., through BAC-end sequencing) (Shultz et al, 2006a; 2006b). Soybean cultivar 'Williams 82' was selected by the soybean genomics community as the representative cultivar for soybean genomics research and over 80% of EST sequences have been generated from this cultivar. This genotype is also being used for whole genomic sequencing. Therefore, construction of a physical map for the soybean cultivar 'Williams 82' is underway. Genome-wide physical mapping strategy based on BAC restriction fingerprinting has revolutionized genomics of large, complex genomes. The global fingerprinting approach employs a restriction fragment profile of a clone as its fingerprint and assumes that clones sharing a large number of, but not all, fragments are overlapping clones. Comparison of the fingerprints in a large population of clones is used to identify overlapping clones and is the basis for the construction of contigs. To generate fingerprints of DNA fragments cloned in BACs, two methods have been applied in soybean physical mapping, one is polyacrylamide (or DNA sequencing) gel-based fingerprinting (PAGFP) method for the 'Forrest' physical map (Wu et al. 2004a), and another one is a capillary-sequencer-based fingerprinting method, known as high information content fingerprinting (HICF) (Luo et al. 2003), for the 'Williams 82' physical map (unpublished data). The HICF method generates a more accurate and detailed clone fingerprint, and therefore produces better assemblies with less redundancy needed in the clone coverage (Nelson et al. 2005). The fingerprint data can be directly uploaded into an adapted version of Fingerprint Contig (Soderlund et al. 1997; 2001) software to assemble contigs based on restriction digest fingerprints, and the contigs are ordered and assigned to chromosomes based on anchored markers.

Integrated genetic and physical maps are extremely valuable for map-based gene isolation, comparative genome analysis, as sources of sequence-ready clones for genome sequencing projects and as important references for assembling of whole genome shot-gun sequences for the complex genomes. Accordingly, the public efforts are to develop resources to access the complete soybean genome through producing a high-resolution genetic map densely populated with SSR and SNP markers, constructing physical map segments with deep-coverage libraries of BACs, and through molecular markers to integrate the genetic and physical maps. In order to reach the goals, two important tools have being developed in the National Center for Soybean Biotechnology (NCSB) at University of Missouri-Columbia. The first tool is the six-dimensional BAC DNA pools, which is used for anchoring genetic

markers onto the physical map. The 49,152 clones of a Williams 82 BAC library that covers about 6 genome equivalents were pooled in six different dimensions creating 208 pools of BAC DNA. A PCR approach is used to screen these pools to identify BAC clone(s) containing genetically-mapped SSR markers and genes of interest. This resource will be fundamental to integrate the genetic and physical maps in an efficient and cost-effective manner, provide an invaluable resource for mapping genes or EST unigenes, and afford targeting genomic regions underlying various soybean traits including resistance to soybean cyst nematode, tolerance to abiotic stresses, and seed compositions for positional cloning. The second tool is a high density genetic map using a population of more than 700 RILs derived from the cross of Forrest and Williams 82 for integration of genetic and physical map in a fine scale.

The integrated physical, genetic, and cytogenetic map of the soybean genome will provide a 'highway' for anchoring gene sequences on physical map, isolation of a large number of genes and for many genetic and biological studies. The physical map can be readily accessed and will provide virtually unlimited numbers of DNA markers from any chromosome region or closely linked gene of interest. Synteny between soybean and legume species could also be used for isolation of agronomically important genes in other species such as *Medicago* or *Lotus* (Cannon et al. 2003; Yan et al. 2004; Mudge et al. 2005), for which whole genome sequences are available in GenBank.

3. FUNCTIONAL GENOMIC TOOLS AND RESOURCES

Plant functional genomics has emerged as a new and rapidly evolving scientific discipline to study the functions of genes. With the rapid advancement in the complete sequencing of genomes of an increasing number of plants (*Arabidopsis*, rice, *Medicago*, etc.), research priority has shifted from the identification of genes to the elucidation of their functions. The ultimate goal of functional genomics is the improved understanding of cell organization at different levels, from individual genes to groups of bio-molecules and complete genomes. Breakthroughs in understanding the relationship between genotype and phenotype will happen only through the integrated functional genomics approach. Recently, legume genomics has been focused primarily on the development of resources and information of two species considered to be model legumes, *Medicago truncatula* and *Lotus japonicus*, and soybean, the legume of principal economic importance in the United States (VandenBosch and Stacey 2003). A Joint Genome Institute (JGI) and soybean research community collaborative effort to sequence the soybean genome is underway and this will facilitate progressive completion of gene function analysis in soybean.

Very promising genomic tools including extensive EST resources and the availability of different kinds of microarray platforms are instrumental to build momentoum toward gene function studies in soybean, and recently included reference species among legumes. In soybean, over 390,000 EST sequences are publicly available through various projects (http://www.plantgdb.org) and these

are expressed in different tissue and organ systems including developing seeds, seed coats, leaves, pods, roots, and various stages of plants regenerating from tissue culture. This resource includes the recently added drought stress specific and root tissue type specific ~15,000 EST sequences generated from the normalized, subtracted soybean root libraries from authors research group. Analysis of these ESTs contributed nearly 36,000 Unigene for soybean (www.tigr.org). Even though ESTs are useful for making a catalogue of expressed genes and the information, genome-scale collections of the full-length cDNAs of expressed genes are important for the analysis of structure and function of genes (Seki et al. 2004). So, in order to understand the gene regulatory network and also for various marker development projects it is necessary to build a strong soybean full length cDNA resource.

DNA microarray assays (Schena et al., 1995; Shalon et al. 1996) have provided an unprecedented opportunity for the generation of gene expression data on a whole-genome scale. The production of microarrays and the global transcript profiling of crop plants has revolutionized the study of gene expression and provides a unique snapshot of how these plants are responding to a particular stress/event. Both cDNA and oligoarrays are available for soybean gene expression profiling studies. The cDNA arrays available for soybean were constructed from low redundancy 'unigene' set of 27,513 clones that represent a variety of soybean cDNA libraries made from a wide array of source tissue and organ systems, developmental stages, and stress or pathogen-challenged plants (Vodkin et al. 2004). The construction of new sets of soybean long oligo arrays (70mer oligo sequence) is in pipe line from the above research group and these arrays contain a total of ~37,000 transcripts. At the same time GeneChip® soybean genome arrays are available to study gene expression of over 37,500 soybean transcripts. This short oligo array also contains transcripts for studying two pathogens important for soybean research. Specifically, the array includes probe sets to detect approximately 15,800 transcripts for the fungal pathogen of Phytophthora root rot (*Phytophthora sojae*) as well as 7,500 soybean cyst nematode (*Heterodera glycines*) transcripts (http://www.affymetrix.com/products/arrays/specific/soybean.affx).

The first generation of soybean cDNA microarrays containing approximately 4,100 Unigene ESTs derived from axenic roots, to evaluate their application and utility for functional genomics of organ differentiation in legumes (Maguire et al. 2002). Using a 9,280-cDNA clone array, Thibaud-Nissen et al. (2003) have identified 495 cDNA clones showing modulation of expression in response to 2,4-D treatment during the development of somatic embryos. Clustering the clones by similarity of expression profile over the course of this study allowed determining the timing of molecular events taking place during embryogenesis. In the same studies, they also found that oxidative stresses influence arrangement of new cells into organized structures which may depend on a genetically controlled balance between cell proliferation and cell death. Transcript profiling experiments conducted during *Pseudomonas syringae* pv. *glycinea* susceptible (S) and hypersensitive response-associated resistance (R) interactions in soybean identified differential expression pattern of plant-pathogen responsive genes and reported that the

R response generally showed a significantly greater degree of differential expression (Zou et al. 2005). Recently, Ainsworth et al. (2006) investigated the transcriptome response of soybean to elevated carbon dioxide [CO_2] in growing and fully expanded leaves using the cDNA microarrays. This study suggested that at the transcript level, elevated [CO_2] also stimulates the respiratory breakdown of carbohydrates, which likely provides increased fuel for leaf expansion and growth. Recently, the author's research group has identified several drought stress responsive and root and leaf tissue specific transcripts in soybean (Valliyodan et al. unpublished data) using GeneChip® soybean genome arrays from Affymetrix (www.affymetrix.com).

The use of microarray tool to analyze gene expression changes within the target plant is important for revealing the transcriptional regulatory networks which in turn contribute to understanding the responses of crop plants to various stresses and developmental changes, and crop improvement. Although the uses of gene expression profiling using microarrays are increasingly common, the integration of transcriptomics, proteomics and metabolomics represents the frontier of integrated functional genomics.

4. DISSECTING THE GENETICS OF THE TRAITS AND APPLICATIONS

4.1. Disease Resistance

Similar to other crop plants, several agronomically important traits in soybean, such as grain yield, disease resistance, and seed composition, are controlled by quantitative trait loci (QTL). These complex traits are environmentally sensitive; thus, it is often difficult to make progress in conventional breeding programs (Lee 1995). This is especially true for traits with low heritability. The development of DNA molecular marker technologies has facilitated the construction of genetic maps of many crop genomes, including soybean. A genetic analysis of the association of phenotypic trait and genotypic marker data from a population of lines segregating for genes conditioning quantitative traits provided the unique ability to dissect the complex trait into a set of discrete QTL (Tanksley et al. 1989), in which the contribution of an allele from each parent at a particular locus can be determined separately from the other loci affecting the trait. These important attributes provide an effective approach to pyramid desirable alleles for a complex trait into an adapted genetic background (Orf et al. 2004). Therefore, marker-based plant breeding coupled with traditional breeding techniques will be more effective and more efficient to introduce important agronomic traits into soybean.

4.1.1. Soybean cyst nematode

Soybean cyst nematode (SCN, *Heterodera glycines* Ichinohe) is among the most economically destructive pests of soybean in the United States and the world. The disease was first reported in Japan (1915), in Korea (1936), and in the US (1954) (Winstead et al. 1955). In the US, the most recent estimate shows that SCN causes

a 2 to 6% annual yield loss worth $1 billion each year (Wrather et al. 2003). Early inheritance studies indicated that SCN resistance is genetically controlled by three recessive genes designated *rhg1*, *rhg2*, and *rhg3* (Caldwell et al. 1960) and one dominant gene designated *Rhg4* (Matson and Williams 1965). An additional dominant gene was later identified in PI88788 and designated as *Rhg5* (Rao-Arelli 1994). But genetic analysis of different sources of resistance showed that inheritance of SCN resistance was oligogenic and complex (Anand and Rao-Arelli 1989). Multiple alleles at a single locus could be involved in SCN resistance (Hancock et al. 1987; Hartwig 1985). Since then, many efforts have been made to screen soybean germplasm for resistance to SCN. Among the resistant sources identified, soybean line 'Peking' and plant introduction 'PI88788' were quickly deployed to incorporate resistant genes into new varieties in soybean breeding programs (Arelli et al. 2000). However, it was reported that certain field populations of SCN could overcome resistant cultivars, revealing the complex nature of host-pathogen interaction in soybean and SCN, which required the establishment of race-specific interrelationship for each resistant genotype and SCN race.

Advances in molecular genetics and biotechnology make it possible to identify and map genomic regions conditioning SCN resistance. As one of early molecular genetics analysis of SCN resistance, Concibido et al. (1994) reported two RFLP markers, pA85 and pB32, were significantly associated with SCN disease response and tentatively mapped to LGs A and K. In addition, a previously identified RFLP maker, pK69 (Boutin et al. 1992), was also found to be associated with SCN resistance and mapped to LG G (Concibido et al.1994). Since then, several additional putative SCN resistant QTL were identified from a number of resistance sources, such as plant introductions, exotic germplasm, wild annual species (*Glycine soja*). Recently, when reviewing early QTL analysis of SCN resistance, Concibido et al. (2004) summarized a total of 16 putative QTL responsible for SCN resistance mapped on 17 different LGs of the soybean integrated molecular linkage groups (Cregan et al. 1999a). Of these, the QTL on LG G has been well characterized and proven to be the most important QTL because it contains the gene *rhg1*, which is capable of resistance to most of existing SCN races, 1, 2, 3, 5, 6, and 14 (Concibido et al. 1997; Yue et al. 2001a; 2001b). In addition to RFLP, AFLP, and RAPD markers reported to be associated with different genomic loci on LG G (Chang et al. 1997; Concibido et al. 1994; Kabelka et al. 2005; Meksem et al. 2001), Cregan et al. (1999b) mapped two SSR markers, BARC-Satt309 and BARC-Sat_168, at a distance of 0.4 cM to the gene *rgh1* and concluded that these two markers were highly effective in differentiating resistance and susceptible genotypes in SCN soybean breeding process. When mapping QTL for SCN resistance from PI90763 and PI404198A, Guo et al. (2006) also found the same genetic linkage of these SSR markers on LG G. It indicated that marker-assisted selection for SCN resistance can be highly efficient by deploying a small set of the SSR markers flanking within interval of the *rhg1* locus.

The second important QTL on LG A2 was located at the dominant gene *Rgh4*, which was reported to play a distinct role in resistance to SCN race 3

(Concibido et al. 1994; Heer et al. 1998; Webb et al. 1995). Similar to genetic linkage of the gene *rgh1*, various DNA molecular markers were shown to be associated with this LGA2-QTL containing the *Rhg4* gene (Concibido et al. 2004). Especially, Cregan et al. (1999b), Guo et al. (2006) mapped the SSR marker Satt632 close to this genomic region. In addition, the *i* locus, a morphological marker conditioning seed coat color, was also proven to be linked to *Rhg4* at a genetic distance of less than 1 cM (Webb et al. 1995). Using AFLP technology and bulk segregant analysis (BSA) to analyze resistant cultivar 'Forrest', Meksem et al. (2001) constructed a high-density genetic map of AFLP markers for the intervals carrying the genes *rhg1* and *Rhg4* and concluded that these two genes together explained over 98% of the resistance to SCN race 3. The findings were similar to the conclusion of a previously study that both these two genes are required for expression of complete resistance to SCN race 3 (Webb et al. 1995).

In attempts to characterize functional genomics of these two genes, Hauge et al. (2001) have reportedly cloned candidate genes for *rhg1* and *Rhg4* using positional cloning techniques. Both candidate genes are *Xa21*-like receptor kinases, which belong to R-genes conferring resistance to bacterial blight in rice (Zaitsev et al. 2001). Structures of these genes have an extracellular leucine-rich repeat (LRR) domain, showing similarity at the amino acid level to the *Cf-2* disease resistance gene, which confers resistance to leaf mold in tomato (Dixon et al. 1996). The results were confirmed in another patent application (Lightfoot and Meksem 2002). Even so, neither research group has reported their complementation studies to confirm that they have correct candidate genes. It is obvious that cloning and characterization of the candidate genes of *rhg1* and *Rhg4* will provide invaluable insight into the mechanisms of genetic resistance to SCN and might provide the knowledge needed to engineer novel resistance genes (Concibido et al. 2004).

With the availability of molecular markers closely linked to SCN resistance genes, such as Satt309, Satt163, Sat_168 (LG G), or Satt632, Sat_157 (LG A2), soybean breeders are capable of selecting resistant lines by genotyping breeding populations in a few days instead of growing progenies for greenhouse bioassay taking a month (Concibido et al. 2004). Each year around 20,000 soybean lines were genotyped for MAS in the SCN resistance breeding program of University of Missouri-Columbia (UMC) (Sleper, per. com.). Obviously, MAS enables breeders increasing the efficiency and speed of development of SCN resistant cultivars compared to conventional SCN breeding. However, in some occasions MAS for the *rhg1* gene has some difficulty due to lack of polymorphisms between parental lines for markers linked to this locus. Lately, the development of SNP genetic markers in soybean (Zhu et al. 2003) has provided soybean breeders an innovative approach to overcome the difficulty encountered. Wu et al. (2005) reported that several SNPs were identified for the *rhg1* and *Rhg4* loci in 30 parental genotypes commonly employed in the SCN resistance breeding program of UMC. With a sequenced region containing the *rhg1* gene, Hyten and Cregan (2006) demonstrated association analysis and discovered 67 SNPs throughout the region being significantly associated with SCN resistance. These findings indicate that SNP for SCN resistance would be molecular markers of choice in soybean

breeding programs and become more efficient as high-throughput SNP genotyping platforms such as Luminex flow cytometry (http://www.luminexcorp.com) and Illumina Beadstation (http://www.illumina.com) are employed.

4.1.2. Phytophthora

Phytophthora root and stem rot caused by fungal pathogen *Phytophthora sojae* was reported to occur in the US and other soybean production regions in the world, such as Brazil, Argentina, Australia, China, and South Korean (Grau et al. 2004). Studying the host resistance to this pathogen indicated that there are three types of resistance based upon inoculation experiments performed in host plant positions. For instance, the hypocotyl inoculation test detects single dominant genes, known as *Rps*, while the root inoculation test detects partial resistance, which is expressed as fewer rotted roots and slower disease progress (Tooley and Grau 1984). To date, eight dominant genes, designated as *Rps1* to *Rps8*, have been identified to convey resistance to 55 physiologic races of *P. sojae* (Burnham et al. 2003). Among these, *Rps1* and *Rps3* were reported to have multiple alleles (Grau et al. 2004). This indicated race-specificity of host resistance and genetic complexity of *P. sojae* population in soybean.

The availability of genomic genotyping technologies has facilitated the identification and mapping of DNA molecular markers associated with resistance to Phytophthora root and stem rot on the soybean genetic linkage map (Cregan et al. 1999a). Screening mapped RFLP markers in 'Williams' near-isogenic lines, Diers et al.(1992b) mapped the loci *Rps1*, *Rps2*, and *Rps3* on linkage groups (LG) K, L, and E, respectively. However, Demirbas et al. (2001) utilized SSR markers for genetic linkage analysis and mapped the loci *Rps1*, *Rps2*, *Rps3*, and *Rps4* on LGs N, J, F and G, respectively. Both research groups reported genomic locations of *Rps5* and *Rps6* were unclear. Lohnes and Schmitthenner (1997) used another set of RFLP markers to map locus *Rps7* in a 'Clark x Harosoy' isoline mapping population and reported this locus was closely linked to the RFLP marker R022-1 on LG N. The genomic location of the loci *Rps1* and *Rps7* on LG N was subsequently confirmed by Demirbas et al. (2001) and Weng et al. (2001), who utilized SSR markers and compared their linkage group to a consensus linkage group of the soybean (Cregan et al. 1999a). Burnham et al. (2003) reported a novel Phytophthora resistant locus, *Rps8*, in PI399073 from South Korea and mapped the locus on LG A2. Subsequently, however, the locus was re-located on LG F based on linkage analysis with SSR and RFLP markers (Gordon et al. 2006).

Among single-dominant *Rps* genes, *Rps1* locus was reported to carry five alleles (*Rps1-a*, *Rps1-b*, *Rps1-c*, *Rps1-d*, *and Rps1-k*) and confer resistance against a large number of *P. sojae* races (Schmitthenner et al. 1994). Thus, many efforts have been made to investigate molecular basis of this locus. By the means of primary trisomic analysis of cytogenetics, Gardner et al. (2001) was able to locate the locus *Rps1-k* on chromosome 3, corresponding to LG N of the integrated genetic linkage map of soybean. Bhattacharyya et al. (1997) and Kasuga et al. (1997) reported the high-resolution genetic and physical maps of the *Rps1-k* region and mapped

three AFLP markers tightly linked to the locus *Rps1-k* at a genetic distance less than 0.5 cM. The authors concluded that the *Rps1-k* was located at the end of an introgressed region. From a disease resistance gene-like sequence, Bhattacharyya et al. (2005) successfully cloned a family of CC-NBS-LRR-type sequences from the genomic region containing the locus *Rps1-k*. Among 12 candidate signaling proteins interacting with the resistance gene *Rps1-k*, Bhattacharyya and Gao (2006) reported that *Rps1-k-2* interactor regulates GmMcII, a gene responsible for hypersensitive cell death, which is associated with the expression of Phytophthora resistance.

Unlike the extensive studies conducted for single-dominant *Rps* genes, genetic information of partial resistance to *P. sojae* is limited. Recently, Burnham et al. (2003) developed three RIL mapping populations derived from the crosses of three susceptible parents and a partial resistant cultivar Conrad. The authors reported two QTL associated with *P. sojae* resistance were mapped to LG F and D1b+W of the soybean linkage groups. However, the authors also indicated that genomic locations of the QTL were different from the regions where the known *Rps* genes were mapped. This will pose a challenge to soybean breeders.

Employing single-dominant *Rps* genes for development of resistant cultivars has proven to be an economical and efficient measure to control this pathogen. However, as in many other host-pathogen systems, the fungal pathogen *P. sojae* has been reported to quickly adapt to specific *Rps* genes and as the result the host plant becomes vulnerable to a new physiologic race of the pathogen population. This is a great challenge to soybean breeders who have to continue accessing soybean germplasm to mine novel *Rps* genes for breeding programs. Besides single dominant gene-based breeding to control the disease, partial resistance has been shown to be effective against all races of *P. sojae* (Tooley and Grau 1986; Schmitthenner 1985). Anderson and Buzzell (1992) and Walker and Schmitthenner (1984) investigated resistance to *P. sojae* in several soybean lines and concluded that partial resistance is highly heritable and a quantitative trait. The authors suggested that selection for partial resistance combined with single-dominant *Rps* genes would provide long-term disease management of *P. sojae* in soybean.

Linkage analysis of DNA molecular markers and various *Rps* genes conditioning resistance to *P. sojae* in soybean can be useful to soybean breeders and pathologists to implement cost-efficient and successful MAS for Phytophthora resistance. Moreover, in addition to the identifications of molecular "tags" closely linked to the genes, markers flanking the loci of interest would provide an efficient tool for map-based cloning of the resistance genes, which in turn facilitates the development of SNP markers, a new generation of DNA molecular marker for soybean.

4.2. Abiotic Stress

Among the different abiotic stresses, drought, salinity, waterlogging/flooding and temperature stress (high and low) are the main factors reducing crop production worldwide and accounts for more than 50% yield looses (Boyer 1982; Bray et al.

2000). When plants are subjected to abiotic stresses, they activate different mechanisms to survive and sustain growth and yield. Thus, it is important to know what mechanisms exist to help plants surviving under severe stress. Plant traits linked to stress tolerance are multigenic, i.e. these traits are controlled by QTL. Here, progress on molecular mapping and QTLs analysis only for drought, salinity and flooding or waterlogging tolerance in soybean and their application in MAS for soybean improvement will only be presented.

Drought is a major limitation for soybean production. Crops often experience a water-limiting environment, thus the genetic improvement of a crop for drought resistance is a sustainable and economically feasible solution to reduce the problem of drought (Blum 1988). Drought tolerance is a complex phenomenon; dissection into individual component makes it easier to understand its genetic bases and also to apply this knowledge in MAS programs. Different authors described a number of drought related physiological traits in soybean from different perspectives (Raper and Kramer 1987; Nilsen and Orcutt 1996; Orcutt and Nilsen 2000; Purcell and Specht 2004). Purcell and Specht (2004) viewed drought from agro-physiological (i.e., plant biomass and grain yield), rather than eco-physiological (i.e., plant survival or competitiveness). Their consideration was based on soybean yield, thus a stress-related trait to be manipulated to lessen the reduction in soybean yield caused by the stress. They summarized a list of traits or physiological mechanisms impacting soybean yield via transpiration (T), water use efficiency (WUE) or harvest index (HI).

Molecular markers have been used to map the genomic location of genes/QTL for different traits related to drought, salinity and submergence tolerance in soybean (Table 1). As of August 2006, a total of 480 soybean genes and 1174 QTLs have been reported in the USDA-ARS soybean database (http://soybase.org). So far, very little progress has been made in the areas of abiotic stress tolerance but significant progress has been made for mapping many agronomic, pest resistant and seed composition related traits in soybean. Mian et al. (1996; 1998), using two mapping populations, cv.Young x PI416937 and S100 x Tokyo, identified a total of 5 QTLs for WUE, but only two QTLs linked to RFLP markers cr392-1n (LG J) and A489H (LG L) explained 13 and 14% of phenotypic variation, respectively. These two QTLs were found independently in two populations. Marker loci A489H (WUE2-1) also harbor QTL for leaf traits such as length, weight, width, shape, and area, plant yield and height, as well as seed oil content (http://soybase.org). Although yet to be confirmed in any other populations, WUE is no doubt an important locus to be used in MAS. An RFLP marker, A063E, was found common in both populations but phenotypic effect was less than 10%. Bhatnagar et al. (2005) identified a major QTL for slow wilting trait related to increased drought tolerance in soybean. The major QTL linked to SSR marker Sat_044 in linkage group K explained 17% of phenotypic variation. Specht et al. (2001) used 236 RILs developed from a cross between Minsoy x Noir 1 to determine the genetic basis of beta and carbon isotope discrimination (CID). They reported a QTL for CID on LG C2 with phenotypic contribution of <10% and with no effect on beta.

Table 1. Quantitative trait loci (QTL) identified for soybean drought, salinity and submergence/ waterlogging tolerance

Reference	Mapping population, Number of lines and Population type	Trait studied number of QTL (in parenthesis)	Linked marker (s), linkage group (LG) and phenotypic variation (R^2%)
Drought			
Bhatnagar et al. (2005)	Jackson x KS4895 81 RILs	leaf wilting (1)	SSR marker; Sat_044, LG K, R^2=17
Specht et al. (2001)	Minsoy x Noir 1 256 RILs	yield (1)	SSR marker Satt205-Satt489, LG C2, $R^2 \sim 7$
Mian et al. (1998)	S100 x Tokyo 116 F_2	water use efficiency (1) (WUE)	RFLP marker A489H, LG L, R^2 =14
Mian et al. (1996)	Young x PI416937 120 F_4	water use efficiency (4) (WUE)	4 RFLP markers, B031-1n, A089-1, cr497-1, K375-1n; $R^2 \sim$ 8–13
Salt tolerance			
Lee et al. (2004)	S100 x Tokyo, 100 F_2	salt tolerance (1)	SSR marker, Sat_091 LG N, $R^2 \sim$ 41-79
Flooding/waterlogging			
Reyna et al. (2003)	A5403 x Archer, RILs 9461 x Archer, NILs	flooding/ waterlogging QTL effect evaluation	No effect of marker Sat_064
VanToai et al. (2001)	Archer x Minsoy 122 RILs	waterlogging (1)	In both the populations, marker Sat_064 linked for improved plant growth
	Archer x Noir 1 86 RILs	waterlogging (1)	$R^2 \sim$ 11-18, and grain yield $R^2 \sim$ 47-80

Cultivated soybean (*Glycine max*) is considered a salt sensitive crop whereas wild soybean (*Glycine soja*) is found to be tolerant (Singh and Hymowitz 1999; Luo et al. 2005; Kao et al. 2006); thus wild soybean (*G. soja*) can be used as resistance source for varietal improvement in soybean (*G. max*). So far, a major QTL for salt tolerance has been mapped in soybean (Lee et al. 2004). More than 100 F_2 populations were developed from a cross between S100 (salt tolerant) x Tokyo (salt sensitive). This QTL was linked to the SSR marker Sat_091 on LG N accounting for 41, 60, and 79% of the genetic variation for salt tolerance in the field, greenhouse, and combined environments, respectively. This data needs to be confirmed in another population before using in MAS program.

Waterlogging is also a major environmental stress reducing soybean growth and yield. Different authors have reported that yield can be reduced 17–43% if soybean is subjected to waterlogging at the vegetative stage, but greater yield reduction

(50–55%) occurs under waterlogging stress at the reproductive stage (Oosterhuis et al. 1990; Scott et al. 1989). Tolerance to waterlogging could increase yield and tolerance in soybean has been identified (VanToai et al. 1994). One major QTL for waterlogging has been reported and confirmed in two mapping population of soybean (VanToai et al. 2001). The QTL linked to SSR marker, Sat_064 on LG G explained phenotypic variance for plant growth 11–18% and yield 47–180%. Reyna et al. (2003) used Sat_064 genotypic data to create several NILs and these NILs were evaluated in southern background and found no effect for Sat_064. This variation may be due in part to the environment differences and the fact Sat_064 was originally identified in a northern genetic background. Several mapping populations have been developed at UMC for QTL mapping with the objectives to discover new QTLs and/or genes that confer tolerance to waterlogging under different genetic backgrounds and apply them for marker-assisted crop breeding.

Marker-assisted selection (MAS), has been successfully applied into soybean breeding program for SCN and other pest resistance traits, oil content and other traits. Monsanto and other private industries have incorporated MAS in their breeding programs and are making millions of selections annually, largely for insect and disease resistance genes using MAS. At the same time, public sectors are also using MAS for breeding programs on a limited scale. High density genetic map (saturated with molecular markers), tightly linked QTLs/genes with markers and confirmed QTLs/genes across the environments and in different genetic backgrounds are the most important criteria for effective MAS. MAS for abiotic stress related traits like drought, salinity and submergence tolerance have been initiated but not to extent of biotic stresses. Both private and public sectors are putting efforts on molecular breeding for abiotic stress tolerance.

4.3. Seed Composition

As earlier mentioned, soybean are grown as a source of protein and oil for agricultural, consumer, and industrial uses. In recent years, the value of soybean seed composition has extensively received attention for industrial applications and medical effects. These include increasing oil content for fitting the demand of biodiesel production, altering fatty acid or amino acid profiles, and developing healthy soybean seeds with high oleic acid, high isoflavone, low linolenic acid, or low phytate. However, the value of soybean to the producer is traditionally based on yield, and this trait has been the primary focus of most soybean breeding programs (Nichols et al. 2006). Recently traits that improve the value and functionality of soybeans such as modification of protein and oil to give greater utility in food, health, and industrial uses have been emphasized by soybean breeders and industries. Soybean breeding research has been directed toward modifying the fatty acid profile in the oil to expand uses in food and industrial applications. Modifying soybean oil to lower saturates to less than 7%, increasing oleic acid content to 50 to 60%, and lowering linolenic acid content to <3% would greatly improve soybean quality for greater functionality in food and industrial applications. Other areas of

emphasis include improving protein by modifying the amino acid profile, improving digestible phosphorus, and reducing anti-nutritional factors such as trypsin inhibitors and allergenic factors (Wilson 2004).

There are many traits whose genetic potential is masked by epistatic interactions between genes or by genes in the repulsion linkage phase. Molecular markers can assist selection for favorable alleles, dissect the QTL regions conferring favorable seed traits and breakdown the unfavorable linkage relationship at the genetic level (Lee 1995). DNA marker analysis can measure the effect of an individual locus at the allelic level. For example, the contribution of alleles from each parent at a particular locus can be determined separately from other loci affecting the trait. This provides an effective method to pyramid desired genes for complex traits or traits with low heritability into adapted varieties (Orf et al. 2004). Therefore, high-throughput molecular breeding techniques, coupled with traditional breeding approaches, will be more effective and efficient for introducing important quantitative traits into soybean.

So far, a large number of QTL related to soybean seed composition have been detected by QTL analysis with molecular markers. All major QTL with more than 10% of phenotypic contribution were summarized in Table 2. The QTL identified could be useful for developing soybean varieties with desirable component in the seed through MAS. QTLs conditioning protein content in soybean were widely investigated on 17 soybean populations and found to be located on the soybean genome except for LG B1, D1b, D2, J, and O (Table 2). A total of 61 QTLs for protein content were reported and 16 putative QTLs (~25%) accounting for more than 10% phenotypic variation were identified. Among them, ten QTLs were detected in more than two soybean populations. The trait for protein content is highly heritable, as shown in the locus on LG I (Satt127) with a maximum phenotypic contribution of 65% (Table 2). Fine-mapping studies have been used to clone the gene close to the Satt127 marker on LG I and resulted in QTL localization within 1.1 cM using BAC libraries (Joseph et al. 2004; Nichols et al. 2006). MAS of lines with or without positive alleles at the QTL on LG I was applied to compare the allelic effect and confirm the QTL resulting in a 56 g kg^{-1} increase in protein content in lines carrying homozygous alleles from the high protein parent (Yates et al. 2004).

In addition to high protein content, soybean is also an important oil source for food and biodiesel. QTL for oil content are distributed throughout the soybean genome except for LG D1a, D1b, F, J, N, and O (Table 2). Thus, three LGs (D1b, J, O) do not carry genes for both protein and oil content. Of 19 QTLs identified, 11 (69%) and 19 (58%) QTLs were found to be associated with protein and oil content, respectively, commonly control both traits. Because strong negative correlations exist between protein and oil content, it needs to be determined genetically whether the negative relationship is due to pleiotropism or repulsion linkages between protein and oil alleles. Of 53 oil QTLs reported in SoyBase (http://soybase.agron.iastate.edu), 19 putative QTLs explain more than 10% of the phenotypic variation (Table 2).

Table 2. Major QTL ($R^2 \geqq 10\%$) related to soybean seed composition

Trait	Population	LG	Flanking loci (cM) (a)	R^2 (%)	LOD	Reference
Protein content	A81-356022 x PI 468916 60 $F_{2:3}$	I	K011	42	–	Diers et al. 1992a
		E	SAC7_1	24	–	
		B2	A242_1	19	–	
		L	A023_1	16	–	
		G	A245_2	12	–	
	PI 437088A x Asgrow A3733 76 RILs	I	OPAJ13a-OPAW13a	36	20	Chung et al. 2003
	F_2 eight populations (b)	A2	A505_1	11	–	Brummer et al. 1997
		C1	A063_1	17	–	
		D1a	A398_1	28	–	
		E	B174_1	11	–	
		G	A890_1	16	–	
		I	A144_1	28	–	
	Young x PI 416937 120 F_4 (b)	C1	Gc97_1	13	–	Lee et al. 1996
		N	A071_2	11	–	
		B2	B142_1	10	–	
	PI 97100 x Coker 237 111 F_2 (b)	H	A566_2	14	–	Lee et al. 1996
		K	A065_3	11	–	
	Peking x Essex 200 $F_{2:3}$	H	B072	32	–	Qui et al. 1999
		F	B148	17	–	
	Noir 1 x Archer 240 RILs	C1	Satt578	12	–	Orf et al. 1999
		L	Satt166	11	–	
	Minsoy x Archer 233 RILs	A1	T155_1	15	–	Orf et al. 1999
		C1	SoyGPATR	12	–	
	A81-356022 x soja PI 468916 BC_3	I	Satt127	65	–	Sebolt et al. 2000
	Minsoy x Noir1236 RILs	M	Satt567	27	12.8	Specht et al. 2001

(continued)

Table 2. (continued)

Trait	Population	LG	Flanking loci (cM) (a)	R^2 (%)	LOD	Reference
	Essex x Williams 131 RILs	C2	Satt277-Satt202	28	9.8	Hyten et al. 2004
		F	Satt335-Satt144	18	4.4	
		K	Satt539-Satt102	24	4.3	
		M	Satt540-Satt463	13	3	
	N87-984-16 x TN93-99101 RILs	G	Satt570	20	3.5	Panthee et al. 2005
	Kefeng No. 1 x Nannong 1138-2 184 RILs	B2	A953_1H-Satt560	12	3.5	Zhang et al. 2004
Cysteine	N87-984-16 × TN93-99101 RILs	F	Satt252	11	2.2	Panthee et al. 2006
		G	Satt235	13	2.8	
Methionine	N87-984-16 × TN93-99101 RILs	F	Satt252	15	2.8	Panthee et al. 2006
		G	Satt564	19	2.6	
		M	Satt590	23	2.4	
Oil content	A81-356022 x PI 468916 60 $F_{2:3}$	I	A407_1	28	–	Diers et al. 1992a
		E	SAC7_1	43	–	
		B2	A242_1	39	–	
		L	A023_1	32	–	
	PI 27890 x PI 290136 69 $F_{2:5}$	A2	T153_1-A111_1 (18)	36	5.5	Mansur et al. 1993
		K	BC1-A315_1 (26)	24	2.9	
	F_2 eight populations (b)	A1	A104_1	19	–	Brummer et al. 1997
		B1	A109_1	31	–	
		G	A584_1	19	–	
		H	A069_1	18	–	
		K	K387_1	16	–	
	Young x PI 416937 120 F_4 (b)	D2	Cr142_1	13	–	Lee et al. 1996
	PI 97100 x Coker 237 111 F_2 (b)	C1	A063_1	13	–	Lee et al. 1996
		G	L154_1	17	–	
	Minsoy x Noir 1 240 RILs	A1	T155_1	13	3.4	Orf et al. 1999
	Archer x Minsoy 233 RILs	A1	Satt174	10	4	
		C1	SOYGPATR	11	3.3	

	Noir 1 x Archer 240 RILs	C2	Satt432	11	3.3	
		L	A489_1	19	6.1	
	Peking x Essex $F_{2:3}$	H	B072	21	–	Qui et al. 1999
	PI 437088A x Asgrow A3733 76 RILs	I	OPAJ13a-OPAW13a	26	14	Chung et al. 2003
	Essex x Williams 131 RILs	C2	Satt277-Satt460	32	12	Hyten et al. 2004
		L	Satt166-Dt1	10	3.3	
		M	Satt540-Satt463	12	3.6	
	N87-984-16 x TN93-99 101 RILs	D1b	Satt274	12	3	Panthee et al. 2005
		O	Satt420	15	3.5	
		O	Satt479	12	3.1	
Linolenic acid	A81-356022 x PI 468916 60 $F_{2:3}$	E	SAC7_1	31	–	Diers and Shoemaker 1992
		L	A023	26	–	
		K	A065_3	20	–	
Linoleic acid	A81-356022 x PI 468916 60 $F_{2:3}$	A1	A082_1	38	–	Diers and Shoemaker 1992
		E	A242_2	21	–	
		B1	A118_1	20	–	
Palmitic acid	A81-356022 x PI 468916 60 $F_{2:3}$	B2	A343_1	24	–	Diers and Shoemaker 1992
		J	K375_1	18	–	
	Cook x N87-2122-4	A1	Satt684	33	5–7	Li et al. 2002
Oleic acid	A81-356022 x PI 468916 60 $F_{2:3}$	A1	A170_1	23	–	Diers and Shoemaker 1992
		E	Pb	21	–	
		B2	A619_2	19	–	
Oligosaccharide	Keunolkong x Iksan 10 115RILs	L	Satt278	14	–	Kim et al. 2005
Cell wall polysaccharides	Minsoy x Archer 108 RILs	A1	Satt174–Satt211	17	4	Stombaugh et al. 2004

(continued)

Table 2. (continued)

Trait	Population	LG	Flanking loci (cM) (a)	R^2 (%)	LOD	Reference
Fucose	Minsoy x Archer 108 RILs	A2	Satt119–Satt233	38	11	Stombaugh et al. 2004
Galactose	Minsoy x Archer 108 RILs	A1	Satt211–T155_1	27	7	Stombaugh et al. 2004
Glucose	Minsoy x Archer 108 RILs	O	Satt153–Scaa001	13	3	Stombaugh et al. 2004
Pectin	Minsoy x Archer 108 RILs	A1	Satt174–Satt211	22	6	Stombaugh et al. 2004
Sucrose content	V71-370 x PI 407162149 $F_{2:3}$	I	A144	12	–	Maughan et al. 2000
Phytate	Boggs-RR x CX1834-1-2 166 F2	L	Satt561	11	–	Walker et al. 2006
		N	Satt237	41	–	
Total isoflavone	AC756 x RCAT Angora 207 RILs	A1	Satt200	12	2	Primomo et al. 2005
		M	Satt201	26	5.5	
Daidzein	AC756 x RCAT Angora 207 RILs	A1	Satt200	15	2.1	Primomo et al. 2005
		M	Satt201	18	3.4	
	Essex x Forrest 100 RILs	A1	Satt276	10	2.7	Kassem et al. 2004
		N	Satt080	10	3.2	
Genistein	AC756 x RCAT Angora 207 RILs	M	Satt201	31	7.6	Primomo et al. 2005
	Essex x Forrest 100 RILs	B2	Satt063	38	2.9	Kassem et al. 2004
Glycitein	AC756 x RCAT Angora 207 RILs	F	Satt516	39	2.9	Primomo et al. 2005
	Essex x Forrest 100 RILs	B1	Satt251	50	11	Kassem et al. 2004
		N	Satt237	11	2.3	

(a) : Only independent QTL contributing >10% to the particular trait variation.
(b) : single factor analysis was employed.

There were reports on seven, six, five, and six QTLs for linolenic, linoleic, palmitic, and oleic acid, respectively, in soybean fatty acid (FA) composition in SoyBase (http://soybase.agron.iastate.edu), including three QTLs for linolenic acid on LG E, K, and L; three QTLs for linoleic acid on LG A1, B1, and E; three QTLs for palmitic acid on LG A1, B2, and J; and three QTLs for oleic acid on LG A1, B2, and E. Only one QTL on LG E conditions the content of three FAs such as linolenic, linoleic, and oleic acids. Recently, six QTLs associated with increased oleic acid were mapped on LG A1, D2, G, and L with SSR markers in a population of G99-G725 (low) x N00-3350 (high), and four of them were confirmed in a different population of G99-G3438 x N00-3350 (Monteros et al. 2004).

Soybean also has desirable carbohydrates (i.e., glucose, fructose, and sucrose) and undesirable carbohydrates (i.e., raffinose, stachyose). Of 17 markers significantly associated with seed sucrose content on seven LGs, only QTL on LG I was found to have a phenotypic contribution greater than 10% (Table 2). Based on a comparison of the reported QTLs for seed composition traits, it is likely to be a gene cluster or a major QTL with pleiotropic effect (Maughan et al. 2000). Among the seven genomic regions underlying QTLs for sucrose content, four regions (LG E, F, I, L) also affect protein content, and three regions (LG A2, I, L) are associated with oil content as well.

There are many bioactive soybean components that may contribute to the hypothesized health benefits of soybean but most attention has recently focused on the isoflavones because isoflavones in soybean seeds are believed to have preventive effects for several major hormone-dependent diseases in humans. Numerous publications and research studies on isoflavones have prompted a nationwide increase in the consumption of soy-based foods and supplements in the United States. QTL analysis identified several genomic regions (LGs A1, B1, B2, F, M and N) associated with isoflavone contents (Primomo et al. 2005; Kassem et al. 2004).

5. CONCLUSIONS AND PERSPECTIVES

As a valuable source of seed protein and vegetable oil for human and animal nutrient, soybean is the most economically important legume crop in many countries in the world. In addition, the values of soybean seed composition have received a sincere attention by soybean researchers to investigate for industrial applications and medical uses. Steadily increased world production and meal consumption of soybean for last ten years indicated that soybean crop continues to be an important leguminous crop plants significantly contributing to various aspects of agricultural production, animal industry, and in particular human heath.

For decades, many efforts have been made to develop new soybean variety with improved nutritional and agronomic characteristics through conventional breeding approaches; however, the recent advances of DNA molecular marker technologies, such as SSR and RFLP, led to successful construction of the integrated genetic linkage map of soybean, in which genetic linkage of molecular markers were established in soybean genome. Newly gene-based SNP markers were integrated to

the pre-existing DNA molecular linkage map of the soybean would provide soybean researchers efficient genomic tools for identification and characterization valuable genes/QTL conditioning important agronomic traits and facilitating MAS programs.

Genome sequencing, in combination with various forward and reverse genetic approaches and "omics technologies", has made possible the identification of more than 30,000 genes in soybean, the understanding of the molecular basis of soybean growth and seed development, which is beginning to revolutionize the way in which plant breeders have utilized for cultivar improvement. Functional genomics has become an important discipline to identify gene, gene function, and to understand the relationship between phenotype and genotype via integrated functional genomic approach. Important functional genomic tools are ESTs, microarray, proteomics, metabolomics, translational genomics and bioinformatics. The availability of EST sequences generated from cultivar 'Williams 82', a representative soybean genotype, facilitated entire genome sequencing and construction of a physical map, which is based upon a novel strategy of 6-D BAC DNA pools. In combination of a genetic map with high density of SSR and SNP markers, a high resolution physical map will be invaluable resources for map-based gene isolation, targeting genomic regions underlying various traits of interest such as disease resistance, abiotic stress, and seed composition.

Recently, a collaborative effort led by Department of Energy and Joint Genome Institute (DOE-JGI) has been initiated aiming to sequence the entire soybean genome in hopes that the sequence information can be available by the end of 2008. With the availability of the soybean sequence and integrated genetic and high quality physical maps, soybean researchers will be able to efficiently implement linkage disequilibrium and association analysis. As a result, QTL mapping, positional cloning, as well as MAS will be accordingly accelerated.

ACKNOWLEDGEMENT

The authors would like to thank to Mr. Sean Blake for critical reading of the manuscript. Funding support to our research program from the Missouri soybean Merchandising Council, V.S. Department of Agricultrue, National Science Foundation, and the United Soybean Board is gratefully acknowledged.

REFERENCES

Akkaya MS, Bhagwat AA, Cregan PB (1992) Length polymorphism of s sequence repeat DNA in soybean. Genetics 132:1131–1139

Anand SC, Rao-Arelli AP (1989) Genetic analyses of soybean genotypes resistance to soybean cyst nematode race 5. Crop Sci 29:1181–1184

Anderson TR, Buzzell RI (1992) Inheritance and linkage of the *Rps*7 gene for resistance to Phytophthora rot of soybean. Plant Dis 76:958–959

Apuya NR, Frazier BL, Keim P, Roth EJ, Lark KG (1988) Restriction fragment length polymorphisms as genetic markers in soybean, *Glycine max* (L.) Merrill. Theor Appl Genet 75:889–901

Arelli PR, Sleper DA, Yue P, Wilcox JA (2000) Soybean reaction to races 1 and 2 of *Heterodera glycines*. Crop Sci 40:824–826

Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. Plant Mol Biol Rep 9:208–219

Bhatnagar S, King CA, Purcell L, Ray JD (2005) Identification and mapping of quantitative trait loci associated with crop responses to water-deficit stress in soybean [*Glycine max* (L.) Merr.]. The ASA-CSSA-SSSA International Annual Meeting (Abstract), November 6–10, p 9

Bhattacharyya M, Gao H (2006) Cloning and characterization of *Rps1-k* interactors. (Abstract) The 11th Biennial Conference on Molecular and Cellular Biology of the Soybean, August 5–8, Lincoln, NE

Bhattacharyya MK, Gonzales RA, Kraft M, Buzzell RI (1997) A copia-like retrotransposon Tgm *r* closely linked to the *Rps1-k* allele that confers race-specific resistance of soybean to *Phytophthora sojae*. Plant Mol Biol 34:255–264

Bhattacharyya MK, Narayanan NN, Gao H, Santra DK, Salimath SS, Kasuga T, Liu Y, Espinosa B, Ellison L, Marek L, Shoemaker R, Gijzen M, Buzzell RI (2005) Identification of a large cluster of coiled coil-nucleotide binding site-leucine rich repeat-type genes from the *Rps1* region containing Phytophthora resistance genes in soybean. Theor Appl Genet 111:75–86

Birt DF, Hendrick S, Alekel DL, Anthony M (2004) Soybean and the prevention of chronic human disease. In: Boerma HR, Specht, JE (eds) Soybeans: improvement, production, and uses. Agron Monogr, 3rd edn. No. 16, ASA-CSSA-SSSA, Madison, WI, pp 1047–1117

Blum A (1988) Plant breeding for stress environments. CRC Press, Boca Raton, FL, p 209

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314–331

Boutin S, Ansari H, Concibido V, Denny R, Orf J, Young N (1992) RFLP analysis of cyst nematode resistance in soybean. Soybean Genet Newsl 19:123–127

Boyer JS (1982) Plant productivity and environment. Science 218:443–448

Bray EA, Bailey-Serres J, Weretilnyk E (2000) Responses to abiotic stresses. In: Gruissem W, Buchannan B, Jones R (eds) Biochemistry and molecular biology of plants. Am Soc Plant Physiologists, pp 1158–1249

Broun P, Tanksley SD (1996) Characterization and genetic mapping of simple sequence repeat in the tomato genome. Mol Gen Genet 250:39–49

Brummer EC, Graef GL, Orf J, Wilcox JR, Shoemaker RC (1997) Mapping QTL for seed protein and oil content in eight soybean populations. Crop Sci 37:370–378

Burnham KD, Dorrance AE, Francis DM, Fioritto RJ, Martin SK St. (2003) *Rps8*, a new locus in soybean for resistance to *Phytophthora sojae*. Crop Sci 43:101–105

Caetano-Anolles G, Bassam BJ, Gresshoff PM (1992) Primer-template interactions during DNA amplification fingerprinting with single arbitrary oligonucleotides. Mol Gen Genet 235:157–165

Caldwell BE, Brim CA, Ross JP (1960) Inheritance of resistance of soybeans to the cyst nematode, *Heterodera glycines*. Agron J 52:635–636

Cannon SB, McCombie WR, Sato S, Tabata S, Denny R, Palmer L, Katari M, Young ND, Stacey G (2003) Evolution and microsynteny of the apyrase gene family in three legume genomes. Mol Genet Genomics 270:347–361

Chalhoub BA, Thibault S, Laucou V, Rameau C, Hofte H, Cousin R (1997) Silver staining and recovery of AFLP amplification products on large denaturing polyacrylamide gels. BioTechniques 22:216–220

Chang SJC, Doubler TW, Kilo VY, AbuThredeih J, Prabhu R, Freire V, Suttner R, Klein J, Schmidt ME, Gibson PT, Lightfoot DA (1997) Association of loci underlying field resistance to soybean sudden death syndrome (SDS) and cyst nematode (SCN) race 3. Crop Sci 37:965–971

Cho YG, Ishii T, Temnykh S, Chen X, Lipovich L, McCouch SR, Park WD, Ayres N, Cartinhour S (2000) Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.) Theor Appl Genet 100:713–722

Chung J, Cregan PB, Shoemaker RC, Specht JE, Lee DJ, Babka HL, Graef GL, Staswick PE (2003) The seed protein, oil, and yield QTL on soybean linkage group I. Crop Sci 43:1053–1067

Collins FS, Brooks LD, Charkravarti A (1998) A DNA polymorphism discovery resource for research on human genetic variation. Genome Res 8:1229–1231

Concibido VC, Denny RL, Boutin SR, Hautea R, Orf JH, Young ND (1994) DNA marker analysis of loci underlying resistance to soybean cyst nematode (*Heterodera glycines* Ichinohe). Crop Sci 34:240–246

Concibido VC, Young ND, Lange DA, Denny RL, Orf JH (1996) RFLP mapping and marker–assisted selection of soybean cyst nematode resistance in PI 209332. Crop Sci 36:1643–1650

Concibido VC, Lange D, Denny RL, Orf J, Young N (1997) Genome mapping on soybean cyst nematode resistance genes in 'Peking', PI 90763, and PI 88788 using DNA markers. Crop Sci 37:258–264

Cregan PB, Bhagwat AA, Akkaya MS, Rongwen J (1994) Microsatellite fingerprinting and mapping of soybean. Methods Cell Mol Biol 5:49–61

Concibido VC, Diers BW, Arelli PR (2004) A decade of QTL mapping for cyst nematode resistance in soybean. Crop Sci 44:1121–1131

Cregan PB, Jarvik T, Bush AL, Shoemaker RC, Lark KG, Kahler AL, Kaya N, VanToai TT, Lohnes DG, Chung J, Specht JE (1999a) An integrated genetic linkage map of the soybean. Crop Sci 39: 1464–1490

Cregan, PB, Mudge J, Fickus EW, Danesh D, Denny R, Young ND (1999b) Two simple sequence repeat markers to select for soybean cyst nematode resistance conditioned by the *rhg1* locus. Theor Appl Genet 99:811–818

Cregan PB, Choi IY, Hyten DL, Song QJ, Matukumalli LK, Yoon MS, Yi SI, Reiter RS, Lee MS, Chase K, Lark KG, Shoemaker RC, Specht JE (2006) A SNP-based soybean genome map and applications in soybean breeding and genetics. (Abstract) The 11th Biennial Conference on Molecular and Cellular Biology of the Soybean, August 5–8, Lincoln, NE

Dekkers JCM, Hospital F (2002) The use of molecular genetics in the improvement of agricultural populations. Nat Rev Genet 3:22–32

Demirbas A, Rector BG, Lohnes DG, Fioritto RJ, Graef GL, Cregan PB, Shoemaker RC (2001) Simple sequence repeat markers linked to the soybean *Rps* genes for Phytophthora resistance. Crop Sci 41:1220–1227

Diers BW, Shoemaker RC (1992) Restriction fragment length polymorphism analysis of soybean fatty acid content. J Am Oil Chem Soc 69:1242–1244

Diers BW, Keim P, Fehr WR, Shoemaker RC (1992a) RFLP analysis of soybean seed protein and oil content. Theor Appl Genet 83:608–612

Diers BW, Mansur L, Imsande J, Shoemaker (1992b) Mapping Phytophthora resistance loci in soybean with restriction fragment length markers. Crop Sci 32:377–383

Dixon MS, Jones DA, Keddie JS, Thomas CM, Harrison K, Jones JDG (1996) The tomato *Cf-2* disease resistance locus comprises two functional genes encoding leucine-rich repeat proteins. Cell 84:451–459

Eujayl I, Sorrells ME, Baum M, Wolters P (2002) Isolation of EST-derived mircosatellite markers for genotyping the A and B genomes of wheat. Theor Appl Genet 104:399–407

Ferreira RA, Foutz AB, Keim P (2000) Soybean genetic map of RAPD markers assigned to an existing scaffold RFLP map. J Hered 91:392–396

Gardner ME, Hymowitz T, Xu SJ, Hartman GL (2001) Physical map location of the *Rps1-k* allele in soybean. Crop Sci 41:1435–1438

Gordon SG, Steven KM, Dorrance AE (2006) *Rps*8 maps to a resistance gene rich region on soybean molecular linkage group F. Crop Sci 46:168–173

Grau CR, Dorrance AE, Bond J, Russin J (2004) Fungal diseases. In: Boerma HR, Specht JE (eds) Soybeans: improvement, production, and uses. Agron Monogr, 3rd edn. No. 16, ASA-CSSA-SSSA, Madison, WI, pp 679–763

Gu Z, Hiller L, Kwok PY (1998) Single nucleotide polymorphism hunting in cyberspace. Hum Mutat 12:221–225

Guo B, Sleper DA, Nguyen HT, Arelli PR, Shannon JG (2006) Quantitative trait loci underlying resistance to three soybean cyst nematode populations in soybean PI404198A. Crop Sci 46:224–233

Hancock JA, Hancock FG, Caviness CE, Riggs RD (1987) Genetics of resistance in soybean to "Race X" of soybean cyst nematode. Crop Sci 27:704–707

Hans CS, Jackson S (2006) A cytogenetic approach to assign linkage groups to chromosomes. (Abstract) The 11th Biennial Conference on Molecular and Cellular Biology of the Soybean, August 5–8, Lincoln, NE

Hartwig EE (1985) Breeding productive soybean with resistance to soybean cyst nematode. In: Shibles RA (ed) World Soybean Research Conference III. Westview Press, Boulder, CO, pp 394–399

Hauge BM, Wang ML, Parsons JD, Parnell LD (2001) Nucleic acid molecules and other molecules associated with soybean cyst nematode resistance. US Pat Appl Publ No. 20030005491

Heer JA, Knap HT, Mahalingam R, Shipe ER, Arelli PR, Matthews BF (1998) Molecular markers for resistance to *Heterodera glycines* in advanced soybean germplasm. Mol Breed 4:359–367

Hermann FJ (1962) A revision of the genus *Glycine* and its immediate allies. USDA Tech Bull 1268, USDA, Washington DC

Hurburgh CR Jr, Brumm TJ, Guinn JM, Hartwig RA (1990) Protein and oil patterns in US and world soybean markets. J Am Oil Chem Soc 67:966–973

Hymowitz T (2004) Speciation and cytogenetics. In: Boerma HR, Specht JE (eds) Soybeans: improvement, production, and uses. Agron Monogr, 3rd edn. No. 16, ASA-CSSA-SSSA, Madison, WI, pp 97–136

Hymowitz T, Singh RJ (1987) Taxonomy and speciation. In: Wilcox JR (ed) Soybeans: improvement, production, and uses. Agron. Monogr, 2nd edn. No. 16, ASA-CSSA-SSSA, Madison, WI, pp 23–45

Hymowitz T, Singh RJ, Kollipara KP (1998) The genomes of the glycine. Plant Breed Rev 16:289–317

Hyten DL, Cregan PB (2006) Saturation of the *rhg1* genomic region with SNP markers to determine linkage drag in resistant soybean cultivars and to demonstrate association analysis in soybean. In: Abstracts of plant & animal genome XIV conference, Jan 14–18, San Diego, CA, p 205

Hyten DL, Pantalone VR, Sams CE, Saxton AM, Landau-Ellis D, Stefaniak TR, Schmidt ME (2004) Seed quality QTL in a prominent soybean population. Theor Appl Genet 109:552–561

Jeong SC, Maroof MAS (2004) Detection and genotyping of SNPs tightly linked to two disease resistance loci, *Rsv1* and *Rsv3*, of soybean. Plant Breed 123:305–310

Joseph B, Specht J, Shoemaker RC (2004) Cloning the gene(s) underlying a major protein QTL on soybean linkage group-I. In: Soy2004 10th Biennial conference of the cellular and molecular biology of the soybean, Columbia, MO, p 73

Kabelka EA, Carlson SR, Diers BW (2005) Localization of two loci that confer resistance to soybean cyst nematode from *Glycine soja* PI 468916. Crop Sci 45:2473–2481

Kahl G, Mast A, Tooke N, Shen R, van de Boom D (2005) Single nucleotide polymorphisms: Detection techniques and their potential for genotyping and genome mapping. In: Meksem K, Kahl G (eds) The handbook of plant genome mapping. Wiley-VCH, Verlag GmbH & Co., KGaA, pp 75–107

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

Kao WY, Tsai TT, Tsai HC, Shih CN (2006) Response of three *Glycine* species to salt stress. Environ Expt Bot 56:120–125

Kassem MA, Meksem K, Iqbal MJ, Njiti VN, Banz WJ, Winters TA, Wood A, Lightfoot DA (2004) Definition of soybean genomic regions that control seed phytoestrogen amounts. J Biomed Biotechnol 1:52–60

Kasuga T, Salimath SS, Shi JR, Gijzen M, Buzzell RI, Bhattacharyya MK (1997) High resolution genetic and physical mapping of molecular markers linked to the *Phytophthora* resistance gene *Rps1-k* in soybean. MPMI 10:1035–1044

Kawabe A, Yamane K, Miyashita NT (2000) DNA polymorphism at the cytosolic phosphoglucose isomerase (*PgiC*) locus of the wild plant *Arabidopsis thaliana*. Genetics 156:1339–1347

Keim P, Shoemaker RC, Palmer RG (1989) Restriction fragment length polymorphism diversity in soybean. Theor Appl Genet 77:786–792

Keim P, Diers BW, Olson TC, Shoemaker RC (1990) RFLP mapping in soybean: association between marker loci and variation in quantitative traits. Genetics 126:735–742

Keim P, Schupp JM, Travis SE, Clayton K, Zhu T, Shi L, Ferreira A, Webb DM (1997) A high-density soybean genetic map based on AFLP markers. Crop Sci 37:537–543

Kim MY, Ha BK, Jun TH, Hwang EY, Van K, Kuk YI, Lee SH (2004) Single nucleotide polymorphism discovery and linkage mapping of lipoxygenase-2 gene (*Lx2*) in soybean. Euphytica 135:169–177

Kim HK, Kang ST, Cho JH, Choung MG, Suh DY (2005a) Quantitative trait loci associated with oligosaccharide and sucrose contents in soybean (*Glycine max* L.). J Plant Biol 48:106–112

Kim MY, Van K, Lestari P, Moon JK, Lee SH (2005b) SNP identification and SNAP marker development for a *GmNARK* gene controlling supernodulation in soybean. Theor Appl Genet 110:1003–1010

Kuittinen H, Aguade M (2000) Nucleotide variation at the *CHALCONE ISOMERASE* locus in *Arabidopsis thaliana*. Genetics 155:863–872

Kwok PY, Deng Q, Zakeri H, Nickerson DA (1996) Increasing the information content of STS-based genome maps: identifying polymorphisms in mapped STSs. Genomics 31:123–126

Lackey J (1980) Chromosomes numbers in the Phaseoleae (Fabaceae: Faboideae) and their relation to taxonomy. Am J Bot 67:595–602

Lai Z, Livingstone K, Zou Y, Church SA, Knapp SJ, Andrews J, Rieseberg LH (2005) Identification and mapping of SNPs from ESTs in sunflower. Theor Appl Genet 111:1532–1544

Lee M (1995) DNA markers and plant breeding programs. Adv Agron 35:265–344

Lee SH, Bailey MA, Mian MAR, Carter TE, Shipe ER, Ashley DA, Parrott WA, Hussey RS, Boerma HR (1996) RFLP loci associated with soybean protein and oil content across populations and locations. Theor Appl Genet 93:646–657

Lee GJ, Carter TE Jr, Li Z, Gibbs MO, Boerma HR, Villagarcia MR, Zhou X (2004) A major QTL conditioning salt tolerance in S-100 soybean and descendent cultivars. Theor Appl Genet 109: 1610–1619

Li YC, Korol AB, Fahima T, Beiles A, Nevo E (2002a) Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. Mol Ecol 11:2453–2465

Li Z, Wilson RF, Rayford WE, Boerma HR (2002b) Molecular mapping genes conditioning reduced palmitic acid content in N87-2122-4 soybean. Crop Sci 42:373–378

Lightfoot D, Meksem K (2002) Isolated polynucleotides and polypeptides relating to loci underlying resistance to soybean cyst nematode and soybean sudden death syndrome and methods employing same. US Pat Appl Publ No. 2002144310

Lindbad-Toh K, Winchester E, Daly MJ, Wang DC, Hirschhorn JN et al (2000) Large-scale discovery and genotyping of single nucleotide polymorphisms in the mouse. Nat Genet 24:381–385

Linnaeus C (1737) Hortus cliffortianus, historiae naturalis classica. In: Cramer J, Swann HK (eds) Vol. 63 Reprint 1968. Stechert-Hafner Service Agency, NY, p 499

Litt M, Luty JA (1989) A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. Am J Hum Genet 44:397–401

Lohnes DG, Schmitthenner AF (1997) Position of the Phytophthora gene *Rps7* on the soybean molecular map. Crop Sci 37:555–556

Lu H, Redus MA, Coburn JR, Rutger JN, McCouch SR, Tai TH (2005) Population structure and breeding patterns of 145 US rice cultivars based on SSR marker analysis. Crop Sci 45:66–76

Luo MC, Thomas C, You FM, Hsiao J, Ouyang S, Buell CR, Malandro M, McGuire PE, Anderson OD, Dvorak J (2003) High-throughput fingerprinting of bacterial artificial chromosomes using the snapshot labeling kit and sizing of restriction fragments by capillary electrophoresis. Genomics 82:378–389

Luo Q, Yu B, Liu Y (2005) Differential sensitivity to chloride and sodium ions in seedlings of *Glycine max* and *G. soja* under NaCl stress. J Plant Physiol 162:1003–1012

Lusas EW (2004) Soybean processing and utilization. In: Boerma HR, Specht JE (eds) Soybeans: improvement, production, and uses. Agron Monogr, 3rd edn. No. 16, ASA-CSSA-SSSA, Madison, WI, pp 949–1036

Maguire TL, Grimmond S, Forrest A, Iturbe-Ormaetxe I, Meksem K, Gresshoff P (2002) Tissue-specific gene expression in soybean (*Glycine max*) detected by cDNA microarray analysis. J Plant Physiol 159:1361–1374

Mansur LM, Lark KG, Kross H, Oliveira A (1993) Interval mapping of quantitative trait loci for reproductive, morphological, and seed traits of soybean (*Glycine max* L.). Theor Appl Genet 86: 907–913

Matson AL, Williams LF (1965) Evidence of a fourth gene for resistance to the soybean cyst nematode. Crop Sci 5:477

Maughan PJ, Saghi Maroof MA, Buss GR (1995) Microsatellite and amplified sequence length polymorphism in cultivated and wild soybean. Genome 38:715–723

Maughan PJ, Maroof MAS, Buss GR (2000) Identification of quantitative trait loci controlling sucrose content in soybean (*Glycine max*). Mol Breed 6:105–111

McCarthy JJ, Hilfiker R (2000) The use of single nucleotide polymorphism maps in pharmacogenetics. Nat Biotechnol 18:505–508

Meksem K, Pantazopoulos P, Njiti VN, Hyten LD, Arelli PR, Lightfoot DA (2001) 'Forrest' resistance to the soybean cyst nematode is bigenic: saturation mapping of the *Rhg1* and *Rhg4* loci. Theor Appl Genet 103:71–717

Merrill ED (1917) An interpretation of Rumphius's Herbarium Amboinense. Bureau of Printing, Manila, Philippines

Mian MAR, Mailey MA, Ashley DA, Wells R, Carter TE, Parrot WA, Boerma HR (1996) Molecular markers associated with water use efficiency and leaf ash in soybean. Crop Sci 36:1252–1257

Mian MAR, Ashley DA, Boerma HR (1998) An additional QTL for water use efficiency in soybean. Crop Sci 38:390–393

Michelmore RW, Paran I, Kesseli RV (1991) Identification of marker 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 88:9828–9832

Monna L, Ohta R, Masuda H, Koike A, Minobe Y (2006) Genome-wide searching of single-nucleotide polymorphisms among eight distantly and closely related rice cultivars (*Oryza sativa* L.) and a wild accession (*Oryza rufipogon* Griff). DNA Res 13:43–51

Monteros MJ, Burton JW, Boerma HR (2004) SSR analysis and confirmation of oleic acid QTL in N00-3350. In: Soy2004 the 10th biennial conference of the cellular and molecular biology of the soybean, Columbia, MO, p 58

Morgante M, Olivieri AM (1993) PCR-amplification microsatellites as markers in plant genetics. Plant J 3:175–182

Morgante M, Rafalski JA, Biddle P, Tingey S, Olivieri AM (1994) Genetic mapping and variability of seven soybean simple sequence repeat loci. Genome 37:763–769

Mudge J, Cannon SB, Kalo P, Oldroyd GE, Roe BA, Town CD, Young ND (2005) Highly syntenic regions in the genomes of soybean, *Medicago truncatula*, and *Arabidopsis thaliana* [Online]. BMC Plant Biol. Available at http://www.biomedcentral.com/1471–2229/5/15 (Verified August 30, 2006)

Muehlbauer GJ, Staswick PE, Specht JE, Graef GL, Shoemaker RC, Keim P (1991) RFLP mapping using near-isogenic lines in the soybean [*Glycine max* (L.) Merr.]. Theor Appl Genet 81:189–198

Myburg AA, Remington DL, O'Malley DM, Sederoff RR, Whetten RW (2001) High-throughput AFLP analysis using infrared dye-labeled primers and an automated DNA sequencer. Biotechnologies 30:348–357

Nelson WM, Bharti AK, Butler E, Wei F, Fuks G, Kim H, Wing RA, Messing J, Soderlund C (2005) Whole-genome validation of high-information-content fingerprinting. Plant Physiol 139:27–38

Nguyen HT, Wu X (2005) Molecular marker systems for genetic mapping. In: Meksem K, Kahl G (eds.) The handbook of plant genome mapping. Wiley-VCH, Verlag GmbH & Co., KGaA, pp 23–50

Nichols DM, Glover KD, Carlson SR, Specht JE, Diers BW (2006) Fine mapping of a seed protein QTL on soybean linkage group I and its correlated effects on agronomic traits. Crop Sci 46:834–839

Nilsen ET, Orcutt DM (1996) The physiology of plants under stress. Vol. 1: Abiotic factors. Wiley, New York, p 704

Ohashi H, Tateishi Y, Huang TC, Chen TT (1984) Leguminosae in Taiwan. Sci Rep Tohoku University 4th Ser. (Biology) 38:315

Olsen KM, Womace A, Garrett AR, Suddith JI, Purugganan MD (2002) Contrasting evolutionary forces in the *Arabidopsis thaliana* floral development pathway. Genetics 160:1641–1650

Oosterhuis DM, Scott HD, Hampton RE, Wullschleger SD (1990) Physiological response of two soybean [*Glycine max*, L. Merr.] cultivars to short-term flooding. Env Exp Bot 30:85–92

Orcutt DM, Nilsen ET (2000) The physiology plants under stress: soil and biotic factors. Wiley, New York, p 696

Orf JH, Chase K, Jarvik T, Mansur LM, Cregan PB, Adler FR, Lark KG (1999) Genetics of soybean agronomic traits. I. Comparison of three related recombinant inbred populations. Crop Sci 39:1642–1651

Orf JH, Diers BW, Boerma HR (2004) Genetic improvement: conventional and molecular strategies. In: Boerma HR, Specht JE (eds) Soybeans: improvement, production, and uses. Agron Monogr, 3rd edn. No. 16, ASA-CSSA-SSSA, Madison, WI, pp 417–480

Pagel J, Walling JG, Young ND, Shoemaker RC, Jackson SA (2004) Segmental duplications within the *Glycine max* genome revealed by fluorescence in situ hybridization of bacterial artificial chromosomes. Genome 47:764–768

Panthee DR, Pantalone VR, West DR, Saxton AM, Sams CE (2005) Quantitative trait loci for seed protein and oil content, and seed size in soybean. Crop Sci 45:2015–2022

Panthee DR, Pantalone VR, Sams CE, Saxton AM, West DR, Orf JH, Killam AS (2006) Quantitative trait loci controlling sulfur containing amino acids, methionine and cysteine, in soybean seeds. Theor Appl Genet 112:546–53

Primomo VS, Poysa V, Ablett GR, Jackson CJ, Gijzen M, Rajcan I (2005) Mapping QTL for individual and total isoflavone content in soybean seeds. Crop Sci 45:2454–2464

Purcell LC, Specht JE (2004) Physiological traits for ameliorating drought stress. In: Boerma HR, Specht JE (eds) Soybeans: improvement, production, and uses. Agron Monogr, 3rd edn. No. 16, ASA-CSSA-SSSA, Madison, WI, USA, pp 569–620

Putnay SD, Herlihy WC, Schimmel P (1983) A new troponin T and cDNA clones for 13 different muscle proteins, found by shortgun sequencing. Nature (London) 302:718–721

Qui BX, Arelli PR, Sleper DA (1999) RFLP markers associated with soybean cyst nematode resistance and seed composition in a 'Peking' × 'Essex' population. Theor Appl Genet 98:356–364

Rafalski A, Tingey S (1993) RFLP map of soybean (*Glycine max*). In: O'Brien SJ (ed) Genetic maps: locus maps of complex genomes. Cold Spring Harbor Laboratory Press, New York, pp 149–156

Rao-Arelli, AP (1994) Inheritance of resistance to *Heterodera glycines* race 3 in soybean accessions. Plant Dis 78:898–900

Raper CD Jr, Kramer PJ (1987) Stress physiology. In: Wilcox JR (ed) Soybeans: improvement, production and uses. Agron Monogr, 2nd edn. No. 16, ASA, CSSA, and SSSA, Madison, WI, pp 589–642.

Reyna N, Cornelious B, Shannon JG, Sneller CH (2003) Evaluation of a QTL for waterlogging tolerance in southern soybean germplasm. Crop Sci 43:2077–2082

Rongwen J, Akkaya MS, Lavi U, Cregan PB (1995) The use of microsatellite DNA marker for soybean genotype identification. Theor Appl Genet 90:43–48

Rostoks N, Mudie S, Cardle L, Russell J, Ramsay L, Booth A, Svensson JT, Wanamaker SI, Walia H, Rodriguez EM, Hedley PE, Liu H, Morris J, Close TJ, Marshall DF, Waugh R (2005) Genome-wide SNP discovery and linkage analysis in barley based on genes responsive to abiotic stress. Mol Genet Genomics 274:515–527

Sanchez de la Hoz MP, Davila JA, Loarce Y, Ferrer E (1996) Simple sequence repeat primers used in polymerase chain reaction amplification to study genetic diversity in barley. Genome 39:112–117

Scallon BJ, Dickinson CD, Nielsen NC (1987) Characterization of a null-allele for the *Gy4* glycinin gene from soybean. Mol Gen Genet 208:107–113

Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270:467–470

Schmitthenner AF (1985) Problems and progress in control of Phytophthora root rot of soybean. Plant Dis 69:362–368

Schmitthenner AF, Hobe M, Bhat RG (1994) *Phytophthora sojae* races in Ohio over a 10-year interval. Plant Dis 78:269–276

Scott HD, DeAngulo J, Daniels MB, Wood LS (1989) Flood duration effects on soybean growth and yield. Agron J 81:631–636

Sebolt AM, Shoemaker RC, Diers BW (2000) Analysis of a quantitative trait locus allele from wild soybean that increases seed protein content in soybean. Crop Sci 40:1438–1444

Seki M, Satuo M, Sakurai T, Akiyama K, Lida K, Ishida J, Nakajima M, Enju A, Narusaka M, Miki Fujita M, Oono Y, Ayako Kamei A, Yamaguchi-Shinozaki K, Shinozaki K (2005) RIKEN *Arabidopsis* full-length (RAFL) cDNA and its applications for expression profiling under abiotic stress conditions. J Exp Bot 55:213–223

Shalon D, Smith S, Brown P (1996) A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. Genome Res 6:639–645

Shan X, Blake TK, Talbert LE (1999) Conversion of AFLP markers to sequence-specific PCR markers in barley and wheat. Theor Appl Genet 98:1072–1078

Shoemaker RC, Olson TC (1993) Molecular linkage map of soybean (*Glycine max* L. Merr.). In: O'Brien SJ (ed) Genetic maps: locus maps of complex genomes. Cold Spring Harbor Press, New York, pp 6131–6138

Shoemaker RC, Hatfield PM, Palmer RG, Atherly AG (1986) Chloroplast DNA variation in the genus *Glycine* subgenus *Soja*. J Hered 87:308–313

Shoemaker RC, Polzin K, Labate J, Specht J, Brummer EC, Olson T, Young N, Concibido V, Wilcox J, Tamulonis JP, Kochert G, Boerma HR (1996) Genome duplication in soybean (*Glycine* subgenus *soja*). Genetics 144:329–338

Shoemaker RC, Cregan PB, Vodkin LO (2004) Soybean genomics. In: Boerma HR, Specht JE (eds) Soybeans: improvement, production, and uses. Agron Monogr, 3rd edn. No. 16, ASA-CSSA-SSSA, Madison, WI, pp 235–263

Shultz JL, Kurunam D, Shopinski K, Iqbal MJ, Kazi S, Zobrist K, Bashir R, Yaegashi S, Lavu N, Afzal AJ, Yesudas CR, Kassem MA, Wu C, Zhang HB, Town CD, Meksem K, Lightfoot DA (2006a) The Soybean Genome Database (SoyGD): a browser for display of duplicated, polyploid, regions and sequence tagged sites on the integrated physical and genetic maps of *Glycine max*. Nucleic Acids Res 34:D758–65

Shultz JL, Yesudas CR, Yaegashi S, Afzal J, Kazi S, Lightfoot DA (2006b) Three minimum tile paths from bacterial artificial chromosome libraries of the soybean (*Glycine max* cv. 'Forrest'): Tools for structural and functional genomics [Online]. Plant Methods Available at http://www.plantmethods.com/content/2/1/9 (Verified August 30, 2006)

Singh RJ, Hymowitz T (1989) The genomic relationship among *Glycine soja* Sieb and Zucc, *G max* (L.) Merr and '*G. gracillis*' Skvortz. Plant Breed 103:171–173

Singh RJ, Hymowitz T (1999) Soybean genetic resources and crop improvement. Genome 42:605–616

Skvortzow BV (1927) The soybean-wild and cultivated in Eastern Asia. Proc Manchurian Res Soc Pub Ser. A Nat History Sec 22:1–8

Soderlund C, Longden I, Mott R (1997) FPC: a system for building contigs from restriction fingerprinted clones. CABIOS 13:523–535

Soderlund C, Engler F, Hatfield J, Blundy S, Chen M, Yeisoo Y, Wing R (2001) Mapping sequence to rice FPC. In: Computational biology and genome Informatics, World Scientific Publishing

Song QJ, Marek LF, Shoemaker RC, Lark KG, Concibido VC, Delannay X, Specht JE, Cregan PB (2004) A new integrated genetic linkage map of the soybean. Theor Appl Genet 109:122–128

Specht JE, Germann M, Markwell JP, Lark KG, Orf JH, Macrander M, Chase K, Chung J, Graef GL (2001) Soybean response to water: a QTL analysis of drought tolerance. Crop Sci 41:493–509

Stombaugh SK, Orf JH, Jung HG, Chase K, Lark KG, Somers DA (2004) Quantitative trait loci associated with cell wall polysaccharides in soybean seed. Crop Sci 44:2101–2106

Tanksley SD, Young ND, Paterson AH, Bonierbale MW (1989) RFLP mapping in plant breeding: new tools for an old science. Biotechnology 7:257–264

Tenaillon MI, Sawkins MC, Long AD, Gaut RL, Doebley JF et al (2001) Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* spp *mays* L.). Proc Natl Acad Sci USA 98:9161–9166

Thibaud-Nissen F, Shealy RT, Khanna A, Vodkin LO (2003) Clustering of microarray data reveals transcript patterns associated with somatic embryogenesis in soybean. Plant Physiol 132:118–136

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

Tooley PW, Grau CR (1984) The relationship between rate-reducing resistance to *Phytophthora megasperma* f. sp. *glycinea* and yield of soybean. Phytopathology 74:1209–1216

Tooley PW, Grau CR (1986) Microplot comparison of rate-reducing and race-specific resistance to *Phytophthora megasperma* f. sp. *glycinea* in soybean. Phytopathology 76:554–557

Van K, Hwang EY, Kim MY, Park HJ, Lee SH, Cregan PB (2005) Discovery of SNPs in soybean genotypes frequently used as the parents of mapping populations in the United States and Korea. J Hered 96:529–535

VandenBosch KA, Stacey G (2003) Summaries of legume genomics projects from around the globe. Community resources for crops and models. Plant Physiol 131:840–865

VanToai TT, Beuerlein JE, Schmitthenner AF, Martin SK St (1994) Genetic variability for flooding tolerance in soybeans. Crop Sci 34:1112–1115

VanToai TT, Martin SK, Chase K, Boru G, Schnipke V, Schmitthenner AF, Lark KG (2001) Identification of a QTL associated with tolerance of soybean to soil waterlogging. Crop Sci 41:1247–1252

Vodkin LO, Khanna A, Shealy R, Clough SJ, Gonzalez DO, Philip R, Zabala G, Thibaud-Nissen F, Sidarous M, Stromvik M et al (2004) Microarrays for global expression constructed with a low redundancy set of 27,500 sequenced cDNAs representing an array of developmental stages and physiological conditions of the soybean plant. BMC Genomics 5:73

Vos P, Hogers R, Bleeker M, Rijans M, van der Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) A new technique for DNA fingerprinting. Nucleic Acids Res 23:4407–4414

Vroh Bi I, McMullen MD, Sanchez-Villeda H, Schroeder S, Gardiner J, Palacco M, Soderlund C, Wing R, Fang Z, Coe EH Jr (2006) Single nucleotide polymorphisms and insertion–deletions for genetic markers and anchoring the maize fingerprint contig physical map. Crop Sci 46:12–21

Walker AK, Schmitthenner AF (1984) Heritability of tolerance to Phytophthora rot in soybean. Crop Sci 24:490–491

Walker DR, Scaboo AM, Pantalone VR, Wilcox JR, Boerma HR (2006) Genetic mapping of loci associated with seed phytic acid content in CX1834-1-2 soybean. Crop Sci 46:390–397

Webb DM, Baltazar BM, Rao-Arelli PA, Schupp J, Clayton K, Keim P, Beavis WD (1995) Genetic mapping of soybean cyst nematode race-3 resistance loci in soybean PI437.654. Theor Appl Genet 91:574–581

Welsh J, McClelland M (1990) Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Res 18:7213–7218

Weng C, Yu K, Anderson TR, Poysa V (2001) Mapping genes conferring resistance to *Phytophthora* root rot of soybean, *Rps1a* and *Rps7*. J Hered 92:442–446

Williams JKG, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res 18:6531–6535

Wilson RF (2004) Seed composition. In: Boerma HR, Specht JE (eds) Soybeans: improvement, production, and uses. Agron Monogr, 3rd edn. No. 16, ASA-CSSA-SSSA, Madison, WI, pp 621–677

Winstead NN, Skotland CB, Sasser JN (1955) Soybean cyst nematode in North Carolina. Plant Dis Rep 39:9–11

Wrather JA, Koenning SR, Anderson TR (2003) Effect of diseases on soybean yields in the United States and Ontario (1999–2003) [Online]. Plant Heath Progress. doi:10.1094/PHP-2003-0325-01-RV. Available at http//www.plantmanagementnetwork.org/ (Verified August 28, 2006)

Wu XL, He CY, Chen SY, Zhuang BC, Wang KJ, Wang XC (2001) Phylogenetic analysis of interspecies in the genus *Glycine* through SSR markers (in Chinese). Act Genet Sin 28:359–366

Wu C, Sun S, Nimmakayala P, Santos FA, Meksem K, Springman R, Ding K, Lightfoot DA, Zhang HB (2004a) A BAC- and BIBAC-based physical map of the soybean genome. Genome Res 14:319–326

Wu X, Lee GJ, Huang S, Wan J, Stacey G, Nguyen H (2004b) Six dimensional BAC DNA pools – a new resource for soybean genome mapping. In: Soy2004 the 10th biennial conference of the cellular and molecular biology of the soybean, Columbia, MO, p 57

Wu X, Blake S, Pyatek K, Sleper D, Shannon G, Nguyen H (2005) SNP marker development for *rhg1* and *Rhg4* conferring SCN resistance in soybean. Life Sciences Week, University of Missouri-Columbia (poster)

Xu SJ, Singh RJ, Kollipara KP, Hymowitz T (2000) Primary trisomics in soybean: origin, identification, breeding behavior, and use in linkage mapping. Crop Sci 40:1543–1551

Xue ZT, Xu MI, Shen W, Zhuang NL, Hu WM, Shen SC (1992) Characterization of a *Gy4* glycine gene from soybean *Glycine max* cv. Forrest. Plant Mol Biol 18:897–908

Yan HH, Mudge J, Kim DJ, Shoemaker RC, Cook DR, Young ND (2004) Comparative physical mapping reveals features of microsynteny between *Glycine max*, *Medicago truncatula*, and *Arabidopsis thaliana*. Genome 47:141–55

Yates JL, Harris DK, Boerma HR (2004) Marker-assisted selection around a major QTL on LG-I increases seed protein content in backcross-derived lines of soybean. In: Soy2004 the 10th biennial conference of the cellular and molecular biology of the soybean, Columbia, MO, p 67

Young WP, Schupp JM, Keim P (1999) DNA methylation and AFLP marker distribution in the soybean genome. Theor Appl Genet 99:785–792

Yue P, Arelli PR, Sleper DA (2001a) Molecular characterization of resistance to *Heterodera glycines* in soybean PI 438489B. Theor Appl Genet 102:921–928

Yue P, Sleper DA, Arelli PR (2001b) Mapping resistance to multiple races of *Heterodera glycines* in soybean PI 89772. Crop Sci 41:1589–1595

Zaitsev VS, Naroditsky BS, Khavkin EE (2001) Homologs of the genes for receptor kinase proteins conferring plant resistance to pathogens. Comparative analysis of the homologs of the rice gene *Xa21* in Triticeae Institute of Agriculture Biotechnology, 42 Timiryazevskaya ul., Moscow 127550, Russia (http://www.ncbi.nlm.nih.gov)

Zhang WK, Wang YJ, Luo GZ, Zhang JS, He CY, Wu XL, Gai JY, Chen SY (2004) QTL mapping of ten agronomic traits on the soybean [*Glycine max* (L.) Merr.] genetic map and their association with EST markers. Theor Appl Genet 108:1131–1139

Zhao Q, Zhang Y, Cheng Z, Chen M, Wang S et al (2002) A fine physical map of the rice chromosome 4. Genome Res 12:817–823

Zhu T, Shi L, Doyle JJ, Keim P (1995) A single nucleotide locus phylogeny of soybean based on DNA sequence. Theor Appl Genet 90:991–999

Zhu YL, Song QJ, Hyten DL, Van Tassell CP, Matukumalli LK, Grimm DR, Hyatt SM, Fickus EW, Young ND, Cregan PB (2003) Single-nucleotide polymorphisms in soybean. Genet 163:1123–1134

Zou JJ, Singh RJ, Lee J, Xu SJ, Cregan PB, Hymowitz T (2003) Assignment of molecular linkage groups to soybean chromosomes by primary trisomics. Theor Appl Genet 107:745–750

Zou J, Rodriguez-Zas S, Aldea M, Li M, Zhu J, Gonzalez DO, Vodkin LO, DeLucia E, Clough SJ (2005) Expression profiling soybean response to *Pseudomonas syringae* reveals new defense-related genes and rapid HR-specific down regulation of photosynthesis. Mol Plant Microbe Interact 18:1161–1174

CHAPTER 12

APPLICATION OF GENOMICS TO FORAGE CROP BREEDING FOR QUALITY TRAITS

THOMAS LÜBBERSTEDT*

University of Århus, Faculty of Agricultural Sciences, Research Centre Flakkebjerg, 4200 Slagelse, Denmark

Abstract: Forage quality depends on the digestibility of fodder, and can be directly measured by the intake and metabolic conversion in animal trials. However, animal trials are time-consuming, laborious, and thus expensive. It is not possible to study thousands of plant genotypes, as required in breeding programs. Therefore, several indirect methods including near-infrared reflectance spectroscopy (NIRS) have been established to overcome this limitation. However, the ideal indirect system for the prediction of forage performance would be based on gene-derived "functional" DNA markers, allowing early selection ultimately without need of field trials, and being environment independent. In addition, once identified relevant genes controlling forage quality are targets for transgenic approaches. Substantial progress has recently been achieved in the development and application of genomic tools both in model species and major forage crops such as ryegrass and alfalfa. Key genes involved in developmental and biochemical pathways affecting forage quality such as cell-wall, lignin, fructan, and tannin biosynthesis have been isolated and characterized. For some of these genes, allelic variation has been studied in detail and sequence motifs with likely effect on forage quality have been identified by association studies. Moreover, transgenic approaches substantiated the effect of several of these genes on forage quality. Perspectives and limitations of these findings for forage crop breeding are discussed given expected further progress in forage crop genomics, but also the complexity of the trait complex forage quality, since typically species mixtures of heterogeneous and heterozygous genotypes are grown in the field.

1. INTRODUCTION

1.1. Direct Measurement of Forage Quality

In most forage grasses and legumes, above-ground parts are harvested before or during flowering. An exception is forage maize, harvested after seed-set. Animal feeding trials allow direct evaluation of forage crop quality via the target product

*Address for correspondence: Thomas.Luebberstedt@agrsci.dk.

R.K. Varshney and R. Tuberosa (eds.), Genomics Assisted Crop Improvement: Vol. 2: Genomics Applications in Crops, 281–306.

such as milk or beef. Different parameters were developed such as Mega Joule "Net-Energy-Lactation" ((MJ) NEL) (Groß 1979, Weißbach 1993) for milk production, "Kilo Starch-Units" (kStE) (Zimmer et al. 1980) for beef production and "Metabolizable Energy" (ME) (Menke and Huss 1987), all reflecting the energy density (J/kg) of forage dry matter (Boberfeld 1986). However, animal trials are time-consuming, laborious, and thus expensive. It is not possible to study thousands of plant genotypes, as required in breeding programs.

1.2. Indirect Measurement of Forage Quality

Several indirect biological, chemical, and physical methods for quality evaluation have been developed. Biological methods for quality evaluation can be subdivided into field, *in vitro*, and enzymatic methods. Field methods score the expected nutritive value of plant communities (Boberfeld 1986) based on the species composition. In case of forage maize, the proportion of ears in total dry matter has been used for quality evaluation (Zscheischler 1990). A widely used *in vitro* rumen digestion analysis was developed by Tilley and Terry (1963) using a two-step procedure - first rumen liquor and subsequently peptic hydrochloric acid to estimate the *in vitro* digestibility of organic matter. Another *in vitro* test employing rumen liquor determines gas production, protein and fat content to estimate NEL or StE (Menke and Steingaß 1987). Enzymatic methods use cellulase together with peptic hydrochloric acid to estimate NEL and StE (Kirchgessner and Kellner 1981).

Since digestibility is mainly limited by poorly digestible cell wall components, chemical methods for forage quality evaluation focus on the breakdown and characterization of cell wall fractions within the organic matter. Using detergents, Van Soest (1974) separated cell complexes into soluble cell content and insoluble "neutral detergent fibre" (NDF) representing mainly the cell wall fraction. By acidic detergents further fractionation into a lignin ("acid detergent lignin": ADL) and a non-lignin fraction (ADF-ADL; ADF: "acid detergent fibre") is possible. ADF values can be converted into NEL and StE estimates (Kirchgessner and Kellner 1981).

All above mentioned approaches are too laborious for routine quality evaluation of large numbers as required in plant breeding. In contrast, by near-infrared reflectance spectroscopy (NIRS) (Norris et al. 1976) large sample numbers can be investigated with low effort. Infrared spectra of ground materials (1400 to 2600 nm) can be employed to estimate a number of quality parameters, once calibrations to animal trials, biological, or chemical methods have been established. For most of the above mentioned quality characters, NIRS calibrations have been established for the major forage crops (e.g., Ralph et al. 2004, Barriere et al. 2003). Moreover, in view of developing DNA-markers for forage quality traits, NIRS allows quality evaluation at large scale as required for QTL mapping or association studies.

1.3. Trait Complexity

Forage crop breeding is characterized by different layers of complexity. Several factors need to be considered when determining forage quality, such as the fed

animal species, the forage plant species, the cropping system, and the method(s) used for forage quality evaluation. In addition, the optimum diet of animals depends on the product, such as beef or milk for cattle. Ruminants have a much better capability to digest fibrous carbohydrates compared to monogastrics and to convert poor quality protein and nonprotein nitrogen sources (Van Soest 1974). Furthermore, intake and digestion by animals depends on forage properties such as its dry matter content, particle size, and the ensiling process.

At the feed or plant level, a given variety is often only part of the animal diet fed together with minerals, additives etc. If used for grazing, a single variety typically is only part of a mixture of different grasses and legumes. Further aspects adding to the complexity are different ploidy levels within crops such as in ryegrass, and symbiosis with endophytes or root nodule bacteria. In contrast, many genetic experiments are performed under simplified conditions to establish sound phenotype – genotype associations, preferably (i) at the diploid level, (ii) in monoculture, (iii) at *per se* level, and (iv) for single spaced plants. Therefore, a crucial question is, whether results obtained in "artificial" experimental situations are transferable and, thus, valuable to operational breeding programs.

2. FACTORS INFLUENCING FORAGE QUALITY

2.1. Organ, Tissue- and Cell Level

Forage maize breeding in Europe has originally been based on the concept that the best hybrids for grain production are also optimal for forage use. However, coefficients of determination between grain and forage yield ($r^2 < 0.25$) were not sufficiently high to justify selection on the basis of grain yield (Vattikonda and Hunter 1983). In addition, substantial genotypic variation for digestibility and yield of stover was found among hybrids suggesting that the non-grain portion of the maize plant offers considerable potential for the improvement in yield and forage of silage maize (Dhillon et al. 1990; Geiger et al. 1986).

It has been shown for different forage species that especially the extent and mode of lignification is limiting both digestibility and intake (Barriere et al. 2004). Thus, major breeding criteria for forage grasses such as *Lolium perenne* are besides a maximum animal intake per time unit and a high digestibility, (1) a high biomass yield and (2) a minimum need for cuttings. This can be achieved by late- or not-flowering and long-leaved grass types (high leaf elongation ratio) with high quality also in older leaf parts showing lack of senescence (stay green types) (Wilman et al. 2004).

2.2. Cell Wall and Cell Content

Digestibility of the ear fraction in maize is close to 90% in contrast to about 50% of stover (Zimmer et al. 1990). Within stover, the lignin fraction shows by far the lowest digestibility (about 7%) compared to cellulose and hemicellulose. Therefore, a promising target for the improvement of feed quality in silage maize and

forage grasses is the lignin biosynthetic pathway, including monolignol biosynthesis and polymerization of monolignols, the latter being performed by peroxidases and laccases (Ralph et al. 2004). The digestibility of grasses as a general trait becomes markedly reduced during the course of the growing season. This reduction is largely caused by an increase in the content of poorly digestible cell wall structural components. In parallel, there is a decrease in the content of soluble carbohydrates – "sugars". Varieties of ryegrass with a high stable level of carbohydrates in the form of fructans have been shown to retain a high degree of digestibility throughout the growing season. Poorly digestible structural components create an imbalance between carbohydrate and protein levels during ruminant fermentation, leading to a loss of nitrogen (ammonia) to environment.

Forage legumes are highly digestible as compared to forage grasses. However, proteolysis and microbial deamination might lead to protein loss in the rumen, not fully compensated by post-ruminal absorption (Robbins et al. 2002). High digestion rates may result in protein foaming and rumen pasture bloat as a digestive disorder (Gruber et al. 2001). Moreover, stem lignification is also affecting forage legume quality (Reddy et al. 2005).

2.3. Biochemical Pathways

The lignin biosynthetic pathway is well characterized (Boudet et al. 1995), especially the common phenylpropanoid pathway starting with the deamination of phenylalanine and providing hydroxycinnamoyl CoAs. The enzymes involved in the common phenylpropanoid pathway are phenylalanine ammonia-lyase (PAL), cinnamate hydroxylase (C4H), coumarate hydroxylase (C3H), caffeic O-methyltransferases (COMT), ferulate hydroxylase (F5H), and hydroxycinnamate CoA ligases (4CL). In total 34 genes have been identified in the *A. thaliana* genome coding for enzymes in the monlignol biosynthesis (Raes et al. 2003). The end products of this common pathway, the hydroxycinnamoyl CoAs, are precursors of the major classes of phenolic compounds which accumulate in plant tissues, e.g. flavonoids, stilbenes, phenolamides as well as lignins. Subsequently, cinnamoyl CoA reductase (CCR) and cinnamyl alcohol dehydrogenase (CAD) are specifically involved in biosynthesis of the lignin monomers p-coumaryl, coniferyl, and sinapyl alcohol. In maize, genes for COMT (Collazo et al. 1992) and CAD (Halpin et al. 1998) have been isolated. Defect alleles of both genes have been shown to correspond to *brown midrib* mutations (COMT: *bm3* ; CAD: *bm1*) known for long time (Barriere and Argillier 1993). Several independent studies on *bm1* and *bm3* have proved already the concept of increasing silage quality by altering the lignin biosynthetic pathway (Barriere and Argillier 1993) and reducing lignin content. However, application of *bm* mutants in plant breeding has been hampered by their pleiotropic effects on yield characters and lodging.

Lignins exhibit a high degree of structural variability depending on the relative proportion of three monolignols, different types of interunit linkages, and the occurrence of non conventional lignin units within the polymer (Boudet and Grima-

Pettenati 1996). Polymerization of monolignols involves peroxidases and laccases in an oxidation step but is generally not well understood (Boudet et al. 1995). Laccase was the first enzyme shown to be able to polymerize lignin monomers *in vitro* (Freudenberg et al. 1958). Several studies indicated that laccase and laccase-like activities are closely correlated with lignin deposition in developing xylem (Davin et al. 1992). Nersissian and Shipp (2002) identified 19 laccases in the genome of *A. thaliana* . In case of peroxidases, association between allelic variation at one Prx locus with forage digestibility in maize has been demonstrated (Guillet-Claude et al. 2004b). Thus, a minimum of 70–80 genes code for enzymes that are directly involved in lignin formation, based on the small *A. thaliana* genome, which can be considered as candidate genes for cell-wall digestibility. Furthermore, gene families involved in cellulose or hemicellulose formation might as well affect cell wall digestibility (Barriere et al. 2003, Ralph et al. 2004).

Fructans are polymers of fructose, and have a general structure of a glucose linked to multiple fructose units (polyfructosylsucroses). In contrast to the uniform structure of bacterial fructans, plant fructans represent five major classes of structural distinct fructans according to the linkages between the fructose units; inulins, levans, inulin neoseries, levan neoseries, and mixed type levans. It is now known that four out of five different fructosyltransferases (FT), each with their own specificity, are needed to synthesize the wide variety of fructans found in plants; 1-SST (1-sucrose:sucrose fructosyltransferase), 1-FFT (1-fructan:fructan fructosyltransferase), 6-FFT (6-fructan:fructan fructosyltransferase), 6-SFT (Sucrose fructosyl 6-transferase) and 6G-FFT (fructan:fructan 6G-fructosyltransferase). Other enzymes that are involved in fructan degradation are fructanhydrolases (FH), and invertases. Fructanhydrolase is a beta-fructofuranosidase and can uncouple fructose units from fructans with sucrose as an end product. Invertase, which is active in the vacuole, cleaves one sucrose molecule into glucose and fructose. It's also capable of cleaving fructose molecules from smaller fructans. This fructan hydrolysing activity decreases with a higher degree of polymerization of the fructan. The type of fructans and FTs varies among different monocot species. Levan type fructans are abundant in Triticum, Hordeum and Bromus, whereas inulin types are characteristic to e.g. Lolium species. Interestingly, bifurcose (a product of 6-SFT and precursor to the levan type and one of the routes to the levan neoseries) has not been found in Lolium species. According to this, four enzymes would be necessary to account for the synthesis of the fructans identified in *L. perenne*, namely 1-SST, 1-FFT, 6G-FFT and 6FFT (Parvis et al. 2001). Genes coding respective enzymes in forage crops have been isolated recently (Gallagher et al. 2004, Chalmers et al. 2003).

For wool production under grazing conditions, the sulfur-containing amino acids methionine and cysteine are among the most limiting essential amino acids (Rogers 1990). Sunflower seed albumin 8 (SFA8) contains high levels of methionine and cysteine and is resistant to rumen degradation (McNabb et al. 1994). Chimeric constructs directing the SFA8 protein into endoplasmatic reticulum in tall fescue leaves resulted in an up to 0.2% accumulation of SFA8 in the soluble leaf protein fraction (Wang 2001).

A major role for reducing the high digestion rates in forage legumes has been assigned to condensed tannins (Gruber et al. 2001). Condensed tannins are polymeric flavonoids with protein-precipitating properties. Whereas high amounts of condensed tannins are detrimental to ruminant digestion, moderate levels (2–3% of dry matter) improve forage legume quality by reducing ruminal digestion rates and avoidance of protein foaming, and thus lead to higher rates of protein conversion into animal products (Robbins et al. 2002). Highly nutritious species such as white clover and alfalfa have a low level of endogeneous tannins as compared to tanniferous forages like Lotus. The initial steps in condensed tannin biosynthesis belong to the general flavonoid pathway and include enzymes like chalcone synthase. Whereas genes coding for enzymes of the general flavonoid have been isolated for long time, the first genes coding for enzymes of the condensed tannin specific pathway such as leucoanthocyanidin reductase have been isolated more recently (Tanner et al. 2003). As for the lignin biosynthesis pathway, regulatory genes coding for transfactors have been envisaged as targets to manipulate tannin content, such as myb- or myc-like genes (Gruber et al. 2001).

3. QTL FOR FORAGE QUALITY

Quantitative trait loci (QTL) mapping is a forward genetic approach, associating trait variation with genome regions, systematically covered with molecular markers. Several QTL mapping studies for forage traits have been conducted, e.g., in maize (Lübberstedt et al. 1997a, b, 1998; Ralph et al. 2004), and ryegrass (Cogan et al. 2005a).

3.1. Maize

The first published QTL in relation to forage quality were based on whole plant measurements (Lübberstedt et al. 1997a, b, 1998), in order to reflect agronomic performance. However, the ear fraction might mask some of the variation present for cell-wall digestibility in the stover. Nevertheless, analysis of multiple traits including dry matter yield (DMY), dry matter content (DMC), starch content (STC) and yield (STY), metabolizable energy content (MEC) and yield (MEY), and *in vitro* digestibility of organic matter (IVDOM) revealed the "architecture" of forage quality in whole plant samples. In total, 16 QTL were detected for IVDOM and 14 for MEY (Lübberstedt et al. 1997b, 1998) as the two major forage quality characters. MEY was closely correlated with DMY and plant height (PHT), i.e., biomass production, which was also reflected by several common QTL affecting these three characters. In contrast, IVDOM was negatively correlated with PHT, but positively with STC and MEC as also expressed by shared QTL regions. It was thus possible to identify those genome regions controlling IVDOM, independent of the starch (=ear) fraction.

Subsequently QTL for cell-wall digestibility and lignification traits were studied in different recombinant inbred line (RIL) populations (Mechin et al. 2001,

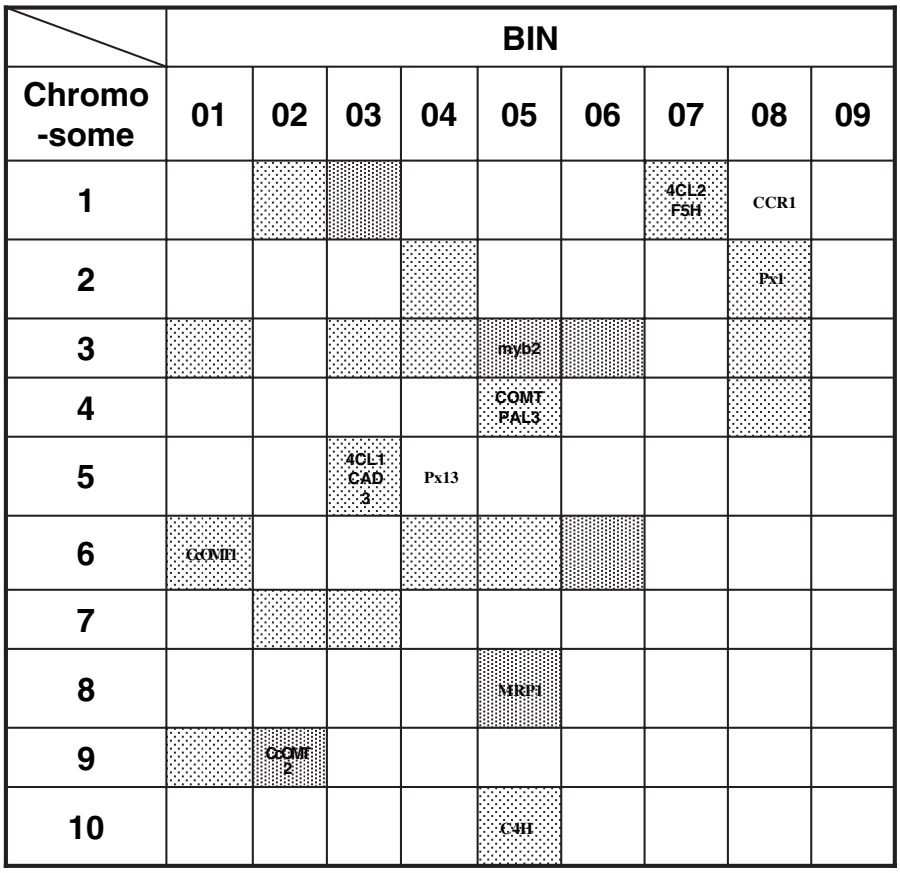


Figure 1. Assignment of mapped forage QTL (gray boxes; according to Ralph et al. 2004) to chromosomal bin regions (Gardiner et al. 1993) in maize. Dark gray boxes indicate "QTL hotspots", i.e., genome regions detected more than once. Colocalized candidate genes (e.g. "lignin genes") where included (abbreviations: see Ralph et al. 2004)

Roussel et al. 2002, Cardinal et al. 2003, Ralph et al. 2004). In total, 22 different QTL were identified for these characters (Figure 1), with some of the genome regions detected repeatedly (chromosomal bins 1.02, 1.03, 2.08, 3.05/06, 4.08, 4.09/10, 6.06, 9.02; Gardiner et al. 1993, Ralph et al. 2004). In many cases, QTL for cell-wall digestibility and lignin content co-localized. Moreover, several genes involved in monolignol biosynthesis mapped to the respective QTL regions including the isolated brown midrib genes coding for COMT (*bm3*) and CAD (*bm1*) (Ralph et al. 2004). However, each bin region of about 20 cM might still contain more than 1000 genes. Thus, additional evidence is required to proof which genes underlie those QTL controlling forage quality or lignification.

3.2. Forage Grasses and Forage Legumes

In forage grasses such as *Lolium perenne* both cell-wall digestibility especially of flower stems and high concentration of water soluble carbohydrates like fructans in leaves are considered as the main forage quality parameters (Yamada and Forster 2005). QTL for herbage traits in *L. perenne* have been determined for progeny of a paircross between a heterozygous and a doubled haploid genotype (Cogan et al. 2005b), forming the p150/112 reference population of the International *Lolium* Genome Initiative (ILGI). As major traits relating to ryegrass forage quality, *in vitro* dry-matter digestibility (IVDMD), neutral detergent fiber (NDF), estimated metabolisable energy (EstME), protein content (CP), and water soluble carbohydrates (WSC) have been determined on spaced plants employing NIRS calibrations. In total 42 QTL have been identified for these quality traits in six different experiments. Major QTL regions were located on linkage groups 3, 5, and 7. The QTL at linkage group 7 were located in the vicinity of three genes coding for enzymes involved in lignin biosynthesis, COMT, CCR1, and CAD.

QTL for WSC and WSC components have been identified in a *L. perenne* F_2 population (Turner et al. 2006), derived from contrasting grandparent plants with regard to WSC content. Plant samples were taken from tiller and leaf bases of spaced plants. While WSC was determined by a colorimetric assay, WSC components including concentrations of sucrose, glucose, fructose, and fructans at different degrees of polymerization were determined by HPLC. In total, 14 QTL for WSC and WSC components were detected, clustering on linkage groups 1, 2, 5, and 6 (Turner et al. 2006). Major QTL explaining more than 30% of the phenotypic variation were identified for WSC and fructan concentration on linkage group 6. Two loci coding for alkaline invertase were mapped to linkage group 6 in the vicinity of QTL for glucose and fructose concentration, which is in agreement with the biochemical role of alkaline invertase to break down sucrose into fructose and glucose.

Grasses can contribute to avoid malnutrition of ruminants, as shown for grass tetany (hypomagnesaemia) caused by grass varieties with low levels of magnesium. Yamada and Forster (2005) reported on the mapping of QTL for concentrations of multiple micro- and macronutrients determined by inductively-coupled plasma mass spectrososcopy in the p150/112 mapping population. In total, 45 QTL were identified, with major clusters on linkage groups 1, 2, 4, and 5, and QTL for magnesium content located on linkage groups 2 and 5.

In other forage grass species such as Festuca sp. and timothy, molecular markers and mapping populations have been developed more recently (e.g., Cai et al. 2003, Alm et al. 2003). Thus, results on forage quality in additional forage grass species can be expected in the near future.

Forage legumes are considered to be beneficial for the quality of forage mixtures including grasses. High digestion rates especially of the protein fraction might result in protein foaming or reduced conversion into ruminant biomass, which can be controlled by moderate levels of condensed tannins (see 2.3). Although marker-aided studies including QTL mapping have been performed in alfalfa (Brummer

et al. 2005), white and red clover (Sawbridge et al. 2003, Herrmann et al. 2006), investigations targeting forage legume quality are lacking so far.

3.3. Transferability of QTL

For some forage crops such as alfalfa, commercial varieties are mainly tetraploid although diploids also exist. In this case, QTL mapping in diploids is much more straightforward. However, QTL detected at diploid level might not be functional at the autopolyploid level. A well known example for differences in gene action at diploid and tetraploid level is the presence of gametophytic self-incompatibility in diploid potatoes, whereas autotetraploid potatoes generally are self fertile (Becker 1993). Furthermore, autopolyploids generally have enlarged cells and vegetative organs as compared to diploid forms (Becker 1993). This implies that tetraploid performance can only partly be predicted based on "diploid information". Similarly, prediction of genotype or family performance to be grown in swards based on information obtained at the single-spaced plant level might be poor. Posselt (1984) reported a generally lower heterosis for agronomic traits for ryegrass in swards as compared to spaced plants. Furthermore, low correlations were found in ryegrass for seed yield components evaluated in plots versus single plants (Elgersma 1990). Thus, depending on the trait of interest, the mode of testing genotype or family performance is essential with regard to the transferability of information for breeding of superior varieties under practical conditions. Another topic relates to trait evaluation in monoculture or mixture with other species. The latter might be closer to agronomic practice but complicate QTL mapping. For all above mentioned questions it is important to know the correlations between experimental and agonomic performance.

Genome regions increasing GCA within a given synthetic are of highest priority for synthetic breeding. Hence, evaluation of testcross rather than per se performance (after cloning of mapped genotypes) will be preferable. In hybrid breeding, per se performance of inbreds is of minor interest compared to that of hybrid performance. In an experiment on mapping of QTL for forage traits in maize, four segregating populations were established within the flint heterotic pool and evaluated for forage traits after testcrossing to elite dent tester inbreds at the hybrid level (Lübberstedt et al. 1997a, b, 1998). The predictive value of QTL was evaluated by comparing QTL results across testers within one population or across populations using the same tester. The three small validation populations had zero, one, or both parent lines in common with the large calibration population. Generally, the number of common QTL across populations increased with the genetic similarity of mapping populations. Almost all QTL detected in the small independent sample were also detected in the calibration population, both derived from the same cross, or could be explained by linked QTL with reversed gene effects. Hence, from a breeders perspective small populations might be sufficient for detecting the most relevant QTL.

For unrelated mapping populations, about 70% of the detected QTL were specific to each population. However, consistency of QTL across populations as well as testers was highly trait-dependent. In conclusion, QTL or genes identified in

different populations or test systems (like plots versus single spaced plants) need re-evaluation in breeding populations under practical relevant conditions. No information is available so far about the transferability of forage grass or legume QTL from spaced plant to plot performance, or the transferability between populations.

4. FUNCTIONAL MARKERS

A concept for definition, development, application, and prospects of functional markers (FM) in plants has recently been published (Andersen and Lübberstedt, 2003). In contrast to anonymous genetic markers such as morphological, isozyme, or random DNA markers (microsatellites, AFLPs, RFLPs, etc), FM are derived from polymorphic sites within genes causally involved in phenotypic trait variation. The major drawback of anonymous genetic markers is that their predictive value depends on the known linkage phase between marker and target locus alleles (Lübberstedt et al. 1998). Thus, (quantitative) trait locus mapping is necessary for each cross de novo, as different subsets of QTL are polymorphic in individual populations, and linkage phases between marker and QTL alleles can disagree even in closely related genotypes. In contrast, once genetic effects have been assigned to functional sequence motifs, FM derived from such motifs can be used for fixation of gene alleles (defined by one or several FM alleles) in a number of genetic backgrounds without additional calibration. FM development requires (1) functionally characterized genes, (2) allele sequences from such genes, (3) identification of polymorphic, functional motifs affecting plant phenotype within these genes, and (4) validation of associations between DNA polymorphisms and trait variation. Finally, application of FM depends on the availability of robust marker assay technologies.

Accumulation of plant nucleotide sequences proceeds rapidly with almost 2.74 million entries deposited for maize at GenBank in April 2006 (ftp://ftp.ncbi.nih.gov/genbank/gbrel.txt). However, only a limited number of genes have so far been assigned to an "agronomic function" (Lübberstedt et al. 2005). Furthermore, while few projects are focused on comprehensive allele sequencing (http://www.maizegenetics.net/), the major effort has been spent on characterizing one or few genotypes per species. The purpose of the following chapter is to describe methods for isolation of "forage quality genes" while including the so far functionally characterized genes.

4.1. Gene Isolation

In case of forage quality, biochemical pathways and genes coding for enzymes of these pathways have been shown to contribute to forage quality (see 2.3). Therefore, these genes qualify for candidate genes to be validated for their role in respect to forage quality by hypothesis-driven "reverse genetics" approaches. Additional genes controlling forage quality can be identified by phenotype-based

"forward genetic" approaches. For both reverse and forward genetics, efficient high-throughput approaches have been developed in the past decade in the area of plant genomics, allowing for parallel evaluation of multiple genes.

4.1.1. Mutants

In maize, *bm* mutants have originally been identified based on their phenotypic effect on the coloration of midribs. Later it has been shown, that *bm* mutants have an increased digestibility, but inferior agronomic performance (Barriere and Argillier 1993). Two of the four *bm* genes have been isolated and shown to code for enzymes in the monolignol biosynthesis (Boerjan et al. 2003). More recently, *brittle stalk-2* (*bk2*) coding for a COBRA-like protein has been shown to mediate cellulose-deposition in secondary cell-walls of maize (Ching et al. 2005). *Bk2* affects both the strength and lignin content of maize stems and by this likely forage quality. Consequently, genes involved in cell-wall biosynthesis and especially those in the monolignol biosynthesis as well as in the polymerization of lignins can be considered as candidate genes with respect to forage quality (Barriere et al. 2003).

Comprehensive knock-out mutant collections mainly based on the Mutator transposon have been and are being established (http://www.mutransposon.org/project/) in maize both for forward and reverse genetic screening of traits and genes. For example, the NSF-funded Maize Gene Discovery Project uses *RescueMu* plasmid rescue to create immortalized collections of insertion sites in *E. coli* (Lunde et al. 2003). Over 70.000 RescueMu flanking sequences have been sequenced, while cataloging mutant seed and cob phenotypes of 23.000 maize ears, 6.200 families of maize seedlings, and 4.000 families of adult maize plants carrying MuDR/Mu and RescueMu insertion allele. To obtain seed, users can search the web site database for insertions into genes of interest and then perform PCR or hybridization on column libraries to ascertain which plant has the mutation (Lunde et al. 2003). Respective knock-out mutant collections have not been reported so far for forage grasses or forage legumes.

4.1.2. Transformation, RNAi, and virus induced gene silencing

Transformation for in vivo validation of gene function is established for the major forage crops (Spangenberg et al. 2001). Transgenic manipulation of lignin biosynthesis in tall fescue confirmed the central role of genes coding for CAD and COMT in controlling forage quality in grasses (Chen et al. 2003, 2004). Down-regulation of both genes led to improved forage quality. Since additional genes involved in monolignol formation and in polymerization of monolignols have already been isolated from forage grasses (Lidgett et al. 2005, Gavnholt and Larsen 2002), validation of these genes with respect to forage quality can be expected in the near future. Likewise, genes in fructan metabolism in ryegrass such as fructosyltransferases or invertases have been isolated (Chalmers et al. 2005). Transgenic modification of fructan metabolism in grasses has so far been restricted to introduction of alien (such as bacterial) genes (e.g., Ye et al. 2001).

In forage legumes, one focus is on generation of high levels of condensed tannins by use of genes involved in anthocyanidine biosynthesis (Robbins et al. 2005, O'Donoghue et al. 2005), using so far genes from the model legumes *Lotus ssp*. and *Medicago trunculata*. With respect to lignification, down-regulation of C4H and C3H showed even stronger effects on digestibility in alfalfa as compared to COMT and CAD (Reddy et al. 2005). The closest correlation to digestibility was found for lignin content, but not lignin composition.

Recently, RNAi (RNA interference) has emerged as method of choice to validate gene function in the context of plant development. The essence of RNAi is the delivery of double-stranded RNA (dsRNA) into an organism, or cell, to induce a sequence-specific RNA degradation mechanism that effectively silences a targeted gene (Waterhouse and Helliwell 2003). An important aspect of using RNAi in plant genomics is the delivery of the silencing-inducing dsRNA. This RNA can be delivered by stably transforming plants with transgenes that encode dsRNA. It can also be transiently delivered by bombarding plants with nucleic-acid-coated beads, by infiltrating plant cells with transgene-carrying *Agrobacterium tumefaciens* or by infecting plants with a virus, either on its own or together with a satellite virus (Waterhouse and Helliwell 2003). At least for genes involved in lignin biosynthesis, application of RNAi for generation of knock-out genotypes for gene function validation appears to be a very promising approach (Capell and Christou 2004). Transformation of maize and forage crops is not possible at high throughput. Alternatively, virus-induced gene silencing (VIGS) has been proposed for rapid *in vivo* gene function tests in maize based on maize streak virus, wheat streak mosaic virus, or barley stripe mosaic virus (Robertson 2004). In addition, VIGS has been established for legume species (Constantin et al. 2004). However, the usefulness of VIGS in relation to forage quality remains to be demonstrated.

4.1.3. Model species

Complete plant genomes have been sequenced for the model species *A. thaliana* and for rice (The *Arabidopsis* Genome Initiative 2000, Goff et al. 2002, Yu et al. 2002). Maize is turning into another plant model species, given the current activities towards completion of whole-genome sequencing and availability of multiple plant genomics related resources (http://www.maizegdb.org/). Within legumes, comprehensive functional genomics projects are underway for *Lotus corniculatus, L. japonica*, and *Medicago truncatula* (VandenBosch and Stacey 2003) as well as grain legumes such as soybean (Gepts et al. 2005).

Brachypodium distachyon has long been discussed as the ideal model species for forage grasses. Brachypodium is a temperate grass with many growth features in common and close phenotypic similarity to forage and turf grasses. Moreover, orthologous ryegrass ESTs show a substantially higher sequence homology to Brachypodium sequences as compared to rice, supporting a close evolutionary relationship between these two species (unpublished data). Brachypodium has many properties in common with the dicot model species *A. thaliana*, including a similar small genome size and 5 chromosomes per haploid genome,

a short life-cycle, and self-fertility (http://www.brachypodium.org/). Important resources (ESTs, BAC libraries, ecotype accessions) and methods (transformation) have already been established for Brachypodium (http://www.brachypodium.org/). Whereas *A. thaliana* has been accepted as model species (as indicated by it's broad use in research) and proven to be useful during the last decade, Brachypodium has not found broad acceptance until recently. However, this will likely change after whole genome sequencing of Brachypodium has been announced (http://www.jgi.doe.gov/sequencing/cspseqplans2007.html). The genome sequencing of Brachypodium is expected to be completed and released to the public domain in 2008 (http://www.jic.bbsrc.ac.uk/corporate/media-and-public/current-releases/060712.htm).

The term "function" in functional genomics relates to some basic characteristics of genes (e.g. mutant phenotype, biochemical properties, expression pattern of selected genotypes). The ambition for *A. thaliana* is to characterize the function of all genes of this species until 2010 (http://www.arabidopsis.org/info/2010_projects/). Genes of interest can, in principle, be identified for any forage crop by exploiting information based on sequence homology or conserved map position provided from model species. While the usefulness of model species in reverse genetics approaches due to synteny relationships among genomes of related species (see 5.4) is obvious, their implementation in forward genetics strategies to identify novel genes (outside the pathways described under 2.3) controlling forage quality ("agronomic function") especially of perennial crops might be limited.

Since genomic resources for major forage species have substantially increased over the past five years, they might serve as "model crops" for a broad range of related forage grasses or legumes. Meanwhile (5/2006) about 40.000 ESTs each have been released for *Festuca arundenacea* and *Trifolium pratense* (http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html). More than 10.000 ESTs have been published for the two Lolium species *L. multiflorum* and *L. temulentum*. Substantial numbers of ESTs have been generated for *L. perenne* (>70.000; Sawbridge et al. 2003, Asp et al. 2006), and *T. repens* (>30.000; Spangenberg et al. 2003), but not released so far. Furthermore, a comprehensive collection of "gene thresher" genomic sequences has been produced for *L. perenne* (http://www.vialactia.co.nz/news/newsitem.asp?id=61), and partly built into GRAMENE (http://www.gramene.org/). International consortia have been formed addressing the development of genomic tools including physical maps and large scale sequencing in Lolium and Festuca species (e.g., European Lolium and Festuca Initiative (ELFIN); http://www.iger.bbsrc.ac.uk/elfin/) as well as Trifolium species (International Trifolium Network (ITN), inaugurated 7/2005 at IGER, Aberystwyth, UK).

4.1.4. Sequence homology and synteny

An important question in forage grasses and legumes is, whether the model species investigated are phylogenetically close enough to extract meaningful information by comparative genomics. This seems to be the case within the grass

family (Devos and Gale 2000, http://ukcrop.net/comparative.html), but also legumes (Choi et al. 2004). Genes identified in, e.g., a single grass species might be directly used in different forage grasses by gene transfer or for identification of orthologous genes to develop allele-specific markers. An example is the conserved organization of vernalization genes in cereals and forage grasses (Yan et al. 2003, 2004, Jensen et al. 2005a, Andersen et al. 2006). However, identification of orthologues is not always straightforward, as illustrated for the *bm3* gene coding for COMT, mapping on chromosome 4 in maize (Figure 1) and linkage group 7 in ryegrass (Cogan et al. 2005b). According to Alm et al. (2003) describing the synteny between several grass species including maize and ryegrass, no orthologous genome regions between maize chromosome 4 and ryegrass linkage group 7 were discovered at the macrosynteny level. This missing relationship might be detected at the microsynteny level, once more sequence information is available. Moreover, exploitation of synteny to identify relevant loci across species might be complicated in case of gene families, such as in the case of laccases and peroxidases involved in the polymerisation of monolignols (Ralph et al. 2004). However, several relevant forage grasses or legumes are closely related, increasing the chance for transferability of synteny-based information within these two groups of species. A promising system for information transfer between different forage grasses as well as rice are *L. perenne* / *Festuca pratensis* introgression families exploiting high levels of polymorphism between both crossable species with markers or sequences assigned to the rice genome sequence for "introgression landing", as demonstrated for the fine mapping and candidate gene identification for the *F. pratensis* stay green locus (King et al. 2005).

The original synteny studies have been based on RFLP mapping across species using cDNA probes (Devos and Gale 2000). Grass genomes were shown to be organized in a limited number of chromosome blocks. More recent availability of complete genome sequences in *Arabidopsis* and rice and sequenced BAC contigs in other species as well as comprehensive (mapped) EST collections led to re-evaluation of the synteny concept (Delseny 2004). These sequencing projects revealed numerous polyploidization events after speciation even in the small genome species *A. thaliana* (Blanc et al. 2003), but also in rice and maize (Salse et al. 2004). Furthermore, large genomes like the maize genome have been invaded by retrotransposons resulting in reshuffling of original ancestral genomes due to duplications, deletions, illegitimate recombination etc. (Delseny 2004). Since these events occurred after speciation of maize, substantial differences have been reported between larger allelic sequence stretches even within maize among different inbred lines (Fu and Dooner 2002, Brunner et al. 2005). An *in silico* alignment of genomic rice sequence with mapped maize EST sequences identified larger collinear chromosome regions between rice and maize in agreement with previous studies (Salse et al. 2004). However, fine-scale analysis revealed besides several duplicated regions numerous internal rearrangements within syntenic chromosome blocks. In conclusion, rice can be used to identify candidate genes in a target region identified in maize, but the order and number of genes might be altered at the microsynteny

level between rice and maize. The high flexibility regarding chromosome mutations (Brunner et al. 2005) might be specific for the maize genome.

4.1.5. Map-based gene isolation

In maize, map-based gene isolation of QTL controlling forage quality will become much more efficient with the availability of the whole genome sequence in a few years. All relevant additional resources needed for map-based gene isolation such as mapping populations, BAC libraries, markers at high density are already publicly available (http://www.maizegdb.org/).

In forage grasses and legumes, BAC libraries have been reported for *L. perenne*, *F. pratensis*, and *T. repens* (Farrar et al. 2005, Spangenberg et al. 2003, Donnison et al. 2005). Combined with the increasing availability of molecular markers in these species (see 3.2; Jensen et al. 2005b, Lübberstedt 2005), map-based gene isolation has generally become possible. In addition, large numbers of SSRs have recently been published for timothy (>300; Cai et al. 2003) and red clover (>1200; Sato et al. 2005).

4.1.6. Expression profiling

Microarray-based expression profiling enables the simultaneous study of several and ultimately all genes of an organism in one experiment. In maize, publicly available microarrays contain currently PCR fragments from more than 10.000 different ESTs (*http://www.maizegdb.org/microarray.php*), about 18.000 probe sets covering 15.000 different genes (*http://www.affymetrix.com/products/arrays/specific/maize.affx*), whereas long oligo microarrays include 58.000 different oligonucleotides (*http://www.maizearray.org/*). Since current maize microarrays do not cover all maize genes, complementary approaches can be employed to identify, e.g., rare transcripts, such as suppression subtractive hybridization (SSH) (Diatchenko et al. 1996). The SSH procedure enriches cDNA libraries for low-abundant and differentially expressed mRNAs by normalization (Diatchenko et al., 1996). Shi et al. (2005) used SSH combined with macroarray hybridization to identify differently expressed genes in near isogenic lines for virus resistance genes in maize.

Expression profiling was employed to identify novel candidate genes involved in forage quality using maize unigene (cDNA) microarrays as well as macroarrays containing SSH derived EST fragments differentially expressed in a *Bm3*/*bm3* isogenic line contrast (Shi et al. 2006, Lübberstedt et al. 2006). About 10% of the sequences on the microarray were mapped *in silico* allowing an assignment of respective differentially expressed genes to maize chromosomes. “Stems” (more precisely: leaf sheaths) of 5–7 week old maize plants were compared in isogenic comparisons in three genetic backgrounds (KWa, KWb: inbreds from KWS Saat AG; public line F2) with respect to *bm1*, *bm2*, and *bm3*. Moreover, extremes of three QTL mapping populations differing in cell-wall digestibility were bulked and compared with respect to differential expression (Lübberstedt et al. 2006; Shi et al., 2007). Summarized over the different comparisons, several 1000 genes showed

significant differential expression in individual experiments. Moreover, most genes involved in monolignol biosynthesis showed differential expression in at least one of the *Bm*/*bm* isogenic comparisons. When comparing *Bm3*/*bm3* isogenic line contrasts in KWa, KWb, and F2, 53 genes were consistently differentially expressed. Although the *bm3* gene is located at chromosome 4, mapped differentially expressed genes in isogenic *bm3* contrasts were distributed across all maize chromosomes. Thus, isogenic contrasts for a well characterized candidate gene or genome region might reveal additional genes of interest based on differential expression involved in the same metabolic pathways or with regulatory function. For further investigations, differentially expressed genes cosegregating with forage quality QTL have high priority. In the study of *Bm*/*bm* isogenic contrasts (Shi et al. 2006), at least one differentially expressed gene was identified per chromosomal bin region (Gardiner et al. 1993) for each of the forage quality QTL described by Ralph et al. (2004).

For ryegrass and white clover, the International Transcriptome Initiative for Forage and Turf (ITIFT) for forage plant gene discovery based on expression profiling has been announced recently (Webster et al. 2005). Currently 15k cDNA unigene microarrays are provided for each ryegrass and white clover. Most likely, such microarrays will be useful for other closely related forage species such as Festuca ssp. Moreover, Affymetrix microarrays are meanwhile available for rice, wheat, and *M. trunculata* (http://www.affymetrix.com/community/research/consortia.affx), which should be useful in heterologous hybridization studies, as demonstrated, e.g., for alfalfa (Aziz et al. 2005, May 2005).

4.2. Allelic Diversity

Once genes affecting forage quality have been identified, the next step towards development of functional markers is the identification of polymorphisms within these genes causally affecting the target trait.

4.2.1. Association studies in maize

One major limitation in "classical" QTL studies is the low number of (typically 2–4) segregating alleles, adding to the QTL transferability problem. More recently, more complex population structures allowing simultaneous evaluation of multiple alleles at different levels of genetic resolution have been proposed (Thornsberry et al. 2001, Churchill et al. 2004). Assignment of an "agronomic function" to short sequence motifs can be achieved by candidate gene based association studies (Risch 2000). This approach is limited by linkage disequilibrium (LD), i.e. haplotype structures in the gene(s) of interest. However, for several genes a generally low LD was detected in maize (Remington et al. 2001; Flint-Garcia et al. 2003), including examples in elite materials (Zein et al. 2007). Thus, candidate gene-based association studies are promising in maize. In heterogeneous genotype collections associations identified for specific sites might be confounded with effects from other genome regions especially in case of population stratification (Pritchard et al. 2000), which needs to be taken into account for interpretation of results from association studies.

In a pioneering study, Thornsberry et al. (2001) demonstrated the feasibility of association studies in maize to identify sequence polymorphisms within genes affecting characters of agronomic significance. While taking population structure into account, nine SNP or INDEL polymorphisms were shown to significantly affect flowering time in a set of 92 diverse maize lines. In part, these results were confirmed in a collection of European elite inbred lines (Andersen et al. 2005). The major reason for non-significance of some of the nine polymorphisms identified very likely the much narrower genetic material investigated by Andersen et al. (2005) as compared to Thornsberry et al. (2001).

First reports on association studies for genes involved in cell wall biosynthesis confirm that these pathways are promising targets for identification of polymorphic sites associated with forage quality, and thus FM development (Barriere et al. 2003). Zein et al. (2006) investigated the sequence variation at the *Bm3* locus in a collection of 42 European maize inbred lines, contrasting with respect to stover DNDF and relevant for hybrid maize breeding in Central Europe. For association with forage quality, stover digestibility was determined in six environments between 2001 and 2003 in Germany (heritability >0.9). One INDEL polymorphism within the intron revealed significant association with stover digestibility (Lübberstedt et al. 2005). In a study of Guillet-Claude et al. (2004a), polymorphisms both in the AldOMT (= *Bm3*) and the CCoAOMT2 but not CCoAOMT1 coding genes showed significant association with maize digestibility. Moreover, polymorphisms in the maize peroxidase gene ZmPox3 were also significantly associated with maize digestibility (Guillet-Claude et al. 2004b). In conclusion, availability of qualified candidate genes can be effectively converted into informative molecular markers by means of association studies. In maize, comprehensive association studies are ongoing in the group of E. Buckler (http://www.maizegenetics.net/), were 18 genes are studied in a panel of 102 maize genotypes, and within Genoplante (e.g. Guillet-Claude et al. (2004a)).

4.2.2. Association studies in forage grasses and legumes

Association studies based on candidate genes are especially promising in species with a generally low linkage disequilibrium (Flint-Garcia et al. 2003), as can be expected for outcrossing forage crops. Studies on systematic allele-sequencing and association studies in ryegrass are currently ongoing in the EU project GRASP (http://www.grasp-euv.dk, Xing et al. 2006), including genes involved in lignin biosynthesis (Lübberstedt et al. 2003). First results from a similar project aiming at development of gene-derived SNP markers in ryegrass, including genes from lignin and fructan biosynthesis, have been presented by Pointing et al. (2005) and Forster et al. (2005). SNP development and association studies have been initiated in white clover as well (Forster et al. 2006).

4.2.3. TILLING

TILLING (McCallum et al. 2000) can be employed to relate sequence polymorphisms with phenotypic variation. Variants for virtually all genes of interest in a

fixed genetic background can be produced by TILLING (McCallum et al. 2000). The advantage of TILLING as compared to association studies is that isogenic lines are compared avoiding statistical artefacts due to population structure effects. The disadvantage of TILLING is that establishing a comprehensive TILLING population covering most genes is quite laborious. Therefore, TILLING populations are usually restricted to one or few genetic backgrounds and the alleles fixed within the respective "background genotypes". Thus, if a knock-out allele is fixed at a locus of interest, it might be not possible to identify revertants. Two TILLING populations for maize have been produced at Purdue (http://genome.purdue.edu/maizetilling/) in B73 and W22 background, available for the maize research community. Within the next few years, 150 maize genes will undergo systematic studies using this resource (http://genome.purdue.edu/maizetilling/). The development of TILLING populations in forage grasses and legumes is complicated by the lack of homozygous diploid inbred lines in most species, since both phenotypic and heteroduplex analysis in segregating offspring of heterozygous genotypes is impaired. Kardailsky et al. (2005) reported the first TILLING population in ryegrass. In the longer run, establishment of homologous recombination as established in moss (*Physcomitrella patens*) (Schaefer and Zryd 1997) would be desirable to generate isogenic genotypes with defined polymorphic differences and, if possible, in any genetic background.

5. IMPLEMENTATION OF MARKERS IN FORAGE CROP BREEDING

Independent of the breeding category, cultivar breeding includes three phases: I) generation of genetic variation, II) development of genetic components for producing new varieties (such as inbred lines in hybrid breeding), and III) testing of experimental varieties (Becker 1993). Molecular breeding benefits all three phases, but is also useful in the context of variety registration and protection as well as for the characterization and management of genetic resources. The predictive value of markers depends on whether they are random DNA markers, gene-derived or functional markers (Andersen and Lübberstedt 2003).

Markers are useful to establish heterotic groups (Riday and Brummer 2003), and to assign genotypes or populations to heterotic groups. This topic might become increasingly relevant, if forage crop breeding moves from population or synthetic breeding to hybrid breeding (Riday and Brummer 2003). Furthermore, markers might assist identification of genetically divergent parent genotypes or populations with a maximum usefulness (Lamkey et al. 1995, Schnell 1983) to generate better varieties. Finally, recurrent selection programs might benefit from the application of markers ensuring, e.g., a sufficient level of genetic variation over several selection cycles.

Marker-assisted selection (MAS) and backcrossing (MAB) are major applications of molecular markers. MAS is especially promising for traits with low heritability (Lande and Thompson 1990), whereas MAB allows tracing of favourable alleles, which is especially useful in case of recessive gene action. In case of MAB,

including transfer of trangenes across genotypes and populations, markers are useful for background selection (Frisch et al. 2001) to recover the elite parent background efficiently in a short time. For MAS, an increasing number of candidate gene-derived or even functional markers (Andersen and Lübberstedt 2003) will become available in the near future, as demonstrated for ryegrass recently (Faville et al. 2004). Functional markers will reduce the risk of a Type 3 error in declaring QTL-marker associations (Dudley 1993), i.e., declaring in case of a significant QTL the wrong marker allele as being linked to a favourable QTL allele. In addition, markers can be employed to predict the performance of components of hybrid or synthetic varieties using BLUP (Bernardo 2002).

Duing the final steps of variety production markers can be used to (i) reduce the amount of experimental testing, (ii) confirm hybridity, and (iii) fulfill DUS criteria in variety registration (Tommasini et al. 2003). Finally, gene bank collections might benefit from molecular markers to describe and maintain genetic resources, as well as to establish core collections.

FM would be superior as compared to anonymous markers for selection of, e.g. parent materials to build segregating populations, as well as subsequent development of inbred lines. FM would also be useful for variety registration based on presence/absence of specific alleles at morphological trait loci currently used to discriminate varieties. Rapid progress in sequencing projects, especially for maize, and genomics will shift the current bottleneck for FM development from the availability of candidate genes and the availability of allele sequence information to the assignment of "agronomic function" with sequence polymorphisms, which is currently not systematically considered in functional genomics projects. First reports on association studies in maize are promising. The number of sequence polymorphisms useful for FM development can be expected to increase substantially within the next 5–10 years with the availability of further association studies and TILLING experiments, initially in maize but subsequently for other forage grasses and legumes.

REFERENCES

Alm V, Fang C, Busso CS, Devos KM, Vollan K, Grieg Z, Rognli OA (2003) A linkage map of meadow fescue and comparative mapping with other Poaceae species. Theor Appl Genet 108:25–40

Andersen JR, Lübberstedt T (2003) Functional markers in plants. Trends Plant Sci 8:554–560

Andersen JR, Schrag T, Zein I, Melchinger AE, Lübberstedt T (2005) Functional marker validation of polymorphisms in the maize *dwarf8* gene affecting flowering time in European elite materials. Theor Appl Genet 111:206–217

Andersen JR, Jensen LB, Asp T, Lübberstedt T (2006) Vernalization response in perennial ryegrass (*Lolium perenne* L) involves orthologues of diploid wheat (*Triticum monococcum*) *VRN1* and rice (*Oryza sativa*) *Hd1*. Plant Mol Biol 60:481–494

Armstead IP, Turner LB, Farrell M, Skot L, Gomez P, Montoya T, Donnison IS, King IP, Humphreys MO (2004) Synteny between a major heading-date QTL in perennial ryegrass (*Lolium perenne* L.) and the Hd3 heading-date locus in rice. Theor Appl Genet 108:822–828

Asp T, Didion T, Nielsen KK, Holm PB, Lübberstedt T (2006) Integrated resources for functional genomics in perennial ryegrass (*Lolium perenne*): EST collection, mapping populations, BAC library,

microarray-based expression profiling, and transformation systems. Posterabstract 8th International Congress of Plant Molecular Biology, Adelaide, August 20–25 2006

Aziz N, Paiva NL, May GD, Dixon RA (2005) Transcriptome analysis of alfalfa glandular trichomes. Planta 221:28–38

Barrière Y, Argillier O (1993) Brown-midrib genes of maize: a review. Agronomie 13:865–876

Barrière Y, Guillet C, Goffner D, Pichon M (2003) Genetic variation and breeding strategies for improved cell wall digestibility in annual forage crops. A review. Animal Res 52:193–228

Barrière Y, Ralph J, Méchin V, Guillaumie S, Grabber JH, Argillier O, Chabbert B, Lapierre C (2004) Genetic and molecular basis of grass cell wall biosynthesis and degradability. II. Lessons from brown-midrib mutants. Comptes Rendus Biol 327:847–860

Becker H (1993) Pflanzenzüchtung. Ulmer Verlag, Stuttgart

Bernardo R (2002) Breeding for quantitative traits in plants. Stemma press, Woodburry, MN

Bevan M, Mayer K, White O, Eisen JA, Preuss D, Bureau T, Salzberg T, Mewes H-W (2001) Sequence and analysis of the *Arabidopsis* genome. Curr Opin Plant Biol 4:105–110

Blanc G, Hokamp K, Wolfe KH (2003) A recent polyploidy superimposed on older large-scale duplications in the *Arabidopsis* genome. Genome Res 13:137–144

Boberfeld WOV (1986) Grünlandlehre. Ulmer Verlag, Stuttgart

Boerjan W, Ralph J, Baucher M (2003) Lignin biosynthesis. Annu Rev Plant Biol 54:519–546

Boudet AM, Grima-Pettenati J (1996) Lignin genetic engineering. Mol Breed 2:25–39

Boudet AM, Lapierre C, Grima-Pettenati J (1995) Biochemistry and molecular biology of lignification. New Phytol 129:203–236

Brummer EC, Robins JG, Alarcon Zuniga B, Luth D (2005) Genetic mapping in tetraploid alfalfa: results and prospects. In: Humphreys MO (ed) Molecular breeding for the genetic improvement of forage crops and turf. Wageningen Academic Publishers, Wageningen, p 149

Brunner S, Fengler K, Morgante M, Tingey S, Rafalski A (2005) Evolution of DNA sequence nonhomologies among maize inbreds. Plant Cell 17:343–360

Cai HW, Yuyama N, Tamaki H, Yoshizawa A (2003) Isolation and characterization of simple sequence repeat markers in the hexaploid forage grass timothy. Theor Appl Genet 107:1337–1349

Capell T, Christou P (2004) Progress in plant metabolic engineering. Curr Opin Biotech 15:148–154

Cardinal AJ, Lee M, Moore KJ (2003) Genetic mapping and comparative analysis of quantitative trait loci affecting fiber and and lignin content in maize. Theor Appl Genet 106:866–874

Chalmers J, Johnson X, Lidgett A, Spangenberg GC (2003) Isolation and characterization of a sucrose: sucrose 1-fructosyltransferase gene from perennial ryegrass. J Plant Physiol 160:1385–1391

Chalmers J, Lidgett A, Cummings N, Cao YY, Forster J, Spangenberg GC (2005) Molecular genetics of fructan metabolism in perennial ryegrass. Plant Biotech J 3:459–474

Chen L, Auth CK, Dowling P, Bell J, Wang ZY (2003) Improving forage quality of tall fescue by genetic manipulation of lignin bioynthesis. In: Hopkins A, Wang Z-Y, Mian R, Sledge M, Barker RE (eds) Molecular breeding of forage and turf. Developments in plant breeding, Kluwer Academic Publishing, Dordrecht, pp 181–188

Chen L, Auth C, Dowling P, Bell J, Lehmann D, Wang ZY (2004) Transgenic down-regulation of caffeic O-methyltransferase (COMT) led to improved digestibility in tall fescue. Funct Plant Biol 31:235–245

Ching A, Dhugga KS, Appenzeller L, Multani DS, Bourett T, Meeley R, Rafalski AJ (2005) Maize brittle stalk 2 (*bk2*) gene determines mechanical strength of tissue by mediating cellulose deposition in secondary cell walls. Maize Genet Conf Abstr 47:P30

Choi H-K, Mun J-H, Kim D-J, Zhu H, Baek J-M, Mudge J, Roe B, Ellis N, Doyle J, Kiss GB, Young ND, Cook DR (2004) Estimating genome conservation between crop and model legume species. Proc Natl Acad Sci USA 101:15289–15294

Churchill et al (2004) The collaborative cross, a community resource for the genetic analysis of complex traits. Nat Genet 36:1133–1136

Cogan NOI, Smith KF, Yamada T, Francki MG, Vecchies AC, Jones ES, Spangenberg GC, Forster JW (2005a) QTL analysis and comparative genomics of herbage quality traits in perennial ryegrass. Theor Appl Genet 110:364–380

Cogan NOI, Vecchies AC, Yamada T, Smith KF, Forster JW (2005b) QTL analysis of mineral content in perennial ryegrass. In: Humphreys MO (ed) Molecular breeding for the genetic improvement of forage crops and turf, Wageningen Academic Publishers, Wageningen, p 153

Collazo P, Montoliu L, Puigdomenech P, Rigau J (1992) Structure and expression of the lignin O-methyltransferase gene from *Zea mays* L. Plant Mol Biol 20:857–867

Constantin GD, Krath BN, MacFarlane SA, Nicolaisen M, Johansen IE, Lund OS (2004) Virus-induced gene silencing as a tool for functional genomics in a legume species. Plant J 40:622–632

Davin LB, Bedgar DL, Katayama T, Lewis NG (1992) On the stereoselective synthesis of (+)-pinoresinol in *Forsythia suspensa* from its achiral precursor, coniferyl alcohol. Phytochemistry 31:3869–3874

Delseny M (2004) Re-evaluating the relevance of ancestral shared synteny as a tool for crop improvement. Curr Opin Plant Biol 7:126–131

Devos KM, Gale MD (2000) Genome relationships: the grass model in current research. Plant Cell 12:637–646

Dhillon BS, Paul C, Zimmer E, Gurrath PA, Klein D, Pollmer WG (1990) Variation and covariation in stover digestibility traits in diallel crosses of maize. Crop Sci 30:931–936

Diatchenko L, Lau YF, Campbell AP, Chenchik A, Moqadam F, Huang B, Lukyanov S, Lukyanov K, Gurskaya N, Sverdlov ED, Siebert PD (1996) Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. Proc Natl Acad Sci USA 93:6025–6030

Donnison IS, O'Sullivan DM, Thomas A, Canter P, Moore B, Armstead I, Thomas H, Edwards KJ, King IP (2005) Construction of a *Festuca pratensis* BAC library for map-based cloning in *Festulolium* substitution lines. Theor Appl Genet 110:846–851

Dudley JW (1993) Molecular markers in plant improvement: manipulation of genes affecting quantitative traits. Crop Sci 33:660–668

Elgersma A (1990) Seed yield related to crop development and to yield components in nine cultivars of perennial ryegrass. Euphytica 49:141–151

Farrar K, Thomas I, Humphreys MO, Donnison I (2005) Construction of a BAC library for *Lolium perenne*. In: Plant animal genome conference XIII abstracts (http://www.intl-pag.org/pag/13/abstracts/PAG13_P052.html)

Faville MJ, Vecchies AC, Schreiber M, Drayton MC, Hughes LJ, Jones ES, Guthridge KM, Smith KF, Sawbridge T, Spangenberg GC, Bryan GT, Forster JW (2004) Functionally associated molecular genetic marker map construction in perennial ryegrass. Theor Appl Genet 110:12–32

Flint-Garcia M, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrium in plants. Annu Rev Plant Biol 54:357–74

Fontaine AS, Barrière Y (2003) Caffeic acid O-methyltransferase allelic polymorphism characterization and analysis in different maize inbred lines. Mol Breed 11:69–75

Forster JW, Cogan NOI, Vecchies AC, Pointing RC, Drayton MC, George J, Dumbsday JL, Spangenberg GC, Smith KF (2005) Gene-associated single nucleotide polymorphism (SNP) discovery in perennial ryegrass. In: Humphreys MO (ed) Molecular breeding for the genetic improvement of forage crops and turf. Wageningen Academic Publishers, Wageningen, p 199

Forster JW et al (2006) Development of functionally-associated genetic markers for genetic improvement of white clover. In: Plant and animal conference XIV, San Diego CA (http://www.intl-pag.org/14/abstracts/PAG14_W37.html)

Freudenberg K, Harkin JM, Rechert M, Fukuzumi T (1958) Die an der Verholzung beteiligten Enzyme. Die Dehydrierung des Subaoinalkohols. Chem Ber 91:581–590

Frisch M, Melchinger AE (2001) Marker-assisted backcrossing for simultaneous introgression of two genes. Crop Sci 41:1716–1724

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

Gallagher JA, Cairns AJ, Pollock CJ (2004) Cloning and characterization of a putative fructosyltransferase and two putative invertase genes from the temperate grass *L. temulentum* L. J Exp Bot 55:557–569

Gardiner JM, Coe EH, Melia-Hancock S, Hoisington DA, Chao S (1993) Development of a core RFLP map in maize using an immortalized F2 population. Genetics 134:917–930

Gavnholt B, Larsen K (2002) Molecular biology of plant laccases in relation to lignin formation. Physiol Plantarum 116:273–280

Geiger HH, Seitz G, Melchinger AE, Schmidt GA (1986) Genotypic correlations in forage maize. I. Relationships among yield and quality traits in hybrids. Maydica 37:95–99

Gepts P, Beavis WD, Brummer EC, Shoemaker RC, Stalker T, Weeden NF, Young ND (2005) Legumes as a model plant family. Genomics for food and feed report of the cross-legume advances through genomics conference. Plant Physiol 137:1228–1235

Goff SA et al (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science 296:92–100

Groß F (1979) Nährstoffgehalt und Verdaulichkeit von Silomais 1.Mitt: Bewertung von Silomais. Das wirtschaftseigene Futter 25:215–225

Gruber MY, Ray H, Blahut-Beatty L (2001) Genetic manipulation of condensed tannin synthesis in forage crops. In: Spangenberg, G (ed) Molecular breeding of forage crops – developments in plant breeding, Kluwer Academic Publishing, Dordrecht

Guillet-Claude C, Birolleua-Touchard C, Manicacci D, Fourmann M, Barraud S, Carret V, Martinant J-P, Barrière Y (2004a) Genetic diversity associated with variation in silage corn digestibility for three O-methyltransferase genes involved in lignin biosynthesis. Theor Appl Genet 110:126–135

Guillet-Claude C, Birolleua-Touchard C, Manicacci D, Rogowsky PM, Rigau J, Murigneux A, Martinant J-P, Barrière Y (2004b) Nucleotide diversity of the ZmPox3 maize peroxidase gene: relationships between a MITE insertion in exon 2 and variation in forage maize digestibility. BMC Genet 5:19–28

Halpin C, Holt K, Chojecki J, Oliver D, Chabbert B (1998) Brown midrib maize (bm1) – a mutation affecting the cinnamyl alcohol dehydrogenase gene. Plant J 14:545–553

Herrmann D, Boller B, Studer B, Widmer F, Kölliker R (2006) QTL analysis of seed yield components in red clover (*Trifolium pratense* L.). Theor Appl Genet 112:536–545

Jensen LB, Andersen J, Frei U, Xing Y, Taylor C, Holm PB, Lübberstedt T (2005a) QTL mapping of vernalization response in perennial ryegrass reveals cosegregation with an orthologue of wheat VRN1. Theor Appl Genet 110:526–537

Jensen LB, Aarens P, Andersen CH, Holm PB, Ghesquiere M, Julier B, Lübberstedt T, Muylle H, Nielsen KK, de Riek J, Roldán-Ruiz I, Roulund N, Taylor C, Vosman B, Barre P (2005b) Development and mapping of a public erence set of SSR markers in *Lolium perenne* L. Mol Ecol Notes 5:551–557

Kardailsky I, Bryan G, Faville M, Forester N, Gagic M, Marshall M, Richardson K, Veit B (2005) Controlled flowering project for Lolium perenne at AgResearch. In: Plant and animal conference XIII, San Diego (http://www.intl-pag.org/13/abstracts/PAG13_W100.html)

King IP, King J, Armstead IP, Harper JA, Roberts IA, Thomas H, Ougham HJ, Jones RN, Thomas A, Moore BJ, Huang L, Donnison IS (2005) Introgression mapiing in grasses. In: Humphreys MO (ed) Molecular breeding for the genetic improvement of forage crops and turf. Wageningen Academic Publishers, Wageningen, pp 31–42

Kirchgessner M, Kellner RJ (1981) Schätzung des energetischen Futterwertes von Grün- un Rauhfutter durch die Cellulasemethode, Landw Forschung 34:276–281

Lamkey KR, Schnicker BJ, Melchinger AE (1995) Epistasis in elite maize hybrids and choice of generation for inbred line development. Crop Sci 35:1272–1281

Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 124:743–756

Lidgett A et al (2005) Gene discovery and molecular dissection of lignin biosynthesis in perennial ryegrass. In: Humphreys MO (ed) Molecular breeding for the genetic improvement of forage crops and turf. Wageningen Academic Publishers, Wageningen, p 180

Lübberstedt T (2005) Objectives and benefits of molecular breeding in forage species. In: Humphreys MO (ed) Molecular breeding for the genetic improvement of forage crops and turf. Wageningen Academic Publishers, Wageningen, pp 19–30

Lübberstedt T, Melchinger AE, Schön CC, Utz HF, Klein D (1997a) QTL mapping in testcrosses of European flint lines of maize: I. Comparison of different testers for forage yield traits. Crop Sci 37:921–931

Lübberstedt T, Melchinger AE, Klein D, Degenhardt H, Paul C (1997b) QTL mapping in testcrosses of European flint lines of maize: II. Comparison of different testers for forage quality traits. Crop Sci 37:1913–1922

Lübberstedt T, Melchinger AE, Fähr S, Klein D, Dally A, Westhoff P (1998) QTL mapping in testcrosses of Flint lines of maize: III. Comparison across populations for forage traits. Crop Sci 38:1278–1289

Lübberstedt T, Andreasen BS, Holm PB et al (2003) Development of ryegrass allele-specific (GRASP) markers for sustainable grassland improvement – a new EU Framework V project. Czech J Genetics and Plant Breeding 39:125–128

Lübberstedt T, Zein I, Andersen JR, Wenzel G, Krützfeldt B, Eder J, Ouzunova M, Shi C (2005) Development and application of functional markers in maize. Euphytica 146:101–108

Lübberstedt T, Shi C, Krützfeldt B, Eder J, Wenzel G, Zein I, Ouzunova M (2006) Forschungsstrategien zur Verbesserung der Futterqualität beim Mais. Vorträge für Pflanzenzüchtung 69:27–34

Lunde CF, Morrow DJ, Roy LM, Walbot V (2003) Progress in maize gene discovery: a project update. Funct Integr Genomics 3:25–32

May GD (2005) Translational genomics for alfalfa varietal improvement. In: Humphreys MO (ed) Molecular breeding for the genetic improvement of forage crops and turf. Wageningen Academic Publishers, Wageningen, pp 55–62

McCallum CM et al (2000) Targeting induced local lesions IN genomes (TILLING) for plant functional genomics. Plant Physiol 123:439–442

McNabb WC, Spencer D, Higgins TJ, Barry TN (1994) *In vitro* rates of rumen proteolysis of ribulose-1,5-bisphosphatase carboxylase (Rubisco) from Lucerne leaves, and of ovalbumin, vicilin and sunflower albumin 8 storage proteins. J Sci Food Agric 64:53–61

Mechin V, Argillier O, Hebert Y, Guingo E, Moreau L, Charcosset A, Barriere Y (2001) Genetic analysis and QTL mapping of cell-wall digestibility and lignification in silage maize. Crop Sci 41:690–697

Menke KH, Huss W (1987) Tierernährung und Futtermittelkunde, 3rd ed. Ulmer Verlag, Stuttgart

Menke KH, Steingaß H (1987) Schätzung des energetischen Futterwertes aus der in vitro mit Pansensaft bestimmten Gasbildung und der chemischen Analyse, II. Regressionsgleichungen. Übersichten zur Tierernährung 15:59–93

Nersissian AM, Shipp EL (2002) Blue copper-binding domains. Adv Prot Chemistry 60:271–340

Norris KH, Barnes RF, Moore JE, Shenk JS (1976) Predicting forage quality by infrared lectance spectroscopy. J Anim Sci 43:889–897

O'Donoghue MTO, Spillane C, Guiney E (2005) Minimising bloat through development of white clover with high levels of condensed tannins. In: Humphreys MO (ed) Molecular breeding for the genetic improvement of forage crops and turf,. Wageningen Academic Publishers, Wageningen, p 227

Parvis N, Boucaud J, Prud'homme MP (2001) Fructans and fructan-metabolizing enzymes in leaves of *Lolium perenne*. New Phytologist 150:97–110

Pointing RC, Drayton MC, Cogan NOI, Spangenberg GC, Smith KF, Forster JW (2005) SNP discovery and haplotype variation in full-length herbage quality genes of perennial ryegrass. In: Humphreys MO (ed) Molecular breeding for the genetic improvement of forage crops and turf. Wageningen Academic Publishers, Wageningen, p 196

Posselt U (1984) Hybridzüchtung bei Lolium perenne. Vorträge für Pflanzenzüchtung 5:87–100

Pritchard JK, Stephens M, Rosenberg NA, Donnelly P (2000) Association mapping in structured populations. Am J Hum Genet 67:170–181

Raes J, Rohde A, Christensen JH, van de Peer Y, Boerjan W (2003) Genome-wide characterisation of the lignification toolbox in *Arabidopsis*. Plant Physiol 133:1051–1071

Ralph J, Guillaumie S, Grabber JH, Lapierre C, Barrière Y (2004) Genetic and molecular basis of grass cell-wall degradability. III. Towards a forage grass ideotype. Biologies 327: 467–479

Reddy MSS, Chen F, Shadle G, Jackson L, Aljoe H, Dixon RA (2005) Targeted down-regulation of cytochrome P450 enzymes for forage quality improvement in alfalfa (*Medicago sativa* L.). Proc Natl Acad Sci USA 102:16573–16578

Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM, Buckler ES (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc Natl Acad Sci USA 98:11479–11484

Riday H, Brummer EC (2003) Dissection of heterosis in alfalfa hybrids. In: Hopkins A, Wang Z-Y, Mian R, Sledge M, Barker RE (eds) Molecular breeding of forage and turf – Developments in Plant Breeding, Kluwer Academic Publishing, Dordrecht, pp 181–188

Risch NJ (2000) Searching for genetic determinants in the new millennium. Nature 405:847–856

Robbins MP, Allison GG, Bettany AJE, Dalton SJ, Davies TE, Hauck B (2002) Biochemical and molecular basis of plant composition determining the degradability of forage for ruminant nutrition. In: Durand J-L, Emile J-C, Huyghe C, Lemaire G (eds) Grassland science in Europe – multifunctional grassland: quality forages, animal products and landscapes. Proceedings of the 19th General Meeting of the European Grassland Federation, La Rochelle, France, pp 37–43

Robbins MP, Allison G, Bryant D, Morris P (2005) Polyphenolic phenomena: transgenic analysis of some of the factors that regulate the cell-specific accumulation of condensed tannins in forage crops. In: Humphreys MO (ed) Molecular breeding for the genetic improvement of forage crops and turf. Wageningen Academic Publishers, Wageningen, p 226

Robertson D (2004) VIGS vectors for gene silencing: many targets, many tools. Annu Rev Plant Biol 55:495–519

Rogers GE (1990) Improvement of wool production through genetic engineering. Trends Biotechnol 8:6–11

Roussel V, Gibelin C, Fontaine AS, Barriere Y (2002) Genetic analysis in recombinant inbred lines of early dent forage maize. II. QTL mapping for cell-wall constituents and cell-wall digestibility from per se value and top cross experiments. Maydica 47:9–20

Salse J, Piegu B, Cooke R, Delseny M (2004) New in silico insight into the synteny between rice (*Oryza sativa* L.) and maize (*Zea mays* L.) highlights reshuffling and identifies new duplications in the rice genome. Plant J 38:396–409

Sargent TD (1987) Isolation of differentially expressed genes. Meth Enzymol 152:423–432

Sato S et al (2005) Comprehensive structural analysis of the genome of red clover. DNA Res 12:301–364

Sawbridge T, Ong E-K, Binnion C, Emmerling M, McInnes R, Meath K, Nguyen N, Nunan K, O'Neill M, O'Toole F, Rhodes C, Simmonds J, Tian P, Wearne K, Webster T, Winkworth A, Spangenberg GC (2003) Generation and analysis of expressed sequence tags in perennial ryegrass. Plant Sci 165:1089–1100

Schaefer DG, Zryd JP (1997) Efficient gene targeting in the moss *Physcomitrella patens*. Plant J 11:1195–1206

Schnell FW (1983) Probleme der Elternwahl – Ein Überblick. In: Arbeitstagung der Arbeitsgemeinschaft der Saatzuchtleiter in Gumpenstein, Austria. November 22–24 1983, Press by Bundesanstalt für alpenländische Landwirtchaft, Gunpenstein, Austria, pp 1–11

Shi C, Invardsen C, Thümmler F, Wenzel G, Melchinger AE, Lübberstedt T (2005) Identification of genes differentially expressed in association with SCMV resistance in maize by combining SSH and cDNA array techniques. Mol Gen Genomics 273:450–461

Shi C, Zein I, Ouzunova M, Wenzel G, Lübberstedt T (2006) Transcriptome analysis in maize brown-midrib isogenic lines. Plant Mol Biol 62:697–714

Shi C, Uzanowsra A, Ouzumova M, Wenzel G, Lübbestedr T (2007) Association between cell wall digestibility and candidate gene expression profiles by application of genetical genomics to a Flint × Flint maize recombinant imbred line population. BMC Genomics 8:22

Spangenberg GC, Kalla R, Lidgett A, Sawbridge T, Ong EK, John U (2001) Breeding forage plants in the genome era. In: Spangenberg G (ed) Molecular breeding of forage crops – developments in plant breeding, Kluwer Academic Publishing, Dordrecht

Spangenberg GC et al (2003) Integrated resources for pastoral functional genomics: EST collections, BAC libraries, VIGS systems and microarray-based expression profiling in perennial ryegrass (*Lolium perenne*), white clover (*Trifolium repens*) and *Neotyphodium* grass endophytes. Plant & animal genomes XI conference (http://www.intl-pag.org/11/abstracts/W21_W131_XI.html)

Tanner GJ, Francki KT, Abrahams S, Watson JM, Larkin PJ, Ashton AR (2003) Proanthocyanidin biosynthesis in plants. J Biol Chem 278:31647–31656

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

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

Tilley JMA, Terry RA (1963) A two-stage technique for the in vitro digestion of forage crops. J Br Grassl and Soc 18:104–111

Tommasini L, Batley J, Arnold GM, Cooke RJ, Donini P, Lee D, Law JR, Lowe C, Moule C, Trick M, Edwards KJ (2003) The development of multiplex simple sequence repeat (SSR) markers to complement distinctness, uniformity and stability testing of rape (*Brassica napus* L.) varieties. Theor Appl Genet 106:1091–1101

Turner LB, Cairns AJ, Armstead IP, Ashton J, Sk t L, Whittaker D, Humphreys MO (2006) Dissecting the regulation of fructan metabolism in perennial ryegrass with quantitative trait locus mapping. New Phytol 169:45–58

Van Soest PJ (1974) Composition and nutritive value of forages. In: Heath ME, Metcalfe DS, Barnes RF (eds) Forages, 3rd edn. Iowa State University Press, Ames, IA, pp 53–63

VandenBosch KA, Stacey G (2003) Summaries of legume genomics projects from around the globe. Community resources for crops and models. Plant Physiol 131:840–865

Vattikonda MR, Hunter RB (1983) Comparison of grain yield and whole-plant silage production of recommended corn hybrids. Can J Plant Sci 63:601–609

Wang Z-Y, Ye XD, Nagel J, Potrykus I, Spangenberg GC (2001) Expression of a sulphur-rich sunflower albumin gene in transgenic tall fescue plants. Plant Cell Rep 20:213–219

Wang Z-Y, Hopkins A, Lawrence R, Bell J, Scott M (2003) Field evaluation and risk assessment of trangenic tall fescue plants. In: Hopkins A, Wang Z-Y, Mian R, Sledge M, Barker RE (eds) Molecular breeding of forage and turf – developments in plant breeding. Kluwer Academic Publishing, Dordrecht, pp 367–379

Waterhouse PM, Helliwell CA (2003) Exploring plant genomes by RNA-induced gene silencing. Nat Rev Genet 4:29–38

Webster T, Nguyen N, Rhodes C, Felittie S, Chapman R, Edwards D, Spangenberg GC (2005) A proposal for an international transcriptome initiative for forage and turf: microarray tools for expression profiling in ryegrass, clover, and grass endophytes. In: Humphreys MO (ed) Molecular breeding for the genetic improvement of forage crops and turf. Wageningen Academic Publishers, Wageningen

Weißbach F (1993) Bewerten wir die Qualität des Maises richtig? Mais 21:162–165

Wilman D, Irianni RAM, Humphreys MO (2004) Stay-green compared with non-stay-green *Lolium perenne* in field swards with different cutting and nitrogen treatments. Ann Appl Biol 144: 95–101

Xing Y, Andreasen BS, Frei U, Lübberstedt T (2006) Development of ryegrass allele-specific markers (GRASP) for rust resistance in *Lolium perenne*. In: Plant and animal conference XIV, San Diego (http://www.intl-pag.org/14/abstracts/PAG14_W35.html)

Yamada T, Forster JW (2005) QTL analysis and trait dissection in ryegrass. In: Humphreys MO (ed) Molecular breeding for the genetic improvement of forage crops and turf. Wageningen Academic Publishers, Wageningen, pp 43–54

Yan L, Loukoianov A, Tranquilli G, Helguera M, Fahima T, Dubcovsky J (2003) Positional cloning of the wheat vernalisation gene *VRN1*. Proc Natl Acad Sci USA 100:6263–6268

Yan L, Loukoianov A, Blecht A, Tranquilli G, Ramakrishna W, SanMiguel P, Bennetzen JL, Echenique V, Dubcovsky J (2004) The wheat VRN2 gene is a flowering repressor down-regulated by vernalisation. Science 303:1640–1644

Ye XD, Wu XL, Zhao H, Frehner M, Nosberger J, Potrykus I, Spangenberg GC (2001) Altered fructan ccumulation in transgenic *Lolium multiflorum* plants expressing a *Bacillus subtilis sacB* gene. Plant Cell Rep 20:205–212

Yu J et al (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science 296:79–92

Zein I, Wenzel G, Andersen JR, Lübberstedt T (2007) Nucleotide sequence diversity at the Caffeic acid O-methyltransferase locus in 42 European elite maize inbred lines. Gen Res Crop Evol, 54:139–148

Zimmer E, Theune HH, Wermke M (1980) Estimation of nutritive value of silage maize by using chemical parameters and in vitro digestibility. In: Pollmer WG, Phipps RH (eds) Improvement of quality traits of maize for grain and silage use. Martinus Nijhoff Publishers, The Hague, Boston, London, pp 447–465

Zimmer E, Gurrath PA, Paul C, Dhillon BS, Pollmer WG, Klein D (1990) Near-infrared lectance spectroscopy analysis of digestibility traits of maize stover. Euphytica 48:73–84

Zscheischler J (1990) Handbuch Mais. DLG-Verlag, Frankfurt am Main, Germany

CHAPTER 13

MOLECULAR MAPPING, MARKER-ASSISTED SELECTION AND MAP-BASED CLONING IN TOMATO

MAJID R. FOOLAD*

Department of Horticulture, The Pennsylvania State University, University Park, PA 16802, USA

Abstract: Significant progress has been made in molecular marker research in tomato, *Lycopersicon esculentum* Mill., including generation of markers, development of maps, mapping of genes and QTLs, and fine-mapping, characterization and map-based cloning of genes and QTLs. Numerous types of molecular markers have been developed in tomato, including RFLPs, RAPDs, AFLPs, SSRs, CAPS, ESTs, COSs and SNPs. Several molecular maps of tomato have been developed based on different interspecific populations, including the saturated linkage map based on a *L. esculentum* × *L. pennellii* cross. Markers and maps have been utilized extensively to map genes and QTLs controlling agriculturally and biologically important traits and for marker-assisted improvement of many simple-inherited traits such as disease resistance. Marker information also has been used for fine mapping and map-based cloning of several major genes and QTLs. Comparatively, little progress has been made in improving complex traits via marker-assisted selection. However, rapid advances in developing more efficient and resolving markers and in refining QTL positions are expected to lead to a greater use of marker technology for crop improvement in tomato.

1. INTRODUCTION

1.1. Botanical Description

The cultivated tomato, *Lycopersicon esculentum* Mill., is a worldwide-grown vegetable crop and the focus of a large agricultural industry. Botanically, tomato is a fruit as it develops from an ovary, however, in most places it is consumed as a vegetable. It belongs to the nightshade family Solanaceae, which is a diverse family consisting of ~96 genera and over 2800 species (Knapp et al. 2004). Solanaceae is the most variable of all crop families in terms of agricultural utility, the third most economically-important after grasses and legumes, and the most valuable in

*Address for correspondence: mrf5@psu.edu

R.K. Varshney and R. Tuberosa (eds.), Genomics Assisted Crop Improvement: Vol. 2: Genomics Applications in Crops, 307–356.

terms of vegetable crops. The family contains many of the commonly cultivated plants, including the tuber-bearing potato (*Solanum tuberosum* L.), fruit-bearing vegetables such as tomato, tomatillo (*Physalis philadelphica* L.), pepper (*Capsicum annum* L.) and eggplant (*Solanum melongena* L.), ornamental plants such as petunia (*Petunia hybrida* hort.) and Nicotiana (*Nicotiana tobacum* L.), medicinal plants such as Jimsonweed (*Datura stramonium* L.), *Capsicum* and *Nicotiana*, plants with edible leaves (e.g. *Solanum aethiopicum* L., *Solanum macrocarpon* L.), and various nightshades weeds. The tomato genus *Lycopersicon* is one of the smallest genera in Solanaceae, though the centerpiece for genetic and molecular research in the family.

The cultivated tomato was originally named *Solanum lycopersicum* by Linnaeus (Linnaeus 1753). In 1754, Miller separated tomatoes and designated the genus *Lycopersicon* and the species *esculentum* (roughly means "edible") for the cultivated tomato (Miller 1754). Genus *Lycopersicon* was initially distinguished from *Solanum* by its distinct characteristics of anthers and leaves. However, phylogenetic relationships between *Solanum* and *Lycopersicon* have been the subject of a great debate for a long time, with many Solanaceae researchers recognizing *Lycopersicon* as a distinct genus while others suggesting its merge with *Solanum* (for a review see Spooner et al., 1993). More recently, a new taxonomic classification of tomato and re-adoption of *S. lycopersicum* for the cultivated tomato has been suggested based on molecular and morphological information (Knapp et al. 2004; Spooner et al. 1993). The other species of *Lycopersicon* also have been assigned or re-assigned to *Solanum*. In this chapter, however, due to the use and citation of numerous historical references and in order to be consistent with much of the literature, I follow the Miller's classification and use original *Lycopersicon* names.

There are 9 known species within *Lycopersicon*, including the cultivated type *L. esculentum* and its wild form *L. esculentum* var. *cerasiforme* (Dun.) Gray, and the 8 wild species *L. pimpinellifolium* (Jusl.) Mill., *L. cheesmanii* Riley, *L. chmielewskii* Rick, Kes., Fob. & Holle, *L. chilense* Dun., *L. parviflorum* Rick, Kes., Fob. & Holle, *L. peruvianum* (L.) Mill., *L. hirsutum* Humb. and Bonpl. and *L. pennellii* (Corr.) D'Arcy (Rick 1976a; Rick 1979b). Although all species produce hermaphrodite flowers, a complete range of mating systems is found, from the almost completely autogamous *L. cheesmanii* and *L. parviflorum* to obligately outcrossed, self-incompatible biotypes of *L. chilense*, *L. hirsutum*, *L. peruvianum,* and *L. pennellii* . Self-fertility with various degrees of facultative outcrossing is found in *L. chmielewskii*, *L. esculentum*, *L. pimpinellifolium* and the self-compatible biotypes of *L. hirsutum* and *L. pennellii* . The 9 species have been grouped into two intra-crossable, inter-incrossable groups (or complexes): the "*esculentum* complex", including *L. esculentum*, *L. esculentum* var. *cerasiforme*, *L. pimpinellifolium*, *L.cheesmanii* (including *L. cheesmanii* f. *minor*), *L. chmielewskii*, *L. parviflorum*, *L. hirsutum* (including f. *typicum* and f. *glabratum*) and *L. pennellii* , and the "*peruvianum* complex" including *L. peruvianum* and *L. chilense* (Rick 1979a). All species within the *esculentum* complex can be hybridized with the cultivated tomato relatively easily and most (except *L. hirsutum* f. *typicum* and some *L. pennellii*) are self-compatible. The two species within the *peruvianum* complex are extremely

diverse and represent a wealth of characteristics potentially valuable to tomato improvement. However, these species have been rather partial in their usefulness to cultivated forms due to various barriers present in sexual hybridization and gene transfer (Rick 1973). For example, they can be hybridized with members of the *esculentum* complex only by the application of techniques such as embryo rescue.

All tomato species are native to a small area in western S. America, between Ecuador and Chile (Rick 1976b). However, their natural habitat is highly variable, from very dry to very wet and from coastal to mountainous areas of more than 3300 m elevations (Warnock 1988). Fruit color varies depending on the species, from red and orange to yellow and green. Among the 9 species, only *L. esculentum* has become a domesticated crop (Rick 1978), which includes the common fresh-market and processing tomatoes, land races and primitive cultivars, as well as the wild cherry. Fruits of the cultivated species come in a wide range of shapes, sizes and colors. Some may be globe, round, flattened, oval, heart, or elongated and even square-shaped, and which may be red, pink, golden, yellow, striped, purple, green or white (Stevens and Rick 1986). The average weight for the garden tomato is 4 to 6 oz., but some varieties weigh up to 2 pounds. Both the wild cherry and currant tomatoes (*L. pimpinellifolium*) produce shiny, bright red berries whereas the Galapagos tomatoes (*L. cheesmanii*) produce dull purplish or yellow-orange berries (diameter $\approx$ 1 cm). The rest of the wild species produce small, inedible green fruit.

All tomato species are diploid (2N = 2X = 24) and are similar in chromosome number and structure. Tomato's 12 chromosomes are highly differentiated and can readily be distinguished at pachytene. Tomato is very sensitive to any chromosome imbalance, and no aneuploid with euchromatic deficiencies can be maintained (Rick and Gill 1973). It is one of the most genetically characterized higher plant species and an excellent model system for basic and applied research. This is due to many reasons, including ease of culture, short life cycle, high self-fertility and homozygosity, great reproductive potential, ease of use for controlled pollination and hybridization, availability of a wide array of mutants and genetic stocks (http://tgrc.ucdavis.edu/; http://www.sgn.cornell.edu/), diploid with a rather small genome (0.86 pg, 950 Mbp) (Arumuganathan and Earle 1991), and amenability to asexual propagation and protoplast, cell and tissue cultures and whole plant regeneration thereof (McCormick et al. 1986). Members of *Lycopersicon* are easily transformed, and transgenic tomatoes are routinely produced using *Agrobacterium tumefaciens* (McCormick et al. 1986). Recent availability of high molecular weight insert genomic libraries (YACs and BACs) has facilitated map-based gene cloning in tomato. Moreover, advances in tomato EST database and genome sequencing have added additional tools for further expansion of basic and applied research in tomato.

1.2. Domestication and Genetic Variation

Peru was earlier widely accepted as the center of tomato domestication, however, "the bulk of the historical, linguistic, archaeological and ethno-botanical evidence favors Mexico as the source of the cultivated tomatoes" (Rick 1976b) and the

most likely center of domestication. It is not known exactly when domestication of tomatoes occurred, however, by the time the Spanish conquered Mexico in 1523, they were already domesticated (Rick 1978). Accordingly, the Spanish explorers found tomatoes in Central America and took seed to Europe, from where it were disseminated to many other parts of the world (Rick 1976b). A comparison of hereditary enzyme variants revealed much greater similarity between the older European cultivars and the primitive cultivars and cherry tomatoes of Mexico and Central America than between the European cultivars and the primitive plants of the Andean regions (Rick 1976b, 1991). The first record of tomatoes in Europe is credited to descriptions published in 1554 by Italian herbalist Pier Andrea Mattioli, while the first recorded mention of it in N. America was in 1710 (Rick 1978). However, in the U.S., commercial production of tomato in a small scale began in 1847 at Easton, Pennsylvania, and it grew to become a major vegetable crop in the mid 20th century.

The cultivated tomato has a narrow germplasm base, largely because of several population bottlenecks that occurred during domestication and evolution of modern cultivars (Rick 1976b). Although higher levels of variability are found in primitive cultivars in the native regions of tomato, only about 5% of the total genetic variation within *Lycopersicon* can be found within *L. esculentum* (Miller and Tanksley 1990; Rick and Fobes 1975). The related wild species of tomato, however, are a rich source of desirable genes and have been utilized extensively to improve agricultural characteristics of tomato, in particular disease resistance. In addition, there are thousands of monogenic mutants (http://tgrc.ucdavis.edu/; http://www.sgn.cornell.edu/), which are useful for genetics and breeding studies in tomato. Recent advancements in molecular markers and marker-assisted selection (MAS) have made tomato breeding via introgression from related wild species and mutants more efficient, as discussed in this chapter.

1.3. Economic Importance

The tomato has become an important crop worldwide in less than one century. Worldwide, it is the 2nd most consumed vegetable per person (after potato) (FAOSTAT 2005) and unquestionably the most popular garden vegetable. In the U.S., it is the 3rd most economically important vegetable crop (with a total farm value of $2.062 B) after potato ($2.564 B) and lettuce ($2.064 B) (http://www.usda.gov/nass/pubs/agr05/agstats2005.pdf). In addition to tomatoes that are eaten directly as raw vegetable or added to other food items, a variety of processed products have gained significant acceptance. Although a tropical plant, tomato is grown in almost every corner of the planet. It is grown in greenhouses where summers are too cool for pollination or fruit set in outdoors. Major production countries in descending orders include China, U.S.A., Russia, Turkey, India and Italy (FAOSTAT 2005). Worldwide, in 2005 a total of 4,528,519 ha of tomato were harvested with a total production of 124,748,292 Mt (FAOSTAT 2005). In N. America, production occurs in the U.S., Canada and Mexico, comprising a total of

310,000 ha. In the U.S., total harvested area in 2004 was estimated to be 170,808 ha (505,60 ha fresh-market tomatoes valued $1.34 B and 120,248 ha processing tomatoes valued $0.72 B) (http://www.nass.usda.gov:8080/QuickStats/index2.jsp). California is by far the leading producer of processing tomatoes followed by Florida, which is also the leading state in producing fresh-market tomatoes (USDA 2005). Per capita consumption in the U.S. includes 31.7 kg of processing and 8.7 kg of fresh tomatoes (http://www.ars.usda.gov/). Although tomatoes do not rank high in nutritional value, by virtue of volume consumed, they contribute significantly to the dietary intake of vitamins A and C and essential mineral and nutrients. In the U.S. diet, for example, tomato ranks first among all fruits and vegetables as a source of vitamins and minerals (Rick 1980). Furthermore, tomatoes are the richest source of lycopene, a phytochemical that protects cells from oxidants that have been linked to cancer (Giovannucci 1999).

1.4. Past Breeding Achievements

Research to develop new cultivars of tomato with improved characteristics started more than 200 years ago in Europe. Effective breeding programs in the U.S., however, started a little over a century ago (Stevens and Rick 1986). Livingston is generally recognized as the first tomato breeder in the N. America, conducting breeding in 1870s (Tigchelaar 1986). Almost until the middle of the 20th century, tomato-breeding objectives included development of multipurpose cultivars to meet several needs, including fresh-market and processing industries. Since then, however, tomato breeding objectives have depended upon method of culture, that is, field or greenhouse-grown tomatoes and whether the product has to be used fresh or processed (Stevens and Rick 1986; Tigchelaar 1986). However, the universal goal of tomato breeding has been to increase fruit yield per unit area. Other essential characteristics common to both fresh-market and processing industries include disease resistance, broad adaptability, earliness in maturity, ability to set fruit at adverse temperatures, resistance to rain-induced cracking, tolerance to major ripe-fruit rots, adequate vine cover, fruit firmness and several other fruit quality characteristics. Specific traits needed in processing cultivars include compact, determinate plant habit suitable for machine harvest, ease of fruit separation from the vine (jointless character), and specific fruit quality characteristics such color, pH, total acidity, soluble solids, total solids, viscosity (consistency). In comparison, specific traits of interest in fresh-market cultivars include large, round fruit with adequate firmness and shelf-life, uniform fruit size, shape and color, appearance, freedom from external blemishes or abnormalities, texture, taste and flavor (Stevens and Rick 1986; Tigchelaar 1986). During the past several decades remarkable progress has been made in improving tomatoes for various important traits. For example, between 1920s and 1990s, fruit yield of processing cultivars in the U.S. increased from 10.1 to 72.4 tons/ha, a 7.2-fold increase (Warren 1998). A recent statistic by the USDA indicated processing tomato yield of ~102 tons/ha in the U.S. in 2004 (http://www.nass.usda.gov:8080/QuickStats/index2.jsp). Furthermore, most

recently developed cultivars possess up to 6 (in true-breeding lines) or 10 (in hybrids) disease-resistance attributes. These mainly include diseases for which major resistance genes have been identified, including fusarium wilt, verticillium wilt, root-knot nematode, alternaria stem canker, gray leaf spot, and some bacterial and viral diseases. All original characterizations and gene introgressions were through phenotypic selection (PS) and traditional breeding protocols. Even today much of the disease resistance breeding in tomato is through the use of the same protocols. However, there are major limitations when using traditional protocols for gene transfer, in particular transfer from wild to the cultivated species. Thus, tomato breeders have consistently sought more effective breeding approaches. During the past two decades, the use of molecular markers and MAS techniques has facilitated identification, mapping and transferring of many genes in tomato. The use of MAS is becoming a routine procedure in tomato improvement, in particular when breeding for disease resistance, as described in below.

1.5. New Breeding Approaches

Traditional protocols of plant genetics and breeding, which are based on PS and progeny testing, have been very effective in improving crop productivity and quality during the past several decades (Duvick 1996). These methods, however, are often time consuming and not without inherent difficulties. For example, the average length of a breeding project for a seed or vegetable crop, from hybridization and selecting the new genetic combinations to testing them in the field and introducing them in the market, is ~10–15 years. This lengthy process may not allow the time-sensitive needed increase in crop productivity in the future. Furthermore, for many desirable agronomic and horticultural characteristics, controlling genes may be found only in exotic genetic backgrounds, utilization of which often encounters difficulties. After interspecific hybridization, a major task becomes eliminating the great bulk of undesirable genes introduced from the wild donor. A series of backcrosses to the cultivated type alternated with concurrent inbreeding are required to select the desired combination of parental characteristics. During this process, however, some genes of interest from the wild donor may be lost or eliminated, limiting the breeding success. These and other problems associated with traditional breeding methods warrant the use of techniques that have higher resolutions.

An alternative approach to improving effectiveness of selection is to discover and use genetic markers that are associated with genes or quantitative trait loci (QTLs) that control the trait(s) of interest. Marker technology can facilitate determination of the number, chromosomal location, and individual and interactive effects of genes or QTLs affecting a trait. Following their identification and localization, useful genes or QTLs can be introgressed into desirable genetic backgrounds via MAS, or isolated via map-based cloning. MAS may not only speed up the process of gene transfer, but it also allow pyramiding of desirable genes and QTLs from different genetic backgrounds. This can be an effective approach to substantial crop improvement, more so than potentially feasible through PS. Furthermore, in *Lycopersicon*, where

most genetic variability is within the wild species, identification and transfer of genes from the wild into the cultivated type may be significantly facilitated by the use of marker technology. In below, the current status of markers and maps, gene and QTL mapping, marker-assisted breeding and map-based cloning in tomato is reviewed and discussed.

2. GENETIC MARKERS AND MAPS

2.1. Classical Markers and Maps

The potential value of genetic markers for use as indirect selection criteria has been known to breeders for more than 80 years, since morphological and physiological markers were used to tag and select for superior phenotypes. Although morphological markers are easy to monitor, their use in genetics and breeding studies is restricted due to factors such as limitation in number, expression of dominance or epistatic interactions, pleiotropic effects, and incomplete penetrance or expressivity. In tomato, there are over 1300 known morphological, physiological and disease resistance genes (Chetelat 2002). During 1970s and 1980s, the second-generation markers, namely isozymes, became popular and were used for genetic mapping in plant species. In tomato, 41 isozymic loci corresponding to 15 unique enzymatic reactions have been characterized (Tanksley 1993). While isozyme markers have several advantages, including the low cost and ease of use, environmental insensitivity, lack of pleiotropic effects and co-dominant inheritance, they are very limited in number and often not polymorphic among closely-related genotypes (Foolad et al. 1993). Although currently isozyme markers are not used widely, before the era of DNA markers they were considered highly useful in genetics and breeding studies.

The first classical linkage map of tomato, showing markers on all 12 linkage groups, was reported in 1968 and included a total of 153 morphological and physiological markers (Butler 1968). For the next several years, the map was expanded and by 1975 more than 258 morphological and physiological markers were assigned to tomato chromosomes (Rick 1975). At that time, tomato had one of the best linkage maps of any plant species. The classical map information in 1970s greatly facilitated mapping of isozyme loci. The first complete isozyme linkage map of tomato was published in 1980, which included 26 isozyme markers (Tanksley and Rick 1980). Currently, there are 36 known isozyme markers in tomato that have been mapped (Tanksley 1993). The latest published classical linkage map of tomato consisted of 285 morphological, physiological, isozyme and disease resistance genes mapped onto the 12 tomato chromosomes (Tanksley 1993; Chetelat 2002).

2.2. Molecular Markers and Maps

With the advent of DNA markers in 1980s and 1990s, many limitations associated with morphological and isozyme markers were overcome and genetic mapping

entered a new exciting and progressive era with the promise to significantly enhance the efficiency of plant genetics and breeding studies. During the past two decades, different types of DNA markers have been developed and evolved, including restriction fragment length polymorphisms (RFLPs) (Botstein et al. 1980), randomly amplified polymorphic DNAs (RAPDs) (Williams et al. 1990), amplified fragment length polymorphisms (AFLPs) (Vos et al. 1995), variable number of tandem repeats (VNTRs or minisatellites) (Jeffreys et al. 1985), simple sequence repeats (SSRs or microsatellites) (He et al. 2003), cleaved amplified polymorphic sequences (CAPS) (Konieczny and Ausubel 1993), sequence characterized amplified regions (SCARs) (Paran and Michelmore 1993), expressed sequence tags (ESTs) (Adams et al. 1991), conserved ortholog set (COS) markers (Fulton et al. 2002), and single-nucleotide polymorphisms (SNPs) and insertion-deletion (InDels) markers (Landegren et al. 1998). Among crop species, tomato is very rich in the number and type of molecular markers. Currently there are over 1000 RFLP markers, most of which mapped to the 12 tomato chromosomes, and more than 162,000 ESTs (http://www.tigr.org/tigr-scripts/tgi/T_index.cgi?species=tomato), of which only a small portion mapped to tomato chromosomes (http://www.sgn.cornell.edu/cgi-bin/search/markers/cos_list.pl). ESTs are derived from different cDNA libraries and their sequences are available on the web site of Solanaceae Genome Network (SGN; http://www.sgn.cornell.edu). In addition, several other markers, including SSRs (Villalta et al. 2005), CAPS (Ganal et al. 1998), RAPDs (Foolad et al. 1993) and AFLPs (Suliman-pollatschek et al. 2002) have been developed in tomato. For example, currently at least148 SSR markers and 77 CAPS have been mapped onto the high-density tomato genetic map (Frary et al. 2005) (http://www.sgn.cornell.edu/cgi-bin/mapviewer/mapTop.pl?map_id=9).

The development and use of PCR-based markers in tomato has recently increased as they are generally cheaper, faster and less labor intensive than RFLPs. However, a significant issue in marker development in tomato is the limited polymorphisms among closely related genotypes. Such limitation restricts the use of DNA markers in many plant-breeding programs which focus on exploitation of intraspecific genetic variations. To overcome this problem, recently significant efforts have been devoted to the discovery and development of high-resolution markers such as SNPs and InDels, which can detect polymorphisms among genotypes within the cultivated species of tomato (Labate and Baldo 2005; Suliman-pollatschek et al. 2002). Such markers can have great utilities in tomato genetic and breeding studies that use populations derived from intraspecific crosses or between the cultivated and closely-related species such as *L. pimpinellifolium*. The growing tomato databases of DNA sequence, in particular the tomato ESTs, is providing information that are being used to develop more resolving genetic markers. It is expected that the availability of such markers will be on the rise for the next few years.

The first molecular map of tomato was published in 1986 containing 18 isozyme and 94 DNA markers (Bernatzky and Tanksley 1986). The first high-density map of tomato, comprising of 1030 markers, was published in 1992 (Tanksley et al. 1992). This map, constructed based on 67 F_2 plants of a *L. esculentum* × *L. pennellii*

Table 1. Genetic linkage maps of tomato (*Lycopersicon spp.*) developed based on different intra- and interspecific crosses

Linkage map	Population type[a]	Population size	Number of markers	Type of markers[b]	References
L. esculentum × *L. esculentum* var. *cerasif.*					
1. Cervil × Levovil	F_7-RIL	153	377	RFLP, RAPD, AFLP	(Saliba-Colombani et al. 2000)
L. esculentum × *L. pimpinellifolium*					
• M82-1-7 × LA1589	BC_1	257	120	RFLP, RAPD, morphological	(Grandillo and Tanksley 1996a)
• NC84173 × LA722	BC_1	119	151	RFLP	(Chen and Foolad 1999)
• Giant Heirloom × LA1589	F_2	200	90	RFLP, CAPS	(Lippman and Tanksley 2001)
• E6203 × LA1589	BC_2F_6-BIL	196	127	RFLP	(Doganlar et al. 2002)
• NC84173 × LA722	F_{10} RIL	119	191	RFLP, RGA	Foolad et al. (unpubl.)
• NCEBR1 × PSLP125	F_2	172	256	RFLP, EST, RGA	Foolad et al. (unpubl.)
• NCEBR1 × PSLP125	F_8-RIL	172	255	RFLP, EST	Foolad et al. (unpubl.)
L. esculentum × *L. cheesmanii*					
1. UC204B × LA483	F_2	350	71	RFLP	(Paterson et al. 1991)
2. UC204B × LA483	F_7-RIL	97	132	RFLP	(Paran et al. 1995)
L. esculentum × *L. parviflorum*					
1. E6203 × LA2133	BC_2	170	133	RFLP, SCAR, morphological	(Fulton et al. 2000)
L. esculentum × *L. chmielewskii*					
1. UC82B × LA1028	BC_1	237	70	RFLP, Isozyme	(Paterson et al. 1988)
L. esculentum × *L. hirsutum*					
1. E6203 × LA1777	BC_1	149	135	RFLP	(Bernacchi and Tanksley 1997)

(continued)

Table 1. (continued)

Linkage map	Population type[a]	Population size	Number of markers	Type of markers[b]	Reference
2. E6203 × LA1777	NIL, BIL	111	95	RFLP	(Monforte and Tanksley 2000b)
3. NC84173 × PI126445	BC_1	145	171	RFLP, RGA	(Zhang et al. 2002)
L. esculentum × *L. pennellii*					
1. VF36 *Tm 2*^{a} × LA716 (high-density map of tomato)	F_2	67	1050	Isozyme, RFLP, morphological	(Pillen et al. 1996; Tanksley et al. 1992)
2. Vendor *Tm2*^{a} × LA716	F_2	432	98		(de Vicente and Tanksley 1993)
3. M82 × LA716	IL	50	375	RFLP	(Eshed and Zamir 1995)
4. VF36 *Tm2*^{a} × LA716	F_2	67	1242	AFLP, RFLP	(Haanstra et al. 1999)
5. E6203 (LA925) × LA716	F_2	83	1,500	COS	(Fulton et al. 2002)
6. E6203 × LA1657	BC_2	175	110	RFLP	(Frary et al. 2004b)
7. E6203 × LA716	F_2	83	152	SSRs, CAPs	(Frary et al. 2005)
L. esculentum × *L. peruvianum*					
1. E6203 × LA1706	BC_3	241	177	RFLP, SCAR	(Fulton et al. 1997)
L. esculentum var. *cerasif.* × *L. pimpinellifolium*					
1. E9 × L5	F_6-RIL	142	132	SSR, SCAR	(Villalta et al. 2005)
L. esculentum var. *cerasif.* × *L. cheesmanii*					
1. E9 × L3	F_6-RIL	115	114	SSR, SCAR	(Villalta et al. 2005)
L. peruvianum × *L. peruvianum*					
1. LA2157 × LA2172	BC_1	152	73	RFLP	(van Ooijen et al. 1994)

[a] BIL: backcross inbred line; NIL: near isogenic line; RIL: recombinant inbred line

[b] AFLP: amplified fragment length polymorphism; CAPS: cleaved amplified polymorphic sequence; EST: expressed sequence tag; RAPD: randomly amplified polymorphic DNA; RFLP: restriction fragment length polymorphism; RGA: resistance gene analog; SCAR: sequence characterized amplified region; SSR: simple sequence repeat

cross, also displayed the chromosomal locations of 100 genes of known function or phenotype, including morphological, isozyme and DNA markers. A more saturated version of this map was published in 1996, reducing the inter-marker space to $\leq$1 cM (Pillen et al. 1996). Furthermore, the density of markers in this map has since increased, currently including about 2000 markers (http://www.sgn.cornell.edu/cview/map.pl?map_id=9). The haploid DNA content of the tomato genome is ~950 Mbp (Arumuganathan and Earle 1991). This means that on average 1cM genetic map distance in tomato amounts to ~750 kb. With the high-density molecular map of tomato, it is likely that any gene of interest, if segregating in this population, would be within one to few map units of at least one marker. However, many agriculturally important traits are not segregating in this population or many of the markers in this map are not polymorphic in other populations of tomato. Thus, during the past two decades several other molecular maps of tomato have been constructed, mostly based on interspecific crosses between the cultivated and various wild species of tomato (Table 1). Most of these maps are of low to moderate density and developed based on RFLP markers from the high-density map, though some other markers, such as RAPDs, ESTs, AFLPs, SSRs and resistance gene analogs (RGAs), also have been utilized. As alluded to before, for interspecific crosses between the cultivated tomato and closely-related wild species *L. pimpinellifolium* and *L. cheesmannii* identification of polymorphic markers is very challenging. For example, only ~30% of the RFLP markers in the high-density map detect polymorphism in *L. esculentum* × *L. pimpinellifolium* crosses (Chen and Foolad 1999; Grandillo and Tanksley 1996a). In fact in a more recent study, < 15% of RFLP markers detected polymorphism between a Mexican accession of *L. pimpinellifolium* and *L. esculentum* (MR Foolad et al., unpubl.). Despite these and other challenges, however, to date linkage maps have been developed based on interspecific crosses between the cultivated and all wild species of tomato, except may be *L. chilense*.

To accelerate the use of markers and maps in tomato genetics and breeding studies, recently efforts have been made to develop maps based on mainly PCR-based markers. One such effort (Frary et al. 2005) has produced a map based on an F_2 population of a *L. esculentum* × *L. pennellii* cross using a set of 76 SSRs and 76 CAPS (Table 1). Another significant effort currently underway in a few laboratories around the world is conversion of tomato RFLP markers to friendly PCR-based markers such as CAPS. Furthermore, a simultaneous effort is development of genetic maps based on intraspecific populations within the cultivated tomato.

3. GENE TAGGING AND QTL MAPPING

Tomato has been at the forefront of many crop species in terms of gene tagging and QTL mapping for agriculturally and biologically important traits. For simply-inherited traits, gene mapping in tomato started in 1930s (MacArthur 1934), however, tagging of single-gene agricultural traits with genetic markers and their use in tomato breeding started in 1970s. Rick and Fobes (1974) reported an association of root-knot nematode (*Meloidogyne incognita*) resistance with a rare form

of isozyme acid phosphatase, *Aps-1*[1]. This association was later determined to be due to a tight linkage between nematode resistance gene *Mi* and the *Aps-1* locus on chromosome 6 (Medina-Filho 1980). Subsequently, linkages were reported between isozyme loci and genes controlling a few important traits in tomato, including male-sterility (Tanksley et al. 1984) and self-incompatibility (Tanksley and Loaiza-Figueroa 1985). Since then, marker tagging of many simple traits in tomato has been reported, as discussed in below.

The use of markers to identify and map genes or QTLs controlling complex traits in tomato started in 1980s. Earlier studies mainly used morphological and isozyme markers and early filial or backcross populations to identify QTLs for different quantitative traits. However, the first comprehensive and systematic use of markers to dissect genetic controls of complex traits in tomato was that of Paterson et al. (1988), in which a rather complete RFLP map was employed to map QTLs for various fruit quality characters. This study demonstrated, for the first time, that genetics of quantitative traits could be resolved into discrete Mendelian factors. Subsequently, QTL mapping became common in tomato genetics and breeding studies, where QTLs have been identified for numerous traits. It is difficult to provide a complete account of all genes and QTLs that have been identified and/or mapped in tomato. Rather a tabulated summary of major genes and QTLs that have been identified and mapped during the past two decades is presented (Tables 2–4) with a summary discussion of some major mapping studies.

3.1. Mapped Disease Resistance Genes and QTLs

Mapping disease resistance genes and QTLs has been the centerpiece of all mapping studies in tomato. Identification of markers associated with disease resistance in tomato started with the pioneering work of C.M. Rick and co-workers who identified a linkage between root-knot nematode resistance gene *Mi* and isozyme locus *Aps-1* (Rick and Fobes 1974). Since then genetic markers, in particular DNA markers, have been used extensively to tag or map major genes or QTLs for resistance to many fungal, bacterial, viral, and nematode diseases in tomato. In Table 2, all known mapped disease resistance genes and QTLs in tomato together with gene/QTL symbol, the causal agent of the disease, genetic source of the resistance and chromosomal location of the resistance gene or QTL are presented. The space limitation does not allow discussion of the procedures/methodologies employed to identify and map genes or QTLs for the different diseases. However, some general comments are as follow:

- With a few exceptions, most disease resistance genes and QTLs in tomato have been identified in the related wild species and mapped using interspecific populations.
- For some diseases, often multiple gene resources have been used to identify and map resistance genes and QTLs. Multiple resistance genes/QTLs from different sources may allow pyramiding of resistance using a MAS approach.
- For most horizontal resistance traits, often multiple QTLs have been identified in each study. In many cases, it has been difficult to determine the precise location

or actual effect or importance of each QTL in the original studies. Many studies have suggested development of near-isogenic lines (NILs) to determine such information.

- In most cases, early filial and backcross populations have been employed for gene/QTL mapping. More recently, however, advanced populations such as RILs, BILs and ILs have been used to provide better mapping resolutions (Table 2).
- Recently MAS has been extensively employed for disease resistance breeding in tomato, in particular in many seed companies. The utility of MAS for resistance breeding appears to have superseded its utility for any other trait in tomato, as described in below.

3.2. Mapped Insect Resistance/Tolerance Genes and QTLs

Comparatively less research has been done to identify or map genes and QTLs for insect resistance than disease resistance in tomato. This is due in part to many challenges involved in conducting controlled experiments on insect resistance. However, some mapping experiments have been conducted, much of which using *L. pennellii* accession LA716 as the resistance source. The multiple pest resistance of this accession is mediated by acylsugars exuded by type-IV glandular trichomes on the surface of the plants (Mutschler et al. 1996), which act as feeding deterrents for insects such as potato aphid, green peach aphid, tomato fruitworm and beet armyworm, and as feeding or oviposition deterrents for leafminer and silverleaf whitefly. Using an F_2 population of a *L. esculentum* × *L. pennellii* (LA716) cross, a total of five QTLs were detected on chromosomes 2, 3, 4 and 11 with association with one or more aspects of acylsugar production (Mutschler et al. 1996). Transferring subsets of 3–5 QTLs to a *L. esculentum* background resulted BC_3 progeny which exhibited accumulation of only low levels of acylsugars, prompting the researchers to speculate presence of other yet unidentified genetic factors controlling acylsugars accumulation (Lawson et al. 1997). In a different study, using an F_2 population of a cross between two *L. pennellii* accessions, a total of 6 QTLs were detected (Blauth et al. 1999), which were different from those identified in the previous study, further suggesting the complexity of genetic controls of acylsugar production. There is no report on transferring of these QTLs to the cultivated tomato.

Specific insect resistance genes often confer resistance to only one insect species or to a closely related species. However, the *Mi* gene, which confers resistance to a root-knot nematode, is an interesting exception. After *Mi* was cloned (Kaloshian et al. 1998) it was determined that it was the same as *Meu1*, a locus conferring resistance to the potato aphid, *Macrosiphum euphorbiae* (Rossi et al. 1998). Currently, *Mi* (*Meu1*) is the only insect resistance gene that has been cloned from a plant species. This gene is a member of nucleotice binding, leucine-rich repeat family of plant resistance genes, many members of which have been found to confer isolate-specific resistance to viruses, bacteria, fungi, and nematodes. However, *Mi* is the first example of a plant resistance gene active against two such distantly related

Table 2. Summary of disease resistance genes and QTLs (Q) mapped in tomato chromosomes

Disease	Gene/QTL	Pathogen	Resistance source	Mapping population[a]	Chromosomal location	References
Alternaria stem canker	Asc	*Alternaria alternata* f. sp. *lycopersici*	*L. pennellii*	F_2	3	(Mesbah et al. 1999; van der Biezen et al. 1995)
Anthracnose ripe rot	*Anthracnose* (Q)	*Colletotrichum coccodes*	*L. esculentum*	F_2	Various Chromosomes	(Stommel and Zhang 1998)
Bacterial canker	*Cm 1.1–10.1* (Q)	*Calvibacter michiganensis* ssp. *michiganensis*	*L. peruvianum*	BC_1	1,6,7,8,9,10	(Sandbrink et al. 1995)
	Three (Q)	*Calvibacter michiganensis* ssp. *michiganensis*	*L. peruvianum*	F_2	5,7,9	(van Heusden et al. 1999)
	Rcm2.0, Rcm5.1 (Q)	*Calvibacter michiganensis* ssp. *michiganensis*	*L. hirsutum*	BC_2S_5	2,5	(Coaker and Francis 2004; Kabelka et al. 2002)
Bacterial speck	*Pto*	*Pseudomonas syringae* pv. *Tomato (Pst)*	*L. pimpinellifolium*	NIL F_2	5	(Martin et al. 1993b)
	Prf	*Required for res. to Pst*	*L. pimpinellifolium*	NIL F_2	5	(Salmeron et al. 1996)
Bacterial spot	*Rx-1, Rx-2, Rx-3 (Q)*	*Xanthomonas campestris* pv. *vesicatoria*	*L. esculentum*	BC_1	1,5	(Yu et al. 1995)
	Bs4	*Xanthomonas compestris* pv. *vesicatoria*	*L. pennellii*	F_2	5	(Ballvora et al. 2001)
	Rx-3 (Q)	*Xanthomonas campestris* pv. *vesicatoria*	*L. esculentum*	F_2 and BC_3	5	(Yang et al. 2005)
Bacterial wilt	Several (Q)	*Ralstonia solanacearum*	*L. pimpinellifolium*	F_2	3,4,6,7,10,11	(Danesh et al. 1994; Thoquet et al. 1996)

	Two (Q)	*Ralstonia solanacearum*	*L. peruvianum*	F_3	6	(Mangin et al. 1999)
Blackmold	*Blackmold (Q)*	*Alternaria alternata*	L. cheesmanii		2,3,9,12	(Robert et al. 2001)
Corky root rot	*Py-1*	*Pyrenochaeta lycopersici*	*L. peruvianum*	NIL F_2	3	(Doganlar et al. 1998)
Cucumber mosaic v.	*Cmr*	CMV	*L. chilense*	BC_1-inbred	12	(Stamova and Chetelat 2000)
Early blight	*11 (Q)*	*Alternaria solani*	*L. hirsutum*	BC_1	1,2,5,8,9,10,11,12	(Foolad et al. 2002)
	13 (Q)	*Alternaria solani*	*L. hirsutum*	BC_1S_1	1,2,3,5,8,9,10,11,12	(Foolad et al. 2002)
	7 (Q)	*Alternaria solani*	*L. hirsutum*	BC_1	1,2,3,5,8,9,10,11,12	(Zhang et al. 2003)
	7 (Q)	*Alternaria solani*	*L. pimpinellifolium*	RILs	1,3,4,5,6,9,11	MR Foolad et al, unpubl
Fusarium crown and root rot	*Frl*	*Fusarium oxysporum* f. sp. *radidicis-lycopersici*	*L. peruvianum*	F_2	9	(Vakalounakis et al. 1997)
Fusarium wilt	*I, I1, I2, I2C, I3*	*Fusarium oxysporum* f. sp. *lycopersici*	*L. pimpinellifolium* *L. pennellii*	Different populations	7,8,11	(Sarfatti et al. 1991; Scott et al. 2004; Simons et al. 1998)
Gray leaf spot	*Sm*	*Stemphylium* spp.	*L. pimpinellifolium*	F_2	11	(Behare et al. 1991)
Late blight	*Ph-1*	*Phytophthora infestans*	*L. pimpinellifolium*	BC_1	7	(Pierce 1971)
Late blight	*Ph-2*	*P. infestans*	*L. pimpinellifolium*	F_2	10	(Moreau et al. 1998)
Late blight	*Ph-3*	*P. infestans*	*L. pimpinellifolium*	F_2	9	(Chunwongse et al. 1998)
Late blight	*Ph-4*	*P. infestans*	*L. pimpinellifolium*	RILs	2	Foolad et al. (unpubl.)
Late blight	*Ph-5*	*P. infestans*	*L. pimpinellifolium*	F_2	1	Foolad et al. (unpubl.)
	lb1-lb12 (Q)	*P. infestans*	*L. hirsutum*	BC_1	All 12 chrs.	(Brouwer et al. 2004)
Leaf mould	*Cf-1, Cf-2, Cf-4, Cf-5, Cf-9*	*Cladosporium fulvum*	*L. hirsutum* *L. pimpinellifolium*	F_2, NIL F_2, BC_1	1,6	(Balint-Kurti et al. 1994; Jones et al. 1993; Lauge et al. 1998)

(continued)

Table 2. (continued)

Disease	Gene/QTL	Pathogen	Resistance source	Mapping population[a]	Chromosomal location	References
Nematode (potato cyst)	*Hero*	*Globodera restochiensis*	*L. pimpinellifolium*	NIL F_2	4	(Ganal et al. 1995)
Nematode (root-knot)	*Mi, Mi-1, Mi-2, Mi-3, Mi-9*	*Meloidogyne* spp.	*L. peruvianum*	F_2, F_3, NIL F_2, $BC_{1,2}$	6,12	(Aarts et al. 1991; Ammiraju et al. 2003; Veremis et al. 1999; Yaghoobi et al. 1995)
Potyviruses	*pot-1*	*Potyviruses*	*L. hirsutum*	ILs	3	(Parrella et al. 2002)
Powdery Mildew	Lv	*Leveillula taurica*	*L. chilense*	F_2	12	(Chunwongse et al. 1994)
	Ol-1, Ol-2	*Oidium lycopersicum*	*L. esculentum* *L. hirsutum*	F_2	4,6	(de Giovanni et al. 2004; Huang et al. 2000)
	3 (Q)	*Oidium lycopersicum*	*L. parviflorum*	F_2	6,12	(Bai et al. 2003)
Tobacco mosaic virus	*Tm-1, Tm-2*[a]	*TMV/ToMV*	*L. hirsutum* *L. peruvianum*	F_2	2,9	(Levesque et al. 1990; Ohmori et al. 1996; Vakalounakis et al. 1997)
Tomato mottle virus	One *Gene*	*ToMoV*	*L. chilense*	F_2	6	(Griffiths and Scott 2001)
Tomato spotted wilt v.	*Sw-5*	*TSWV*	*L. peruvianum*	NILs, F_2	9	(Stevens et al. 1995; Langella et al. 2004)
Tomato yellow leaf curl virus	*Ty-1 (Q)*, two other Q	*TYLCV*	*L. chilense* *L. pimpinellifolium* *L. hirsutum*	BILs, F_2, F_3, F_4	6,11	(Chagué et al. 1997; Hanson et al. 2000; Zamir et al. 1994)
Verticillium wilt	*Ve*	*Verticillium dahliae*	*L. cheesmanii* *L. pennellii*	F_2, RILs, ILs	9	(Diwan et al. 1999; Kawchuk et al. 1998)

[a] BIL: backcross inbred line; NIL: near isogenic line; RIL: recombinant inbred line

organisms belonging to two different phyla. Another study revealed that several isolates of potato aphid and green peach aphid (*Myzus persicae*) can overcome the resistance mediated by *Mi* (*Meu1*), limiting the use of this gene for aphid control in tomato (Goggin et al. 2001). In a more recent study, however, seven BILs, derived from crosses between two aphid susceptible *L. esculentum* lines and two aphid resistant accessions of *L. pennellii* and *L. hirsutum*, were identified exhibiting resistance to both types of aphid (Kohler and St. Clair 2005). These lines can be useful for developing tomato cultivars with genetic resistance to different types of aphid. In summary, although there has been limited research and progress in identification and mapping of genes/QTLs for insect resistance in tomato, the increasing restrictions on the use of pesticides and the advancements in marker technology are expected to change this in the direction of marker-assisted breeding of insect-resistant tomato cultivars.

3.3. Mapped Abiotic Stress Tolerance/Resistance Genes and QTLs

Most abiotic stress tolerance traits are complex, rendering crop improvement for such traits via traditional breeding unrewarding. Identification and use of genetic markers that are associated with stress related traits, however, is a promising approach to improve plant stress tolerance. During the past two decades, considerable efforts have been made to identify and map genes or QTLs conferring tolerance to different abiotic stresses in tomato, including salinity, drought and extreme temperatures (for a review see Foolad, 2005). Table 3 displays the major mapping activities in tomato, and below is a brief discussion of some of the studies.

More mapping studies have been done on salt tolerance than tolerance to any other environmental stress in tomato (Foolad 2005). For example, QTLs have been identified for salt tolerance during seed germination, vegetative growth and later stages (Table 3) and in different interspecific populations, in particular derived from *L. esculentum* × *L. pennellii* and *L. esculentum* × *L. pimpinellifolium* crosses. Comparisons of QTLs across populations and studies indicate that, for each trait, often similar QTLs were identified in populations derived from the same cross whereas across species some QTLs were conserved and others species-specific (Foolad and Chen 1998; Foolad et al. 1998b). Also, it seems that at each stage similar QTLs contribute to tolerance under different stress levels (Foolad and Jones 1991). Compared to germination and vegetative stages, limited research has been conducted to identify QTLs for salt tolerance during reproduction. However, various studies suggest that different QTLs control tolerance at different stages (Foolad 1999a, 2005). Nevertheless, QTLs for tolerance at different stages can be pyramided using a MAS approach to develop tomatoes with improved salt tolerance throughout the plant ontogeny.

Limited research has been conducted to identify genes or QTLs for cold tolerance in tomato. The only published research for cold tolerance during seed germination is that of Foolad et al. (1998a) in which a few QTLs were identified in an interspecific backcross population of a cross between *L. esculentum* and *L. pimpinellifolium*. More recently, two studies were undertaken to identify QTLs for cold tolerance

Table 3. Summary of abiotic stress tolerance/resistance QTLs mapped in tomato chromosomes

Stress	Specific trait[a]	Numberof QTLs	Tolerance source	Mapping population[b]	Chromosomal location	References
Cold (low temp.)	SG	3	*L. pimpinellifolium*	BC_1S_1	1,4	(Foolad et al. 1998a)
		5	*L. pimpinellifolium*	RILs	1,2,3,8,12	MR Foolad et al, unpubl.
	VG	3	*L. hirsutum*	BC_1	6,7,12	(Vallejos and Tanksley 1983)
	Sht. wlt., RAU	10	*L. hirsutum*	BC_1	1,3,5,6,7,9,11,12	(Truco et al. 2000)
Drought	SG	4	*L. pimpinellifolium*	BC_1S_1	1,8,9,12	(Foolad et al. 2003a)
		8	*L. pimpinellifolium*	RILs	1,2,3,4,8,9,12	MR Foolad et al, unpubl.
	WUE	3	*L. pennellii*	BC_1S_1, F_3	Undetermined	(Martin et al. 1989)
Salt	SG	5	*L. pennellii*	F_2	1,3,7,8,12	(Foolad and Jones 1993)
		8	*L. pennellii*	F_2	1,2,3,7,8,9,12	(Foolad et al. 1997)
		8	*L. pennellii*	F_2	1,3,5,6,8,9	(Foolad and Chen 1998)
		7	*L. pimpinellifolium*	BC_1S_1	1,2,5,7,9,12	(Foolad et al. 1998b)
		8	*L. pimpinellifolium*	RILs	1,2,3,4,8,9,12	MR Foolad et al, unpubl.
	VG	4	*L. pimpinellifolium*	BC_1S_1	1,5,9	(Foolad and Chen 1999)
		5	*L. pimpinellifolium*	BC_1	1,3,5,6,9	(Foolad et al. 2001)
		7	*L. pimpinellifolium*	RILs	1,3,4,5,7,8,9	MR Foolad et al, unpubl.
	Ion accumulation	6	*L. pennellii*	F_2	1,2,4,5,6,12	(Zamir and Tal 1987).
	FN, FW, FY	Several	*L. pimpinellifolium*	F_2	Undetermined	(Breto et al. 1994)

[a] FN: fruit number; FW: fruit weight; FY: fruit yield; RAU: root ammonium uptake; SG: seed germination; Sht. wlt.: shoot wilting; VG: vegetative growth; WUE: water use efficiency

[b] RIL: recombinant inbred line

during seed germination, one using a selective genotyping in a large (N = 1000) *L. esculentum* × *L. pimpinellifolium* BC_1 population and the second using an F_9 RIL population (N = 145) of the same cross (MR Foolad et al., unpubl.). These studies verified all of the QTLs identified in the original study (Foolad et al. 1998a) and detected a few additional QTLs. A comparison of QTLs indicated that most QTLs were stable across populations whereas a few were population-specific. QTL mapping studies for cold tolerance during vegetative growth and reproduction are scarce. In one study, using BC_1 population of a cross between a cold-sensitive *L. esculentum* line and a cold-tolerant *L. hirsutum* accession, Vallejos and Tanksley (1983) identified 3 QTLs for growth at low temperatures. In another study, several QTLs were identified associated with shoot wilting and root ammonium uptake under chilling temperatures in a *L. esculentum* × *L. hirsutum* BC_1 population (Truco et al. 2000).

Research to identify genes or QTLs for drought tolerance in tomato has also been limited. The only published report on drought tolerance during seed germination is that of Foolad et al. (2003a) in which 4 QTLs were detected in backcross progeny of a *L. esculentum* × *L. pimpinellifolium* cross. As a follow-up, more recently an F_9 RIL population of the same cross was evaluated for drought tolerance during germination, where several QTLs were identified consistent with the results of the first study (MR Foolad et al., unpubl.). Limited research has been done to discern genetic controls of drought tolerance during vegetative and reproductive stages. To facilitate selection for low Δ ($^{13}C/^{12}C$ discrimination) as a means to improve tomato drought tolerance, 3 QTLs associated with this trait were identified using filial and backcross progeny of a *L. esculentum* × *L. pennellii* cross (Martin et al. 1989). However, it has not been determined whether selection for these QTLs would increase water use efficiency in tomato. There is no published research on QTLs for drought tolerance during reproductive stage.

Although several studies have investigated physiological and genetic relationships among tolerances to different environmental stresses in tomato, only a few studies have used QTL mapping techniques, which mainly focused on seed germination stage (Foolad 1999b, 2000; Foolad et al. 1999; Foolad et al. 2003b; Foolad et al. unpubl.). These studies, however, have indicated the presence of genetic relationships among cold, salt, and drought tolerance during seed germination in tomato. For example, a few QTLs were identified with effects on germination under two or three different stresses, which were referred to as stress-nonspecific QTLs. Also, a few QTLs were identified with effects on germination only under specific stress conditions, which were referred to as stress-specific QTLs. In summary, the various studies have indicated that some genes affect tomato seed germination under different stress conditions while others are more stress-specific.

3.4. Mapped Genes and QTLs for Flower- and Fruit-Related Characteristics

During the past few decades, markers and maps have been used to identify genes or QTLs for many flower- and fruit-related characteristics in tomato, as listed in

Table 4. It is obvious from the table that often several groups conducted research on the same or similar traits, or the same traits were studied using different interspecific populations. For example, QTLs for fruit weight (size), shape and soluble solids contents were identified in at least 20, 13 and 23 studies, respectively (Table 4). In below, the status of marker development for a few major fruit traits in tomato is reviewed and discussed.

3.4.1. Fruit size

Although there are variations in fruit size in the cultivated tomato, differences are much greater between the cultivated (with fruit size as big as 1000 g with 10 or more locules) and wild species (with fruit size as small as 1 g with only 2 locules). Therefore, most QTL mapping studies have employed interspecific populations. Although traditional breeding studies had suggested simple genetic control for tomato fruit size, molecular marker research has revealed the presence of approximately 30 QTLs for this trait, mapping to different chromosomes (Table 4) (Chen et al. 1999). Many studies, however, have identified same QTLs for this trait, and the most recent studies have not detected any novel QTL (Lippman and Tanksley 2001). Also, many studies have indicated co-localization of QTLs for fruit size and total (or soluble) solids contents (e.g. see Chen et al., 1999), confirming the presence of a negative correlation between these two traits (Georgelis et al. 2004). One of the major QTLs for fruit size in tomato, *fw2.2*, which was detected in many studies (Table 4), has been cloned (Frary et al. 2000), as discussed in below.

3.4.2. Fruit shape

There are great variations in fruit shape in tomato, ranging from oblate to round to ovate (blocky or elongated) to pear shape. Earlier studies had identified several genes controlling fruit shape in tomato, including *pr* (pyriform), *o* (ovate), *bk* (beaked tomato), *n* (nipple-tip tomato), *f* (fasciated) and *lc* (for locule number). During the past two decades, however, a few of these genes and several other genes and QTLs affecting fruit shape have been located on tomato chromosomes and/or cloned (Tables 4 and 6). For example, a major fruit-shape QTL (*fs8.1*) differentiating fresh-market (round) and processing (blocky) tomatoes was mapped on chromosome 8 (Grandillo et al. 1996) and later cloned and characterized (Ku et al. 2000). *fs8.1* exert its effect on fruit shape by changing the length of carpels during pre-anthesis, resulting in longer and larger fruit. Similarly, a major fruit-shape QTL (*ovate*), controlling the transition from round to pear-shaped fruit was mapped (Ku et al. 1999) and cloned (Liu et al. 2002). The overall results from different studies indicate that much of the variation in tomato fruit shape is controlled by a few major loci and the great variation observed in the cultivated species is most likely the result of allelic variation at these loci (van der Knaap and Tanksley 2003).

3.4.3. Fruit color/nutritional-value

Color is one of the most important quality characteristics in tomato, and the focus of numerous mapping studies. The attention to fruit color has recently increased as

Table 4. Summary of important characteristics for which genes or QTLs have been mapped in tomato chromosomes

Trait[a]	Wild species used	Mapping population[b]	Genes (G) or QTLs (Q)	Chromosome	References
Carotenoid candidate genes	*L. pennellii*	ILs	23 G	Most chromosomes	(Liu et al. 2003b)
Corolla indentation	*L. hirsutum*	BC_1	2 Q	2,8	(Bernacchi and Tanksley 1997)
Flwr., exerted stigma	*L. peruvianum*	BC_3, BC_4	2 Q	2,9	(Fulton et al. 1997)
Frt. antioxidant capacity	*L. pennellii*	ILs	5 Q	3,6,7,10	(Rousseaux et al. 2005)
Frt. ascorbic acid	*L. pennellii*	ILs	6 Q	3,5,10,12	(Rousseaux et al. 2005)
Frt. citric acid content	*L. pennellii*	ILs	7 Q	4,5,8,9,10	(Causse et al. 2004)
Frt. color	*L. pimpinellifolium*	BC_1	2 Q	2,6	(Grandillo and Tanksley 1996b)
	L. pimpinellifolium	BC_2, BC_3	5 Q	2,3,4,7,8	(Tanksley et al. 1996)
	L. peruvianum	BC_3, BC_4	8 Q	1,6,7,8,9,10,12	(Fulton et al. 1997)
	L. hirsutum	BC_2, BC_3	15 Q	1,2,4,5,7,8,9,10,11	(Bernacchi et al. 1998)
	L. hirsutum	subNILs	1 Q	1	(Monforte and Tanksley 2000a)
	L. parviflorum	BC_3	13 Q	1,2,4,5,7,8,11,12	(Fulton et al. 2000)
	L. chmielewskii	NILs	1 Q	1	(Frary et al. 2003)
	L. pennellii	BC_2/BC_2F_1	1 Q	12	(Frary et al. 2004b)
Frt. color (β-carotene)	*L. cheesmannii*	F_2, ILs	1 G *(B)*	6	(Ronen et al. 2000; Zhang and Stommel 2000)
	L. parviflorum	*BC_3*	6 Q	2,4,8,9,10,11	(Fulton et al. 2000)
	L. pennellii	ILs	1 G *(B)*	6	(Ronen et al. 1999; Zhang and Stommel 2001)
	L. pennellii	ILs	2 Q	6	(Rousseaux et al. 2005)
Frt. color (carotene)	*L. esculentum*	RIL	3 Q	2,3,8	(Saliba-Colombani et al. 2001)

(continued)

Table 4. (continued)

Trait[a]	Wild species used	Mapping population[b]	Genes (G) or QTLs (Q)	Chromosome	References
Frt. color (crimson)	*L. esculentum*	F_2	1 G *(og[c], cr)*	6	(Ronen et al. 2000; Thompson et al. 1967)
Frt. color (external)	*L. parviflorum*	BC_3	9 Q	1,2,4,5,7,8,11,12	(Fulton et al. 2000)
	L. pimpinellifolium	BC_2F_6	2 Q	3,11	(Doganlar et al. 2002)
	L. pennellii	BC_2/BC_2F_1	2 Q	5,12	(Frary et al. 2004b)
Frt. color (high-pigment-1)	*L. cheesmannii*	F_2	*1 G (hp-1)*	2	(Peters et al. 1998; Yen et al. 1997)
Frt. color (high-pigment-2)	*L. pennellii*	BC_1	1 G *(hp-2)*	1	(Mustilli et al. 1999; van Tuinen et al. 1997)
Frt. color (internal)	*L. parviflorum*	BC_3	15 Q	1,2,4,5,7,8,9,10,11,12	(Fulton et al. 2000)
	L. pimpinellifolium	BC_2F_6	3 Q	3,11	(Doganlar et al. 2002)
	L. pennellii	BC_2/BC_2F_1	1 Q	12	(Frary et al. 2004b)
	L. pennellii	ILs	16 Q	Most chromosomes	(Liu et al. 2003b)
Frt. color (lycopene)	*L. pimpinellifolium*	BC_1S_1	8 Q	1,4,5,6,7,10,12	(Chen et al. 1999)
	L. parviflorum	BC_3	5 Q	2,3,5,8,12	(Fulton et al. 2000)
	L. pennellii	ILs	2 Q	3,6	(Rousseaux et al. 2005)
Frt. color (orange)	*L. pennellii*	BC_2/BC_2F_1	2 Q	11,12	(Frary et al. 2004b)
Frt. color (yellow)	*L. parviflorum*	BC_3	1 Q	12	(Fulton et al. 2000)
Frt. diameter	*L. pimpinellifolium*	BC_1	3 Q	1,2,8	(Grandillo and Tanksley 1996b)
	L. pimpinellifolium	BC_1S_1	8 Q	1,2,3,6,7,11	(Chen et al. 1999)
	L. pimpinellifolium	F_2	7 Q	1,2,3,4,7,11	(Lippman and Tanksley 2001)
	L. pimpinellifolium	BC_2F_6	12 Q	2	(Doganlar et al. 2002)
Frt. fructose content	*L. pennellii*	ILs	4 Q	4,5,7,9	(Causse et al. 2004)
Frt. fruc:gluc ratio	*L. hirsutum*	$F_{2,3}$, BC_1	1 G *(Fgr)*	4	(Levin et al. 2000)
Frt. glucose content	*L. pennellii*	ILs	4 Q	4,5,9,12	(Causse et al. 2004)
Frt. length	*L. pimpinellifolium*	BC_1S_1	9 Q	1,2,3,6,7,9,12	(Chen et al. 1999)

	L. pimpinellifolium	F_2	7 Q	1,2,3,4,9,11	(Lippman and Tanksley 2001)
	L. pimpinellifolium	BC_2F_6	5 Q	2,3,8,9,11	(Doganlar et al. 2002)
Frt. locule number	*L. pimpinellifolium*	BC_1	2 Q	1,3	(Grandillo and Tanksley 1996b)
	L. pimpinellifolium	F_2	3 Q	2,11	(Lippman and Tanksley 2001)
	L. pimpinellifolium	F_2	5 Q	2,3,4,10,12	(Frary et al. 2004a)
Frt. malic acid content	*L. pennellii*	ILs	5 Q	3,4,8,12	(Causse et al. 2004)
Frt. organoleptic quality	*L. esculentum*	RILs	Many Q	Various chromos.	(Causse et al. 2002)
Frt. ostwald	*L. parviflorum*	BC_3	1 Q	6	(Fulton et al. 2000)
Frt. pH	*L. chmielewskii*	BC_1, BC_2	5 Q	3,6,7,8,10	(Paterson et al. 1988)
	L. cheesmanii	F_2, F_3	9 Q	1,3,4,6,7,8,10	(Paterson et al. 1991)
	L. chmielewskii	BILs/BC_2F_5	1 Q	7	(Azanza et al. 1995)
	L. pimpinellifolium	BC_2, BC_3	5 Q	1,3,5,7,12	(Tanksley et al. 1996)
	L. peruvianum	BC_3, BC_4	6 Q	2,3,9,10,12	(Fulton et al. 1997)
	L. hirsutum	BC_2, BC_3	10 Q	1,2,3,4,6,8,9,10,12	(Bernacchi et al. 1998)
	L. pimpinellifolium	BC_1S_1	6 Q	1,2,4,5,9,12	(Chen et al. 1999)
	L. parviflorum	BC_3	10 Q	2,3,4,5,6,7,9,12	(Fulton et al. 2000)
	L. esculentum	RIL	2 Q	11,12	(Saliba-Colombani et al. 2001)
	L. pennellii	BC_2/BC_2F_1	2 Q	3,12	(Frary et al. 2004b)
	L. pennellii	ILs	11 Q	2,4,5,8,9,10,11,12	(Causse et al. 2004)
Frt. phenolic content	*L. pennellii*	ILs	9 Q	3,5,6,7,8,9	(Rousseaux et al. 2005)
Frt. ripening	*L. esculentum*	F_2	2 Q	5,12	(Doganlar et al. 2000)
	L. pennellii	F_2	Many loci	All chromosomes	(Kinzer et al. 1990; Slater et al. 1985)
	L. pimpinellifolium	BC_1	3 Q	2,8,9	(Grandillo and Tanksley 1996b)
	L. pimpinellifolium	BC_2, BC_3	4 Q	2,4,7,8	(Tanksley et al. 1996)
	L. peruvianum	BC_3, BC_4	4 Q	2,3,8,9	(Fulton et al. 1997)
Frt. rip. (*Colorless non-rip.*)	*L. cheesmanii*	F_2	*Cnr*	2	(Tr et al. 2002)
Frt. ripening (*never-ripe*)	*L. cheesmanii*	F_2	*Nr*	9	(Yen et al. 1995)

(continued)

Table 4. (continued)

Trait[a]	Wild species used	Mapping population[b]	Genes (G) or QTLs (Q)	Chromosome	References
Frt. ripening (*non-ripening*)	*L. pennellii L. cheesmanii*	F_2	*nor, rin*	5,10	(Giovannoni et al. 1999; Moore et al. 2002; Vrebalov et al. 2002)
Frt. ripening (polygalacturonase)	*L. pimpinellifolium*	BC_1	TOM6	10	(Kinzer et al. 1990)
Frt. ripening (*uniform rip.*)	*L. pimpinellifolium*	BC_1	*u*	10	(Kinzer et al. 1990)
Frt. shape	*L. pimpinellifolium*	BC_1S_1	4 Q	1,9,10,12	(Chen et al. 1999)
	L. hirsutum	subNILs	1 Q	1	(Monforte and Tanksley 2000a)
	L. parviflorum	BC_3	16 Q	All 12 chromosomes	(Fulton et al. 2000)
	L. pimpinellifolium	F_2	1 Q	11	(Lippman and Tanksley 2001)
	L. pimpinellifolium	BC_2F_6	2 Q	1,9	(Doganlar et al. 2002)
	L. pimpinellifolium	F_2	4 Q	2,3,7,11	(van der Knaap et al. 2000)
	L. pennellii	BC_2/BC_2F_1	4 Q	2,8,10,12	(Frary et al. 2004b)
Frt. shape (bumpiness)	*L. pimpinellifolium*	F_2	3 Q	8,9,11	(van der Knaap and Tanksley 2003)
Frt. shape (bell-pepper)	*L. pimpinellifolium*	F_2	3 Q	2,8	(van der Knaap and Tanksley 2003)
	L. pennellii	F_2	1 Q	2	(Ku et al. 1999)
Frt. shape (elongated)	*L. pimpinellifolium*	F_2	2 Q	6,9	(van der Knaap and Tanksley 2003)
Frt. shape (heart)	*L. pimpinellifolium*	F_2	4 Q	1,2,3,7	(van der Knaap and Tanksley 2003)
Frt. shape (pear)	*L. pimpinellifolium*	F_2	2 Q	2,10	(Ku et al. 1999)

Frt. shape (blockiness)	*L. pimpinellifolium*	F_2	7 Q	1,2,3,7,8,12	(van der Knaap and Tanksley 2003)
Frt. soluble solids (SS)	*L. pimpinellifolium*	BC_2, BC_3	12 Q	2,3,4,5,6,7,8,11,12	(Tanksley et al. 1996)
	L. pimpinellifolium	BC_1	3 Q	3,6,9	(Grandillo and Tanksley 1996b)
	L. peruvianum	BC_3, BC_4	9 Q	1,2,7,8,9,10	(Fulton et al. 1997)
	L. hirsutum	BC_2, BC_3	5 Q	3,5,6,9	(Bernacchi et al. 1998)
	L. pimpinellifolium	BC_1S_1	13 Q	1,2,3,7,10,12	(Chen et al. 1999)
	L. hirsutum	subNILs	1 Q	1	(Monforte and Tanksley 2000a)
	L. parviflorum	BC_3	5 Q	4,5,6,9	(Fulton et al. 2000)
	L. esculentum	RIL	3 Q	2,9	(Saliba-Colombani et al. 2001)
	L. pimpinellifolium	BC_2F_6	2 Q	8,9	(Doganlar et al. 2002)
	L. pennellii	BC_2/BC_2F_1	3 Q	4,9,12	(Frary et al. 2004b)
	L. pennellii	ILs	9 Q	1,3,4,5,7,9,10,12	(Causse et al. 2004)
Frt. SS × red yield	*L. pennellii*	ILs	14 Q	Most chromosomes	(Eshed and Zamir 1995)
	L. pimpinellifolium	BC_2, BC_3	4 Q	3,7,9	(Tanksley et al. 1996)
	L. peruvianum	BC_3, BC_4	9 Q	1,2,5,7,8,9,10,12	(Fulton et al. 1997)
	L. hirsutum	BC_2, BC_3	9 Q	1,2,3,4,6,8,11,12	(Bernacchi et al. 1998)
	L. parviflorum	BC_3	2 Q	5,8	(Fulton et al. 2000)
	L. pennellii	BC_2/BC_2F_1	4 Q	3,5,12	(Frary et al. 2004b)
Frt. total acid	*L. parviflorum*	BC_3	4 Q	3,4,7,8	(Fulton et al. 2000)
Frt. viscosity	*L. pimpinellifolium*	BC_2, BC_3	1 Q	9	(Tanksley et al. 1996)
	L. peruvianum	BC_3, BC_4	4 Q	1,2,8,9	(Fulton et al. 1997)
	L. parviflorum	BC_3	3 Q	2,9,10	(Fulton et al. 2000)
	L. pennellii	BC_2/BC_2F_1	4 Q	2,3,9,12	(Frary et al. 2004b)
Frt. weight	*L. pimpinellifolium*	BC_1	7 Q	1,2,8,11	(Grandillo and Tanksley 1996b)
	L. pimpinellifolium	BC_2, BC_3	8 Q	2,3,4,5,7,9	(Tanksley et al. 1996)
	L. peruvianum	BC_3, BC_4	10 Q	1,2,3,7,8,9,10,12	(Fulton et al. 1997)
	L. hirsutum	BC_1	3 Q	1,3	(Bernacchi and Tanksley 1997)
	L. hirsutum	BC_2, BC_3	3 Q	2,3,4	(Bernacchi et al. 1998)

(continued)

Table 4. (continued)

Trait[a]	Wild species used	Mapping population[b]	Genes (G) or QTLs (Q)	Chromosome	References
	L. pimpinellifolium	BC_1S_1	12 Q	1,2,3,4,6,7,8,9,11,12	(Chen et al. 1999)
	L. parviflorum	BC_3	8 Q	2,3,6,7,10,11,12	(Fulton et al. 2000)
	L. esculenum	F_2	2 Q	4,6	(Doganlar et al. 2000)
	L. pimpinellifolium	F_2	6 Q	1,2,3,11	(Lippman and Tanksley 2001)
	L. esculentum	RIL	5 Q	2,3,11,12	(Saliba-Colombani et al. 2001)
	L. pimpinellifolium	BC_2F_6	2 Q	2,3	(Doganlar et al. 2002)
	L. pimpinellifolium	F_2	7 Q	1,2,3,5,6,7,11	(van der Knaap and Tanksley 2003)
	L. pennellii	BC_2/BC_2F_1	3 Q	3,10,12	(Frary et al. 2004b)
Frt. yield (total yield)	*L. pimpinellifolium*	BC_2, BC_3	6 Q	2,3,7,9	(Tanksley et al. 1996)
	L. peruvianum	BC_3, BC_4	10 Q	1,2,6,7,8,9,10,12	(Fulton et al. 1997)
	L. hirsutum	BC_2, BC_3	12 Q	1,2,3,4,5,6,7,8,12	(Bernacchi et al. 1998)
	L. hirsutum	subNILs	1 Q	1	(Monforte and Tanksley 2000a)
	L. parviflorum	BC_3	5 Q	1,2,3,6,8	(Fulton et al. 2000)
	L. pennellii	BC_2/BC_2F_1	6 Q	3,5,8,9,12	(Frary et al. 2004b)
Frt. yield (red yield)	*L. pimpinellifolium*	BC_2, BC_3	2 Q	2,7	(Tanksley et al. 1996)
	L. peruvianum	BC_3, BC_4	12 Q	1,2,3,5,7,8,9,10,12	(Fulton et al. 1997)
	L. hirsutum	BC_2, BC_3	11 Q	1,2,3,5,7,8,10,11,12	(Bernacchi et al. 1998)
	L. parviflorum	BC_3	4 Q	2,5,8	(Fulton et al. 2000)
	L. pennellii	BC_2/BC_2F_1	4 Q	3,5,12	(Frary et al. 2004b)
Frt. yield (green yield)	*L. hirsutum*	BC_2, BC_3	11 Q	2,3,7,8,11,12	(Bernacchi et al. 1998)
	L. pennellii	BC_2/BC_2F_1	3 Q	8,9,12	(Frary et al. 2004b)
Jointless	*L. cheesmanii*	F_2	*j*	11	(Wing et al. 1994; Zhang et al. 1994)
	L. cheesmanii	F_2	*j-2*	12	(Zhang et al. 2000)
Male sterility	*L. pimpinellifolium*	F_2	ms-10	2	(Tanksley et al. 1984)

Plant growth habit	*L. peruvianum*	BC_3, BC_4	6 Q	1,2,3,7,8,9	(Fulton et al. 1997)
	L. pimpinellifolium	BC_2F_6	4 Q	2,3,9,11	(Doganlar et al. 2002)
Plant height	*L. pennellii*	ILs	16 Q	Many chromosomes	(Eshed and Zamir 1995)
	L. pimpinellifolium	BC_1	1 Q	2	(Grandillo and Tanksley 1996b)
	L. hirsutum	BC_1	4 Q	1,2,5,11	(Bernacchi and Tanksley 1997)
	L. cheesmanii	F_8 RIL	7 Q	2,3,4,6,7	(Paran and Zamir 1997)
Phytochrome genes	*L. pennellii*	BC_1	5 G	1, 2, 5, 7, 10	(van Tuinen et al. 1997)
Seed number	*L. pimpinellifolium*	BC_1	4 Q	4,6,7,12	(Grandillo and Tanksley 1996b)
	L. pimpinellifolium	F_2	2 Q	1,11	(Lippman and Tanksley 2001)
	L. pimpinellifolium	BC_2F_6	4 Q	5,6,8	(Doganlar et al. 2002)
	L. pimpinellifolium	F_2	10 Q	1,2,3,4,6,7,9,11,12	(van der Knaap et al. 2000)
Seed weight	*L. pennellii*	BC_1	5 Q	1,2,4,7,8	(Tanksley et al. 1982)
	L. cheesmanii	F_8 RILs	14 Q	1,2,3,4,6,7,9,11,12	(Goldman et al. 1995)
	L. pimpinellifolium	BC_1	4 Q	2,4,10,12	(Grandillo and Tanksley 1996b)
	L. pimpinellifolium	F_2	4 Q	1,2,4	(Lippman and Tanksley 2001)
	L. pimpinellifolium	BC_2F_6	5 Q	1,4,5,7	(Doganlar et al. 2002)
Self-incompatibility	*L. peruvianum*	F_1	*S* locus	1	(Tanksley and Loaiza-Figueroa 1985)
	L. peruvianum	F_1	*S* locus	1	(Bernatzky 1993)
	L. hirsutum	BC_1	*S* locus	1	(Bernacchi and Tanksley 1997)
Self-pruning	*L. esculentum*	F_2	sp	6	(Ito and Currence 1964)
	L. chmielewskii	BC_1	sp	6	(Paterson et al. 1988)
	L. pimpinellifolium	BC_1	sp	6	(Grandillo and Tanksley 1996a)
Transgr. segregation	*L. pennellii*	F_2	74 Q/8 traits	All 12 chromosomes	(de Vicente and Tanksley 1993)
Unilateral incongruity	*L. hirsutum*	BC_1	6 Q	1,2,3,11,12	(Bernacchi and Tanksley 1997)

[a] D = distance; Flwr. = flower; Frt. = fruit; L = length; W = width

[b] NIL: near isogenic line; RIL: recombinant inbred line

the health benefits of lycopene, the major carotenoid responsible for the red color of tomato, has become more obvious (Gerster 1997). Previously, several major genes with significant effect on fruit lycopene (e.g, *hp-1*, *hp-2* , *dg* and Og^c) and other carotenoids contents (e.g. beta-carotene, *B*) had been identified and mapped onto the classical linkage maps of tomato (Wann and Jourdain 1985). During the past two decades, however, many studies identified numerous QTLs and candidate genes with significant effects on fruit color and/or carotenoids content Table 4. Although some QTLs were mapped to the chromosomal locations of known genes in the carotenoid biosynthesis pathway, many were mapped to new locations suggesting presence of additional genes affecting fruit color in tomato (Liu et al. 2003b). Currently, a few research groups around the world conduct research to identify, map and possibly clone new genes involved in determining fruit color in tomato. In addition, there are numerous programs attempting to improve tomato nutritional quality either through traditional breeding or transgenic approaches (Ronen et al. 1999; Zhang and Stommel 2000).

3.4.4. Fruit soluble solids

Fruit soluble solids content (SSC) has long been the focus of numerous tomato breeding programs worldwide. However, despite significant improvement in fruit size and yield, SSC has remained almost unchanged due in part to the negative relationship between fruit yield and SSC. Thus, breeders have sought alternative approaches to improve SSC of high yielding tomato cultivars. During the past two decades, considerable efforts have been devoted to identifying QTLs for SSC in order to facilitate improving fruit SSC without compromising the yield. The hope has been to identify QTLs that may not have any adverse effects on fruit size or yield. Currently, there are more than 20 published studies that have reported QTLs for fruit SSC in tomato (Table 4). Although these studies used different interspecific populations, there have been significant overlaps in the locations of QTLs. Furthermore, many studies have revealed that most QTLs that positively influence SSC are at the same locations as QTLs with negative impact on fruit weight (Chen et al. 1999), an unfortunate discovery. However, few studies have reported the identification of QTLs for SSC with little or no detrimental impact on fruit size (e.g. (Fridman et al. 2000)), though the effects of such QTLs are yet to be verified.

3.4.5. Fruit yield

Because of the complexity of yield, technically it is difficult to identify QTLs which are truly affecting this trait and could be utilized in marker-assisted breeding to improve crop yield. Nevertheless, many QTL mapping studies have tried to identify QTLs for fruit yield in tomato, including traits such as total yield, red yield and green yield, mostly using interspecific segregating populations (Table 4). Unlike QTLs for fruit weight and SSC that have been rather consistent across studies, there has been limited agreement in different studies with regard to yield QTLs (Table 4). This is not surprising considering the very complex nature of the trait and its low heritability. Also, there is no published report on the use or verification of fruit yield

QTLs for marker-assisted breeding in tomato, and it is not expected that such QTLs will have wide utility for improving tomato yield, at least in the near future. It is expected, however, that QTLs for components of fruit yield, such as fruit size, number and SSC, will be useful as indirect selection criteria to increase tomato fruit yield.

3.4.6. Fruit ripening

Traditional genetics and breeding research had resulted in the identification and manipulation of several ripening-related genes in tomato. During the past two decades, however, molecular biology techniques facilitated characterization of such genes and identification and mapping of many other ripening related genes, loci and QTLs (Table 4). Furthermore, a few ripening-related genes have been fine-mapped and cloned (Table 6). More detailed lists of genes with effects on fruit ripening in tomato can be found in Giovannoni (2001). Among the major ripening genes, at least one (*rin*) has been used in marker-assisted breeding (Table 5).

3.5. Mapped Genes and QTLs for Other Traits

In addition to the traits described in above, molecular markers have been employed to tag or map genes or QTLs for many other horticultural characters in tomato, as shown in Table 4. Many of these studies have contributed significantly to our better understanding of the genetic controls of different morphological and physiological characters in tomato. Examples include self-incompatibility, unilateral incongruity, transgressive segregation, self-prunning (determinate type plants), jointless pedicel, and seed size and number. The marker information can be useful for further basic research and for crop improvement through marker-assisted breeding.

3.6. Populations used for Gene and QTL Mapping

As in many other plant species, in tomato early filial and backcross populations have been used most widely for gene and QTL mapping (Tables 2–4). In some cases, however, advanced populations such as recombinant inbred lines (RILs), advanced backcross populations (AB; e.g. BC_2 and BC_3), backcross inbred lines (BILs, a.k.a. inbred backcross lines or IBC, e.g., BC_2S_3, BC_3S_2) and introgression lines (ILs) have been employed for genetic mapping. Currently a few tomato RIL populations are available, including one based on a *L. esculentum* × *L. cheesmanii* cross (Paran and Zamir 1997) and three based on different *L. esculentum* × *L. pimpinellifolium* crosses (Foolad et al. unpubl.; Graham et al. 2004;). RIL populations have been used for mapping QTLs for various traits, including fruit weight and SS (Goldman et al. 1995), morphological characters (Paran and Zamir 1997), abiotic stress tolerance (Foolad 2004), seed weight (Goldman et al. 1995) and disease resistance (Diwan et al. 1999; Foolad et al., unpubl). BILs have been used frequently to map QTLs for many traits (Tables 2–4), including fruit quality (Doganlar et al. 2002; Tanksley et al. 1996; Yang et al. 2005) and disease resistance (Coaker and Francis 2004; Foolad et al. 2002; Robert et al. 2001). Greater levels of homozygosity in the RIL

and BIL populations, compared with early segregating populations, have allowed more precise estimation of the location and effects of QTLs. Another type of population that has been extensively used for mapping in tomato is introgression lines (ILs). The first developed IL population of tomato consisted of 50 lines, each containing a single introgression from *L. pennellii* LA716 in the background of tomato cultivar M-82 (Eshed and Zamir 1995). Since then the number of ILs in this population has increased to 76, which together represent the entire genome of the *L. pennellii* LA716 in homozygous or heterozygous conditions and delimit 107 marker-defined mapping bins, each bin having an average length of 12 cM (Gur et al. 2004; Liu et al. 2003b). In addition to this IL population, a total of 99 NILs and BILs derived from a cross between cultivar E6203 and *L. hirsutum* accession LA1777 has been developed (Monforte and Tanksley 2000b). Most of the lines contain a single defined introgression from *L. hirsutum* in the *L. esculentum* genetic background, and together the lines cover more than 85% of the LA1777 genome. IL populations can be used for MAS pyramiding of important QTLs, as in the case of tomato yield and soluble solids QTLs (Gur and Zamir 2004). Furthermore, some ILs and BILs have been used to develop NILs for fine mapping and map-based cloning of genes and QTLs controlling traits such as self-prunning (Pnueli et al. 1998), color mutants (Isaacson et al. 2002; Ronen et al. 1999), SSC (Fridman et al. 2000; Yates et al. 2004), fruit weight (Frary et al. 2000) and shape (van der Knaap et al. 2004; Yates et al. 2004), stigma exsertion (Chen and Tanksley 2004) and a few others (Table 6).

4. MARKER-ASSISTED SELECTION AND BREEDING

4.1. Current Use of MAS

The use of MAS in tomato breeding started in 1980s, when many seed companies took advantage of the reported linkage between *Mi* and *Aps-1*[1] and used the isozyme marker to select for nematode resistance. Recently, however, MAS has become a routine practice in many seed companies for improving tomatoes for many simple-inherited traits (Table 5). Unfortunately, most of these activities are not reported in public literature. However, a survey by the author indicated that many seed companies in the U.S. (e.g., Seminis Vegetable Seeds (now owned by Monsanto), Syngenta, Harris Moran, Sakata, Asgrow) and in Europe (e.g., Nunhems Zaden, Vilmorin, Seminis Vegetable Seeds Holland, ENZA, RijkZwaan, DeRuiter) routinely employ MAS for tomato disease resistance breeding. Examples include vertical resistance to corky root, fusarium wilt, late blight, nematodes, powdery mildew, bacterial speck, tobacco/tomato mosaic virus, tomato spotted wilt virus, and verticillium wilt (Table 5). MAS is also routinely practiced in seed companies for improvement of a few other simple traits in tomato, including, but not limited to, self-prunning (*sp* gene), jointless pedicel, ripening, and carotenoids contents (lycopene and β-carotene). However, there is little indication of the use of MAS for manipulating QTLs for complex traits, though MAS is being attempted to improve

Table 5. Known traits for which marker-assisted selection and breeding are done in tomato

Trait	Source species	Gene/QTL (Q)	Reference
Bacterial canker	*L. peruvianum*	3 Q	Seed companies
	L. hirsutum	*Rcm2.0*, *Rcm5.1*(Q)	(Coaker and Francis 2004)
Bacterial speck	*L. pimpinelli-folium*	Pto	Seed companies, (Yang and Francis 2005)
Bacterial spot	*L. esculentum*	*Rx-3* (Q)	(Yang and Francis 2005)
Bacterial wilt	*L. esculentum*	2 Q	Seed companies
Blackmold	*L. cheesmanii*	Few Q	(Robert et al. 2001)
Corky root rot	*L. peruvianum*	*Py-1*	Seed companies
Fusarium wilt	*L. pimpinelli-folium*	*I-2C, I-3*	Seed companies
Jointless	*L. cheesmanii*		Seed companies
Late blight	*L. pimpinelli-folium*	*Ph-3*	Seed companies, University researchers
	L. hirsutum	4 Q	(Brouwer and St. Clair 2004)
Lycopene	*L. esculentum*	*Og*^{c}, *cr*	Seed companies
Powdery mildew	*L. chilense, L. hirsutum*	*Lv, Ol-1, Ol-2*	Seed companies
Ripening inhibitor	*L. cheesmanii*	*rin*	Seed companies
Root-knot nematode	*L. peruvianum*	*Mi*	Seed companies, public breeders
Self-prunning	*L. esculentum*	*sp*	Seed companies, public breeders and geneticists
Soluble solids	Not known	*Q*	Seed companies
Tomato spotted wilt	*L. peruvianum*	*Sw-5*	(Langella et al. 2004), seed companies
Tomato YLCV	Different species	Few Q	Seed companies
Tobacco mosaic virus	*L. peruvianum*	*Tm-2*[a]	Seed companies
Verticillium wilt	*L. esculentum*	*Ve*	Seed companies

°Brix as well as quantitative resistance to bacterial wilt, bacterial canker, bacterial wilt, powdery mildew and yellow leaf curl virus.

The use of MAS is less common in public tomato breeding programs, though it has been practiced to improve vertical resistance to a few diseases (Barone 2005) such as late blight (M Mutschler, Cornell University; MR Foolad, Penn State University), bacterial canker, bacterial speck and bacterial spot (Coaker et al. 2004), and horizontal resistance to blackmold (Robert et al. 2001) and late blight (Brouwer and St. Clair 2004). It also has been employed for simple traits such as self-pruning (many tomato genetics and breeding programs) and some complex fruit quality characters (Lecomte et al. 2004). However, based on the published results, in most cases where MAS was employed to transfer QTLs there were significant linkage drag problems. For example, in case of late blight where three near-isogentic lines were developed each containing one resistance QTL from *L. hirsutum*, while all lines showed expected level of resistance they also suffered from undesirable horticultural characteristics

Table 6. Fine-mapped and/or cloned genes and QTLs in tomato

Trait	Gene/QTL	Chromosome	Source species	Fine mapping population	Nature/activity/ function	Reference
Aphid (potato) res.	*Meu*	6	*L. peruvianum*	NIL F_2	NBS-LRR	(Rossi et al. 1998; Vos et al. 1998)
Bacterial speck res.	*1 Pto*	5	*L. pimpinelli-folium*	NIL F_2	Protein kinase	(Gu and Martin 1998; Martin et al. 1993a)
Bacterial spot res.	*Bs4*	5	*L. pennellii*	NIL F_2	TIR-NB-LRR	(Schornack et al. 2004)
Fenthion resistance	*Prf*	5	*L. pimpinelli-folium*	NIL F_2	NBS-LRR Resistance gene	(Martin et al. 1994; Salmeron et al. 1996)
Flower, exerted stigma	*se2.1 (Q)*	2	*L. pennellii*	NIL F_2 and F_3	Affects floral morphology	(Chen and Tanksley 2004)
Frt. color (β-carotene)	*B*	6	*L. pennellii*	NIL F_2	Lycopene β-cyclase	(Ronen et al. 2000; Ronen et al. 1999)
Frt. color (crimson)	og^c, *cr*	6	*L. esculentum*	NIL F_2	Lycopene cyclase null allele	(Ronen et al. 2000; Ronen et al. 1999)
Frt. color (high-pigment-2)	*hp-2*	1	*L. esculentum*	N/A	Homologue of DEETIO-LATED1	(Mustilli et al. 1999)
Frt. color (old gold)	*og*	6	*L. esculentum*	NIL F_2	Lycopene cyclase null allele	(Ronen et al. 2000)
Frt. color (tangerine)	*CRISTO*	10	*L. esculentum*	NIL F_2	Carotenoid isomerase	(Isaacson et al. 2002)
Frt. rip. (Colorless non-rip.)	*Cnr*	2	*L. esculentum* *L. cheesmanii*	F_2 and FISH	Not determined	(Tør et al. 2002)
Frt. ripening (never-ripe)	*Nr*	9	*L. esculentum*	NIL F_2	Blocks ethylene perception	(Lanahan et al. 1994; Wilkinson et al. 1995; Yen et al. 1995)
Fruit ripening (non-ripening)	*nor*	10	*L. esculentum*	NIL F_2	MADS-box	(Giovannoni et al. 1995; Moore et al. 2002; Vrebalov et al. 2002)

Frt. ripening (polygalacturonase)	TOM6	10	*L. esculentum*	N/A	Pectin hydrolyzing	(Dellapenna et al. 1986; Grierson et al. 1986; Slater et al. 1985)
Fruit ripening (ripening-inhibitor)	rin	5	*L. cheesmanii* *L. pennellii*	NIL F_2	MADS-box	(Giovannoni et al. 1995; Moore et al. 2002; Vrebalov et al. 2002)
Fruit shape	*fs8.1 (Q)*	8	*L. pimpinelli-folium*	NIL F_2	Imparts blocky shape	(Ku et al. 2000; Tanksley 2004)
	Sun (Q)	7	*L. esculentum*	NIL F_2 and F_3	Imparts oval shape	(van der Knaap et al. 2004)
	ovate (Q)	2	*L. pimpinelli-folium*	NIL F_2	Plant-growth suppressor	(Ku et al. 1999; Liu et al. 2002)
Fruit weight	*fw2.2 (Q)*	2	*L. pennellii*	NIL F_2	Controls carpel cell number	(Alpert et al. 1995; Alpert and Tanksley 1996; Frary et al. 2000; Liu et al. 2003a)
Fusarium wilt resistance	*I2*	11	*L. esculentum*	NIL F_2	Leucine zipper and LRR-NBS	(Simons et al. 1998)
Growth habit	*PW9-2-5 (Q)*	9	*L. pennellii*	F_2	Semi-det. growth	(Fridman et al. 2002)
Iron uptake	*chloronerva*	1	*L. pennellii*	NIL F_2	Nicotianamine synthase	(Ling et al. 1999)
Jointless	*j*	11	*L. esculentum*	F_2, NIL F_2	Suppresses form. abscission zone	(Mao et al. 2001; Wing et al. 1994; Zhang et al. 1994)
	j-2	12	*L. cheesmanii*	F_2	Suppresses form. abscission zone	(Budiman et al. 2004; Zhang et al. 2000)

(continued)

Table 6. (continued)

Trait	Gene/QTL	Chromosome	Source species	Fine mapping population	Nature/activity/ function	Reference
Leaf mould resistance	*Cf-2,4,5,9*	1,6	*L. peruvianum*	NIL F_2	LRR	(Dixon et al. 1996)
Nematode (root-knot) resistance	*Mi-1.1Mi-1.2*	6	*L. peruvianum*	NIL F_2	NBS-LRR	(Kaloshian et al. 1998; Rossi et al. 1998; Vos et al. 1998)
Self-pruning (SP)	*sp*	6	*L. esculentum*	NIL F_2	Regulates cycle veg./repr. growth	(Pnueli et al. 1998)
SP gene family	*sp21, 3D, 5G, 6A, 9D*	2,3,5,6,9	*L. pennellii*	NIL F_2	Not determined	(Carmel-Goren et al. 2003)
Self-incompatibility	*S*	1	*L. peruvianum*	N/A	RNase activity	(Rivers et al. 1993)
Soluble solids	*Brix9-2-5,Lin5 (Q)*	9	*L. pennellii*	NIL F_2	Apoplastic invertase	(Fridman et al. 2004; Fridman et al. 2000)
Tomato spotted wilt virus	*Sw-5*	9	*L. peruvianum*	NIL F_2	NBS-LRR Resistance gene	(Adkins 2000; Brommenschenkel et al. 2000; Langella et al. 2004)
Verticillium wilt	*Ve1, Ve2*	9	*L. esculentum*	N/AcDNA cloning	Glycoproteins w/ leucine zipper or Pro-Glu-Ser-Thr	(Kawchuk et al. 2001)

(Brouwer and St. Clair 2004). Further inspections of the NILs resulted in the detection of QTLs for other characters such as plant shape, canopy density, maturity, fruit yield and fruit size in the same introgressed regions. The results prompted the investigators to suggest further refining of the late blight QTLs before transferring to adapted genetic backgrounds. Similar conclusions were reached as to MAS transfer of QTLs for blackmold resistance from *L. cheesmanii* to the cultivated tomato, where predominantly negative associations were found between introgressed QTL alleles and horticultural characteristics (Robert et al. 2001). Such undesirable associations have been reported also for many other complex traits in tomato, and in most cases it was not determined whether they were due to genetic linkage or pleiotropic effects. However, with the current rough state of QTL detection, it is not unexpected that similar negative associations be experienced when practicing MAS for other complex traits. It is prudent that issues surrounding the utility of MAS should be addressed before it can become a routine practice for crop improvement.

4.2. Issues in Using MAS

Practical utility of MAS for manipulating simple traits has been well documented. However, despite tremendous investment in finding markers associated with important genes and QTLs, MAS has not become a routine practice in most plant breeding programs, in particular for improving complex traits. The issues associated with the use of MAS include: (1) elevated cost of high-throughput marker genotyping, which is not affordable by most breeding programs considering the large size of populations need to be screened; (2) unavailability of reliable PCR-based markers for many simple and complex traits; (3) lack of QTL validation across genetic backgrounds and environments; (4) large size of QTL intervals and association with undesirable characteristics; (5) unfamiliarity of many traditional plant breeders, who in fact release most of the modern cultivars, with the use of markers or their limited access to molecular marker laboratories; and (6) identification and mapping of genes and QTLs mainly by researchers who are not breeders or who may not have inherent interest in crop improvement. A better cooperation between basic scientists and plant breeders is needed to coordinate meaningful identification and use of gene/QTL-linked markers. Another significant limiting factor is lack of sufficient marker polymorphism in the cultivated species. Many tomato breeders focus on exploitation of genetic variation within the cultivated species, where MAS cannot be easily employed using traditional molecular markers. However, recent efforts in discovering SNPs within the cultivated tomato is expected to reduce this problem and facilitate the use of markers when exploiting intraspecific genetic variations.

5. FINE MAPPING AND MAP-BASED CLONING

5.1. Fine Mapping and Map-Based Cloning of Major Genes

Tomato was the first plant species in which a disease resistance gene, *Pto*, conferring resistance to bacterial speck caused by *Pseudomonas syringae* pv. *Tomato* (*Pst*), was

cloned using map-based cloning approach (Martin et al. 1993a). Further analysis indicated similarity of the open reading frame of *Pto* to serine-threonine protein kinases (Table 6) (Martin et al. 1994). Subsequently, similar map-based cloning strategy was employed and several other tomato genes were cloned, including *Prf*, which is required for *Pto* activity and tomato resistance to *Pst* and which also confers tomato susceptibility to organophosphate insecticide Fenthion (Salmeron et al. 1996), *Sw-5*, conferring resistance to tospovirus (Brommenschenkel and Tanksley 1997), *sp* (self pruning), conferring determinate growth habit (Pnueli et al. 1998), members of *sp* gene family (Carmel-Goren et al. 2003), and *j* and *j-2*, controlling jointless pedicel (Mao et al. *2000*). It should be noted that both *j* and *j-2* are recessive mutants that completely suppresses the formation of pedicel abscission zone (Budiman et al. 2004; Mao et al. *2000;* Zhang et al. 2000; Zhang et al. 1994), which is an essential character and widely used in both processing and fresh-market tomato cultivars. Several other major genes that have been fine-mapped and/or cloned via map-based cloning approach in tomato are listed in Table 6.

5.2. Fine Mapping and Map-Based Cloning of QTLs

Most QTL mapping studies detect QTLs within large marker intervals, usually 10 cM or greater, and often cannot determine whether a detected QTL contains one or more genes. Also, detected QTLs often affect more than one trait and it is unknown whether such effects are due to pleiotropic effects of the same genes/QTLs or genetic linkage of independent loci. In tomato, for example, the lower part of chromosome 1 has been identified to affect many agriculturally important traits, including morphological and fruit quality characteristics as well as resistance to many biotic and abiotic stresses (Frary et al. 2003; Monforte and Tanksley 2000a). Similarly, the long arm of chromosome 4 contains QTLs for many horticulturally important traits such as fruit SSC, shape and lycopene content (Yates et al. 2004). Furthermore, it has been demonstrated that introduction of a small segment of DNA from a wild into the cultivated species of tomato often results in significant changes in several characteristics, and it is unknown whether the same genes/QTLs affect different traits. During the past several years, however, advances in marker technology have facilitated physical characterization of QTL introgressions. For example, a NIL of tomato containing a 40-cM introgression from *L. hirsutum* accession LA1777 at the bottom of chromosome 1 was dissected by developing sub-NILs containing smaller segments (Monforte and Tanksley 2000a). In a similar study, a series of sub-NILs were developed containing introgressions from chromosome 1 of *L. chmielewskii* to fine map loci controlling several fruit quality characteristics important to processing tomato cultivars (Frary et al. 2003). It was determined that independent loci in the lower part of chromosome 1 affected fruit °Brix, yield and shape, whereas genetic factors affecting fruit weight, shoulder pigmentation and external color coincided with the location of a °Brix locus. These results, combined with results of other studies, prompted the investigators to conclude that the base of tomato chromosome 1 contains multiple QTLs whose effects cannot be attributed to the pleiotropic effects

of a single locus (Frary et al. 2003; Monforte and Tanksley 2000a). In another study, NILs containing the lower part of chromosome 4 from either *L. peruvianum* or *L. hirsutum* were dissected by developing series of sub-NILs containing small introgressions from either wild species (Yates et al. 2004). Results indicated presence of multiple, non-allelic loci controlling SSC and fruit weight in addition to other loci controlling fruit shape, weight and epidermal reticulation. In addition to these studies, many other QTL coincidences have been reported in tomato and other crop species, however, in most cases the nature of such coincidences has not been determined.

During the past decade, efforts have been made to clone QTLs and assert that QTLs have the same molecular bases as regular Mendelian genes. In fact, much of such efforts have been made in tomato. For example, the first map-based cloning of a QTL in plants was done for a fruit size QTL *(fw2.2)* in tomato (Frary et al. 2000). While tomato improvement for fruit size has been easy due to its high heritability, both inheritance studies and QTL mapping experiments revealed that this trait was controlled by several loci (Chen et al. 1999). To date, most, if not all, QTLs involved in the evolution of tomato fruit size (from small to large) have been identified and mapped (Table 4) (Lippman and Tanksley 2001). In many studies, one major QTL (*fw2.2*) was found to be associated with large phenotypic variation for fruit size (Chen et al. 1999; Grandillo et al. 1999). While the modern tomato cultivars carry large-fruit alleles at this locus, all wild *Lycopersicon* species examined contain small-fruit alleles (Alpert et al. 1995). Because of its large, consistently-detectable effects, a map-based cloning approach was pursued, which resulted in cloning and sequencing of *fw2.2* (Frary et al. 2000; Liu et al. 2003a). It was determined that *fw2.2* was expressed early in floral development and controlled carpel cell number. Subsequently, similar strategy was employed to fine-map and/or clone a few other QTLs in tomato affecting traits such as SSC, fruit shape and exerted stigma (Table 4). It is expected that with the advances in marker technology and QTL identification, more QTLs will be cloned using map-based cloning approach.

6. TOMATO GENOME ORGANIZATION AND SEQUENCING

The tomato nuclear genome comprises 12 chromosomes and approximately 950 Mb of DNA, containing 59% non-coding sequences, 28% coding sequences, 11% transposons, and 2% organellar sequences (Wang et al. 2005). Approximately 77% of the chromosomal DNA is comprised of centromeric heterochromatic regions, which are devoid of genes. The majority of genes are populated at distal euchromatic regions of the chromosomes, with an approximate gene density of 6.7 kb/gene, similar to that of *Arabidopsis* and rice. This study also indicates that a significant portion of the tomato euchromatin is methylated in the intergenic spacer regions. Currently the 12 tomato chromosomes are being sequenced by an international consortium of 10 countries, including China (chr. 3), France (chr. 7), India (chr. 5), Italy (chr. 12), Japan (chr. 8), Koria (chr. 2), The Netherlands (chr. 6), Spain (chr. 9), England (chr. 4), and the U. S. (chrs. 1, 10 and 11)

(*http://www.sgn.cornell.edu/help/about/tomato_sequencing.pl*.). This effort is part of a larger initiative known as the International Solanaceae Genome Project (SOL): Systems Approach to Diversity and Adaptation. Lunched in 2003, this project has set research goals for the next 10 years (*http://www.sgn.cornell.edu/solanaceae-project/index.pl*.). The major emphasis is on physical, evolutionary and functional genomics of the family Solanaceae. The first cornerstone of the project, however, is to generate a high-quality sequence for the euchromatic portions of the tomato chromosomes, as a reference for the Solanaceae. Concomitantly, however, other genome organizations studies are being conducted in tomato. For example, extensive efforts are being made to expand EST database of tomato, and to date more than 162,000 ESTs developed (*http://www.tigr.org/tigr-scripts/tgi/T_index.cgi?species=tomato*). Although the EST-derived unigene sets of tomato do not represent the entire gene repertoire of this species, analysis of the tomato EST database and several sequenced BAC libraries have led to the prediction that tomato genome encodes ~35,000 genes, largely sequestered in euchromatic regions of the 12 tomato chromosomes, which correspond to less than 25% of the total nuclear DNA in tomato (van der Hoeven et al. 2002; Wang et al. 2005).

7. REFLECTION AND PROSPECTS

Remarkable progress has been made in generation of molecular markers and maps, mapping of genes and QTLs, and fine-mapping, map-based cloning and characterization of genes and QTLs in tomato. Comparatively, much less has been accomplished as to the actual use of marker information in tomato crop improvement, in particular for complex traits. There are several reasons for this discrepancy, as indicated in above. However, as judged based on most recent discoveries and progresses, the prospect for future use of markers in tomato improvement is good. Perhaps the most important element is development of low-cost, high-throughput PCR-based markers that are easy-to-assay and have high resolution. In particular markers that can detect polymorphism within the cultivated species or between the cultivated and closely-related wild species such as *L. pimpinellifolium* and *L. cheesmanii*. As sequencing of the tomato genome progresses, it is expected that the information will be utilized to develop more informative markers. It is also expected that a greater emphasis will be placed on development of functional markers, such as ESTs and candidate gene markers, which will be highly useful in both applied and basic research programs. Because of these and other advancements, it is not unexpected that in a near future MAS becomes a routine procedure in most tomato-breeding programs. Many breeders are convinced that even for many simple traits with high heritability MAS does have an edge over PS because of various potential limitations in phenotypic screening.

Currently, there is little evidence of actual use of markers for improving complex traits in tomato. A major limiting factor is uncertainties about the reliability of detected QTLs, as discussed in above. For many complex traits, such as yield and tolerance to abiotic stresses, obtaining reliable phenotypic data for QTL detection

is not straightforward, often leading to identification of QTLs which may not be trustworthy. Improvements must be made in our techniques to identify more tractable QTLs. One approach is to seek QTLs which control components of the genetic variation, rather than QTLs detected based on screening the ultimate trait itself. For example, partitioning of the total variation for a complex trait into contributing components and identification of QTLs for each component may lead to the detection of more reliable and useful QTLs. A subsequent and complementary step is to further refine and fine-map the identified QTLs by developing NILs and sub-NILs. Such refining may not only determine the actual value of each QTL, but also may establish whether any potential negative association due to linkage drag can be broken before transferring the QTLs. In other words, fine-mapping would allow detection of QTLs that are useful for marker-assisted breeding. The significance of such refinements is well recognized among tomato geneticists and breeders, and many research programs have initiated such activities. Another reason for optimism regarding the application of marker technology in tomato improvement is the increasing use of F_1 hybrid cultivars for commercial production. When developing hybrids, the use of MAS will not only be more practical but also more economical. However, in spite of all expected advancements and the optimism, MAS may not be a "silver bullet" solution to every breeding problem in tomato. Most likely, a combination of traditional protocols and marker-assisted breeding will become a routine procedure in many tomato improvement programs. I expect a long time before traditional breeding protocols become totally obsolete, if they ever become.

REFERENCES

Aarts JMMJG, Hontelez JGJ, Fischer P, Verkerk R, van Kammen A, Zabel P (1991) Acid phosphatase-1^1, a tightly linked molecular marker for root-knot nematode resistance in tomato: from protein to gene, using PCR and degenerate primers containing dexoyinosine. Plant Mol Biol 16:647–661

Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, Merril CR, Wu A, Olde B, Moreno RF, Kerlavage AR, McCombie WR, Venter JC (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. Science 252:1651–1656

Adkins S (2000) Tomato spotted wilt virus – positive steps towards negative success. Mol Plant Pathol 1:151–157

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

Alpert KB, Grandillo S, Tanksley SD (1995) *fw 2.2*: a major QTL controlling fruit weight is common to both red- and green-fruited tomato species. Theor Appl Genet 91:994–1000

Ammiraju JSS, Veremis JC, Huang X, Roberts PA, Kaloshian I (2003) The heat-stable root-knot nematode resistance gene *Mi-9* from *Lycopersicon peruvianum* is localized on the short arm of chromosome 6. Theor Appl Genet 106:478–484

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

Azanza F, Kim D, Tanksley SD, Juvik JA (1995) Genes from *Lycopersicon chmielewskii* affecting tomato quality during fruit ripening. Theor Appl Genet 91:495–504

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

Balint-Kurti PJ, Dixon MS, Jones DA, Norcott KA, Jones JDG (1994) RFLP linkage analysis of the *Cf-4* and *Cf-9* genes for resistance to *Cladospsorium fulvum* in tomato. Theor Appl Genet 88:691–700

Ballvora A, Pierre M, van den Ackerveken G, Schornack S, Rossier O, Ganal M, Lahaye T, Bonas U (2001) Genetic mapping and functional analysis of the tomato *Bs-4* locus governing recognition of the *Xanthomonas campestris* pv. *vesicatoria*. Mol Plant Microbe Interact 14:629–638

Barone A (2005) Molecular marker-assisted selection for pyramiding resistance genes in tomato. Adv Hort Sci 19:147–152

Behare J, Laterrot H, Sarfatti M, Zamir D (1991) RFLP mapping of the *Stemphylium* resistance gene in tomato. Mol Plant Microbe Interact 4:489–492

Bernacchi D, Tanksley SD (1997) An interspecific backcross of *Lycopersicon esculentum* × *L. hirsutum*: Linkage analysis and a QTL study of sexual compatibility factors and floral traits. Genetics 147:861–877

Bernacchi D, Beck-Bunn T, Eshed Y, Lopez J, Petiard V, Uhlig J, Zamir D, Tanksley S (1998) Advanced backcross QTL analysis in tomato. I. Identification of QTLs for traits of agronomic importance from *Lycopersicon hirsutum*. Theor Appl Genet 97:381–397

Bernatzky R (1993) Genetic mapping and protein product diversity of the self-incompatibility locus in wild tomato (*Lycopersicon peruvianum*). Bioch Genet 31:173–184

Bernatzky R, Tanksley SD (1986) Towards a saturated linkage map in tomato based on isozymes and random cDNA sequences. Genetics 112:887–898

Blauth SL, Steffiens JC, Churchill GA, Mutschler MM (1999) Identification of QTLs controlling acylsugar fatty acid composition in an intraspecific population of *Lycopersicon pennellii* (Corr) D'Arcy. Theor Appl Genet 99:373–381

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphism. Am J Hum Genet 32:314–331

Breto MP, Asins MJ, Carbonell EA (1994) Salt tolerance in *Lycopersicon* species: III. Detection of quantitative trait loci by means of molecular markers. Theor Appl Genet 88:395–401

Brommenschenkel SH, Tanksley SD (1997) Map-based cloning of the tomato genomic region that spans the *Sw-5* tospovirus resistance gene in tomato. Mol Gen Genet 256:121–126

Brommenschenkel SH, Frary A, Frary A, Tanksley SD (2000) The broad-spectrum tospovirus resistance gene *Sw-5* of tomato is a homolog of the root-knot nematode resistance gene *Mi*. Mol Plant Microbe Interact 2000:1130–1138

Brouwer DJ, St Clair DA (2004) Fine mapping of three quantitative trait loci for late blight resistance in tomato using near isogenic lines (NILs) and sub-NILs. Theor Appl Genet 108:628–638

Brouwer DJ, Jones ES, St Clair DA (2004) QTL analysis of quantitative resistance to *Phytophthora infestans* (late blight) in tomato and comparisons with potato. Genome 47:475–492

Budiman MA, Chang S-B, Lee S, Yang TJ, Zhang H-B, de Jong H, Wing RA (2004) Localization of jointless-2 gene in the centromeric region of tomato chromosome 12 based on high resolution genetic and physical mapping. Theor Appl Genet 108:190–196

Butler L (1968) Linkage summary. Rep Tomato Genet Coop 18:4–6

Carmel-Goren L, Liu Y-S, Lifschitz E, Zamir D (2003) The *SELF-PRUNING* gene family in tomato. Plant Mol Biol 52:1215–1222

Causse M, Saliba-Colombani V, Lecomte L, Duffé P, Rousselle P, Buret M (2002) QTL analysis of fruit quality in fresh market tomato: a few chromosome regions control the variation of sensory and instrumental traits. J Exp Bot 53:2089–2098

Causse M, Duffe P, Gomez MC, Buret M, Damidaux R, Zamir D, Gur A, Chevaller C, Lemaire-Chamley M, Rothan C (2004) A genetic map of candidate genes and QTLs involved in tomato fruit size and composition. J Exp Bot 55:1671–1685

Chagué V, Mercier JC, Guénard M, de Courcel AGL, Vedel F (1997) Identification of RAPD markers linked to a locus involved in quantitative resistance to TYLCV in tomato by bulked segregant analysis. Theor Appl Genet 95:671–677

Chen FQ, Foolad MR (1999) A molecular linkage map of tomato based on a cross between *Lycopersicon esculentum* and *L. pimpinellifolium* and its comparison with other molecular maps of tomato. Genome 42:94–103

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

Chen FQ, Foolad MR, Hyman J, Clair DA St, Beelman RB (1999) Mapping of QTLs for lycopene and other fruit traits in a *Lycopersicon esculentum* × *L. pimpinellifolium* cross and comparison of QTLs across tomato species. Mol Breed 5:283–299

Chetelat RT (2002) Revised list of monogenic stocks. Rep Tomato Genet Coop 52:41–62

Chunwongse J, Bunn TB, Crossman C, Jinag J, Tanksley SD (1994) Chromosomal localization and molecular-marker tagging of the powdery mildew resistance gene (*Lv*) in tomato. Theor Appl Genet 89:76–79

Chunwongse J, Chunwongse C, Black L, Hanson P (1998) Mapping of *Ph-3* gene for late blight from *L. pimpinellifolium* L3708. Rep Tomato Genet Coop 48:13–14

Coaker GL, Francis DM (2004) Mapping, genetic effects, and epistatic interaction to two bacterial canker resistance QTLs from *Lycopersicon hirsutum*. Theor Appl Genet 108:1047–1055

Danesh D, Aarons S, Mcgill GE, Young ND (1994) Genetic dissection of oligogenic resistance to bacterial wilt in tomato. Mol Plant Microbe Interact 7:464–471

de Giovanni C, Dell'Orco P, Bruno A, Ciccarese F, Lotti C, Ricciardi L (2004) Identification of PCR-based markers (RAPD, AFLP) linked to a novel powdery mildew resistance gene (*ol-2*) in tomato. Plant Sci 166:41–48

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

Dellapenna D, Alexandert DC, Bennett AB (1986) Molecular cloning of tomato fruit polygalacturonase: analysis of polygalacturonase mRNA levels during ripening. Proc Natl Acad Sci USA 83: 6420–6424

Diwan N, Fluhr R, Ehsed Y, Zamir D, Tanksley SD (1999) Mapping of *Ve* in tomato: a gene conferring resistance to the broad-spectrum pathogen, *Verticillium dahliae* race 1. Theor Appl Genet 98:315–319

Dixon MS, Jones DA, Keddie JS, Thomas CM, Harrison K, Jones JDG (1996) The tomato *Cf-2* disease resistance locus comprises two functional genes encoding leucine-rich repeat proteins. Cell 84:451–459

Doganlar S, Dodson J, Gabor B, Beck-Bunn T, Crossman C, Tanksley SD (1998) Molecular mapping of the *py-1* gene for resistance to corky root rot (*Pyrenochaeta lycopersici*) in tomato. Theor Appl Genet 97:784–788

Doganlar S, Tanksley SD, Mutschler MA (2000) Identification and molecular mapping of loci controlling fruit ripening time in tomato. Theor Appl Genet 100:249–255

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

Duvick DN (1996) Plant breeding, an evolutionary concept. Crop Sci 36:539–548

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

FAOSTAT (2005) FAO Statistical Databases. Food and agriculture organization of the United Nations, Statistics Division

Foolad MR (1999a) Comparison of salt tolerance during seed germination and vegetative growth in tomato by QTL mapping. Genome 42:727–734

Foolad MR (1999b) Genetics of salt tolerance and cold tolerance in tomato: quantitative analysis and QTL mapping. Plant Biotechnol 16:55–64

Foolad MR (2000) Genetic bases of salt tolerance and cold tolerance in tomato. Curr Top Plant Biol 2:35–49

Foolad MR (2004) Recent advances in genetics of salt tolerance in tomato. Plant Cell, Tissue Org Cult 76:101–119

Foolad MR (2005) Breeding for abiotic stress tolerances in tomato. In: Ashraf M, Harris PJC (eds) abiotic stresses: plant resistance through breeding and molecular approaches. The Haworth Press, New York, pp 613–684

Foolad MR, Chen FQ (1998) RAPD markers associated with salt tolerance in an Interspecific cross of tomato (*Lycopersicon esculentum ¥ L. pennellii*). Plant Cell Rep 17:306–312

Foolad MR, Chen FQ (1999) RFLP mapping of QTLs conferring salt tolerance during vegetative stage in tomato. Theor Appl Genet 99:235–243

Foolad MR, Jones RA (1991) Genetic analysis of salt tolerance during germination in *Lycopersicon*. Theor Appl Genet 81:321–326

Foolad MR, Jones RA (1993) Mapping salt-tolerance genes in tomato (*Lycopersicon esculentum*) using trait-based marker analysis. Theor Appl Genet 87:184–192

Foolad MR, Jones RA, Rodriguez RL (1993) RAPD markers for constructing intraspecific tomato genetic maps. Plant Cell Rep 12:293–297

Foolad MR, Stoltz T, Dervinis C, Rodriguez RL, Jones RA (1997) Mapping QTLs conferring salt tolerance during germination in tomato by selective genotyping. Mol Breed 3:269–277

Foolad MR, Chen FQ, Lin GY (1998a) RFLP mapping of QTLs conferring cold tolerance during seed germination in an interspecific cross of tomato. Mol Breed 4:519–529

Foolad MR, Chen FQ, Lin GY (1998b) RFLP mapping of QTLs conferring salt tolerance during germination in an interspecific cross of tomato. Theor Appl Genet 97:1133–1144

Foolad MR, Lin GY, Chen FQ (1999) Comparison of QTLs for seed germination under non-stress, cold stress and salt stress in tomato. Plant Breed 118:167–173

Foolad MR, Zhang LP, Lin GY (2001) Identification and validation of QTLs for salt tolerance during vegetative growth in tomato by selective genotyping. Genome 44:444–454

Foolad MR, Zhang L, Khan A, Niño-Liu D, Lin GY (2002) Identification of QTLs for early blight (*Alternaria solani*) resistance in tomato using backcross populations of a *Lycopersicon esculentum* ¥ *L. hirsutum* cross. Theor Appl Genet 104:945–958

Foolad MR, Zhang L, Subbiah P (2003a) Genetics of drought tolerance during seed germination in tomato: inheritance and QTL mapping. Genome 46:536–545

Foolad MR, Zhang LP, Subbiah P (2003b) Relationships among cold, salt and drought tolerance during seed germination in tomato: inheritance and QTL mapping. Acta Hort 618:47–57

Frary A, Nesbitt TC, Frary A, Grandillo S, van der Knaap E, Cong B, Liu J-P, Meller J, Elber R, Alpert KB, Tanksley S (2000) *fw2.2*: a quantitative trait locus key to the evolution of tomato fruit size. Science 289:85–88

Frary A, Doganlar S, Fulton TM, Uhlig J, Yates H, Tanksley SD (2003) Fine mapping of quantitative trait loci for improved fruit characteristics from *Lycopersicon chmielewskii* chromosome 1. Genome 46:235–243

Frary A, Fritz LA, Tanksley SD (2004a) A comparative study of the genetic bases of natural variation in tomato leaf, sepal, and petal morphology. Theor Appl Genet 109:523–533

Frary A, Fulton TM, Zamir D, Tanksley S (2004b) Advanced backcross QTL analysis of a *Lycopersicon esculentum* × *L. pennellii* cross and identification of possible orthologs in the Solanaceae. Theor Appl Genet 108:485–496

Frary A, Xu Y, Liu JP, Mitchell S, Tedeschi E, Tanksley SD (2005) Development of a set of PCR-based anchor markers encompassing the tomato genome and evaluation of their usefulness for genetics and breeding experiments. Theor Appl Genet 111:291–312

Fridman E, Pleban T, Zamir D (2000) A recombination hotspot delimits a wild-species quantitative trait locus for tomato sugar content to 484 bp within an invertase gene. Proc Natl Acad Sci USA 97:4718–4723

Fridman E, Liu YS, Carmel-Goren L, Gur A, Shoresh M, Pleban T, Eshed Y, Zamir D (2002) Two tightly linked QTLs modify tomato sugar content via different physiological pathways. Mol Genet Genomics 266:821–826

Fridman E, Carrari F, Liu Y-S, Fernie AR, Zamir D (2004) Zooming in on a quantitative trait for tomato yield using interspecific introgrssions. Science 305:786–789

Fulton TM, Beck-Bunn T, Emmatty D, Eshed Y, Lopez J, Petiard V, Uhlig J, Zamir D, Tanksley SD (1997) QTL analysis of an advanced backcross of *Lycopersicon peruvianum* to the cultivated tomato and comparison with QTLs found in other wild species. Theor Appl Genet 95:881–894

Fulton TM, Grandillo S, Beck-Bunn T, Fridman E, Frampton AJ, Lopez J, Petiard V, Uhlig J, Zamir D, Tanksley SD (2000) Advanced backcross QTL analysis of a *Lycopersicon esculentum* × *Lycopersicon parviflorum* cross. Theor Appl Genet 100:1025–1042

Fulton TM, Van der Hoeven R, Eannetta NT, Tanksley SD (2002) Identification, analysis, and utilizatioin of conserved ortholog set markers for comparative genomics in higher plants. Plant Cell 14:1457–1467

Ganal MW, Simon R, Brommonschenkel S, Tanksley SD, Kumar A (1995) Genetic mapping of a wide spectrum nematode resistance gene (*Hero*) against *Globodera rostochiensis* in tomato. Mol Plant Microbe Interact 8:886–891

Ganal MW, Czihal R, Hannappel U, Kloos D-U, Polley A, Ling H-Q (1998) Sequencing of DNA clones from the genetic map of tomato (*Lycopersicon esculentum*). Genome Res 8:842–847

Georgelis N, Scott JW, Baldwin EA (2004) Relationship of tomato fruit sugar concentration with physical and chemical traits and linkage of RAPD markers. J Am Soc Hort Sci 129:839–845

Gerster H (1997) The potential role of lycopene for human health. J Am Coll Nutr 16:109–126

Giovannoni JJ (2001) Molecular biology of fruit maturation and ripening. Annu Rev Plant Phys Plant Mol Biol 52:725–749

Giovannoni JJ, Noensie EN, Ruezinsky DM, Lu X-H, Tracy SL, Ganal MW, Martin GB, Pillen K, Alpert KB, Tanksley SD (1995) Molecular genetic analysis of the *ripening-inhibitor* and *non-ripening* loci of tomato: a first step in genetic map-based cloning of fruit ripening genes. Mol Gen Genet 248:195–206

Giovannoni JJ, Yen H, Shelton B, Miller S, Vrebalov J, Kannan P, Tieman D, Hackett R, Grierson D, Klee H (1999) Genetic mapping of ripening and ethylene-related loci in tomato. Theor Appl Genet 98:1005–1013

Giovannucci E (1999) Tomatoes, tomato-based products, lycopene, and cancer; Review of the epidemiologic literature. J Natl Cancer Inst 91:317–331

Goggin FL, Williamson VM, Ullrich SE (2001) Variability in the response of *Macrosiphum euphorbiae* and *Myzux persicae* (Himiptera: Aphididae) to the tomato resistance gene *Mi*. Environ Entomol 30:101–106

Goldman IL, Paran I, Zamir D (1995) Quantitative trait locus analysis of a recombinant inbred line population derived from *Lycopersicon esculentum* × *Lycopersicon cheesmanii* cross. Theor Appl Genet 90:925–932

Graham EB, Frary A, Kang JJ, Jones CM, Gardener RG (2004) A recombinant inbred line mapping population derived from a *Lycopersicon esculentum* × *L pimpinellifolium* cross. Rep Tomato Genet Coop 54:22–25

Grandillo S, Tanksley SD (1996a) Genetic analysis of RFLPs, GATA microsatellites and RAPDs in a cross between *L. esculentum* and *L. pimpinellifolium* . Theor Appl Genet 92:957–965

Grandillo S, Tanksley SD (1996b) QTL analysis of horticultural traits differentiating the cultivated tomato from the closely related species *Lycopersicon pimpinellifolium*. Theor Appl Genet 92:935–951

Grandillo S, Ku J, Tanksley SD (1996) Characterization of fs8.1, a major QTL influencing fruit shape in tomato. Mol Breed 2:251–260

Grandillo S, Ku HM, Tanksley SD (1999) Identifying the loci responsible for natural variation in fruit size and shape in tomato. Theor Appl Genet 99:978–987

Grierson D, Tucker GA, Keen J, Ray J, Bird CR, Schuch W (1986) Sequencing and identification of a cDNA clone for tomato polygalacturonase. Nucl Acids Res 14:8595–8603

Griffiths PD, Scott JW (2001) Inheritance and linkage of tomato mottle virus resistance genes derived from *Lycopersicon chilense* accession LA 1932. J Am Soc Hort Sci 126:462–467

Gu Y-Q, Martin GB (1998) Molecular mechanisms involved in bacterial speck disease resistance of tomato. Phil Trans R Soc Lond 353:1455–1461

Gur A, Zamir D (2004) Unsused natural variation can lift yield barriers in plant breeding. PLoS Biol 2:1610–1615

Gur A, Semel Y, Cahaner A, Zamir D (2004) Real time QTL of complex phenotypes of tomato interspecific introgression lines. Trends Plant Sci 9:107–109

Haanstra JPW, Wye C, Verbaked H, Meijer-Dekens F, Berg Pvd, Odinot P, Heusden AWv, Tanksley S, Lindhout P, Peleman J (1999) An integrated high-density RFLP-AFLP map of tomato based on two *Lycopersicon esculentum* × *L. pennellii* F_2 population. Theor Appl Genet 99:254–271

Hanson PM, Bernacchi D, Green S, Tanksley SD, Muniy-Appa V (2000) Mapping a wild tomato introgression associated with tomato yellow leaf curl virus resistance in a cultivated tomato line. J Am Soc Hort Sci 125:15–20

He C, Poysa V, Yu K (2003) Development and characterization of simple sequence repeat (SSR) markers and their use in determining relationships among *Lycopersicon esculentum* cultivars. Theor Appl Genet 106:363–373

Huang CC, Cui YY, Weng CR, Zabel P, Lindhout P (2000) Development of diagnostic PCR markers closely linked to the tomato powdery mildew resistance gene *Ol-1* on chromosome 6 of tomato. Theor Appl Genet 101:918–924

Isaacson T, Ronen G, Zamir D, Hirschberg J (2002) Cloning of tangerine from tomato reveals a carotenoid isomerase essential for the production of B-carotene and xanthophylls in plants. Plant Cell 14:333–342

Ito P, Currence TM (1964) A linkage test involving *c sp B+ md* in chromosome 6. Rep Tomato Genet Coop 14:14–15

Jeffreys AJ, Wilson V, Thein SL (1985) Hypervariable 'minisatellite' regions in human DNA. Nature 314:67–73

Jones DA, Dickinson MJ, Balint-Kurti PJ, Dixon MS, Jones JDG (1993) Two complex resistance loci revealed in tomato by classical and RFLP mapping of *Cf-2*, *Cf-4*, *Cf-5*, and *Cf-9* genes for resistance to *Cladosporium fulvum*. Mol Plant Microbe Interact 6:348–357

Kabelka E, Franchino B, Francis DM (2002) Two loci from *Lycopersicon hirsutum* LA407 confer resistance to strains of *Clavibacter michiganensis* subsp. *michiganensis* . Phytopathology 92:504–510

Kaloshian I, Yaghoobi J, Liharska T, Hontelez J, Hanson D, Hogan P, Jesse T, Wijbrandi J, Simons G, Vos P, Zabel P, Williamson VM (1998) Genetic and physical localization of the root-knot nematode resistance locus *Mi* in tomato. Mol Gen Genet 257:376–385

Kawchuk LM, Hachey J, Lynch DR (1998) Development of sequence characterized DNA markers linked to a dominant verticillium wilt resistance gene in tomato. Genome 41:91–95

Kawchuk LM, Hachey J, Lynch DR, Kulcsar F, van Rooijen G, Waterer DR, Robertson A, Kokko E, Byers R, Howard RJ, Fischer R, Prufer D (2001) Tomato Ve disease resistance genes encode cell surface-like receptors. Proc Natl Acad Sci USA 98:6511–6515

Kinzer SM, Schwager SJ, Mutschler MA (1990) Mapping of ripening-related or – specific cDNA clones of tomato (*Lycopersicon esculentum*). Theor Appl Genet 79:489–496

Knapp S, Bohs L, Nee M, Spooner DM (2004) Solanaceae – a model for linking genomics with biodiversity. Comp Funct Genom 5:285–291

Kohler GR, Clair DA St (2005) Variation for resistance to aphids (Homoptera: Aphididae) among tomato inbred backcross lines derived from wild *Lycopersicon* species. J Econ Entomol 98:988–995

Konieczny A, Ausubel FA (1993) A procedure for mapping *Arabidopsis* mutations using co-dominant ecotype-specific PCR-based markers. Plant J 4:403–410

Ku H-M, Doganlar S, Chen K-Y, Tanksley SD (1999) The genetic basis of pear-shaped tomato fruit. Theor Appl Genet 99:844–850

Ku H-M, Grandillo S, Tanksley S (2000) *fs8.1*, a major QTL, sets the pattern of tomto carpel shape well before anthesis. Theor Appl Genet 101:873–878

Labate JA, Baldo AM (2005) Tomato SNP discovery by EST mining and resequencing. Mol Breed 16:343–349

Lanahan MB, Yen HC, Giovannoni JJ, Klee HJ (1994) The *Never Ripe* mutation blocks ethylene perception in tomato. Plant Cell 6:521–530

Landegren U, Nilsson M, Kwok PY (1998) Reading bits of genetic information: Methods for single-nucleotide polymorphism analysis. Genome Res 8:769–776

Langella A, Ercolano MR, Monti lM, Frusciante l, Barone A (2004) Molecular marker assisted transfer of resistance to TSWV in tomato elite lines. J Hort Sci Biotechnol 79:806–810

Lauge R, Dmitriev AP, Joosten MHA, De Wit PJGM (1998) Additional resistance gene(s) against *Cladosporium fulvum* present on the *Cf-9* introgression segment are associated with strong PR protein accumulation. Mol Plant Microbe Interact 11:301–308

Lawson DM, Lunde CF, Mutschler MA (1997) Marker-assisted transfer of acylsugar-mediated pest resistance from the wild tomato, *Lycopersicon pennellii*, to the cultivated tomato, *Lycopersicon esculentum*. Mol Breed 3:307–317

Lecomte L, Duffé P, Buret M, Servin B, Hospital F, Causse M (2004) Marker-assisted introgression of five QTLs controlling fruit quality traits into three tomato lines revealed interactions between QTLs and genetic backgrounds. Theor Appl Genet 109:658–668

Levesque H, Vedel F, Mathieu C, de Courcel AGL (1990) Identification of a short rDNA spacer sequence highly specific of a tomato line containing *Tm-1* gene introgressed from *Lycopersicon hirsutum*. Theor Appl Genet 80:602–608

Levin I, Gilboa N, Shen S, Schaffer AA (2000) *Fgr*, a major locus that modulates the fructose to glucose ratio in mature fruits. Theor Appl Genet 100:256–262

Ling H-Q, Koch G, Bâumlein H, Ganal MW (1999) Map-based cloning of *chloronerva*, a gene involved in iron uptake of higher plants encoding nicotianamine synthase. Proc Natl Acad Sci USA 96: 7098–7130

Linnaeus C (1753) Species Planatarium, 1st ed edn. Holmiae, Stockholm, Sweden

Lippman Z, Tanksley SD (2001) Dissecting the genetic pathway to extreme fruit size in tomato using a cross between the small-fruited wild species *Lycopersicon pimpinellifolium* and *L. esculentum* var. Giant Heirloom. Genetics 158:413–422

Liu JP, Cong B, Tanksley SD (2003a) Generation and analysis of an artificial gene dosage series in tomato to study the mechanisms by which the cloned quantiative trait locus *fw2.2* controls fruit size. Plant Physiol 132:292–299

Liu JP, Van Eck J, Cong B, Tanksley SD (2002) A new class of regulatory genes underlying the cause of pear-shaped tomato fruit. Proc Natl Acad Sci USA 99:13302–13306

Liu Y-S, Gur A, Ronen G, Causse M, Damidaux R, Duffe P, Buret M, Hirschberg J, Zamir D (2003b) There is more to tomato fruit colour than candidate carotenoid genes. Plant Biotech J 1:195–207

MacArthur JW (1934) Linkage groups in the tomato. J Genet 29:123–133

Mangin B, Thoquet P, Grimsley NH (1999) Temporal and multiple quantitative trait loci analyses of resistance to bacterial wilt in tomato permit the resolution of linked loci. Genetics 151:1165–1172

Mao L, Begum D, Chuang H-W, Budlman MA, Szymkowlak EJ, Irish EE, Wing RA (2000) *JOINTLESS* is a MAD-box gene controlling tomato flower absicission zone development. Nature 406:910–913

Mao L, Begum D, Goff SA, Wing RA (2001) Sequence and analysis of the tomato *JOINTLESS* locus1. Plant Physiol 126:1331–1340

Martin B, Nienhuis J, King G (1989) Restriction fragment length polymorphisms associated with water use efficiency in tomato. Science 243:1725–1728

Martin GB, Brommonschenkel SH, Chunwongse J, Frary A, Ganal MW, Spivey R, Wu T, Earle ED, Tanksley SD (1993a) Map-based cloning of a protein kinase gene conferring disease resistance in tomato. Science 262:1432–1436

Martin GB, Frary A, Wu T, Brommonschenkel S, Chunwongse J (1994) A member of the *Pto* gene family confers sensitivity to fenthion resulting in rapid cell death. Plant Cell 6:1543–1552

Martin GB, Vicente MCd, Tanksley SD (1993b) High-resolution linkage analysis and physical chraracterization of the *Pto* bacterial resistance locus in tomato. Mol Plant Microbe Interact 6:26–34

McCormick S, Niedermeyer J, Fry J, Barnason A, Worsch R, Fraley R (1986) Leaf disk transformation of cultivated tomato (*L. esculentum*) using *Agrobacterium tumifaciens*. Plant Cell Rep 5:81–84

Medina-Filho H (1980) Linkage of *Aps-1*, *Mi* and other markers on chromosome 6. Rep Tomato Genet Coop 30:26–28

Mesbah LA, Kneppers RJA, Takken FLW, Laurent P, Hille J, Nijkamp HJJ (1999) Genetic and physical analysis of a YAC contig spanning the fungal disease resistance locus *Asc* of tomato (*Lycopersicon esculentum*). Mol Gen Genet 261:50–57

Miller P (1754) The gardeners dictionary, 4th ed, London, UK
Miller JC, Tanksley SD (1990) RFLP analysis of phylogenetic relationships and genetic variation in the genus *Lycopersicon*. Theor Appl Genet 80:437–448
Monforte AJ, Tanksley S (2000a) Fine mapping of a quantitative trait locus (QTL) from *Lycopersicon hirsutum* chromosome 1 affecting fruit characteristics and agronomic traits: breaking linkage among QTLs affecting different traits and dissection of heterosis for yield. Theor Appl Genet 100:471–479
Monforte AJ, Tanksley SD (2000b) Development of a set of near isogenic and backcross recombinant inbred lines containing most of the *Lycopersicon hirsutum* genome in a *L. esculentum* genetic background: a tool for gene mapping and gene discovery. Genome 43:803–813
Moore S, Vrebalov J, Payton P, Giovannoni J (2002) Use of genomics tools to isolate key ripening genes and analyse fruit maturation in tomato. J Exp Bot 53:2023–2030
Moreau P, Thoquet P, Olivier J, Laterrot H, Grimsley NH (1998) Genetic mapping of *Ph-2*, a single locus controlling partial resistance to *Phytophthora infestans* in tomato. Mol Plant Microbe Interact 11:259–268
Mustilli AC, Fenzi F, Ciliento R, Alfano F, Bowler C (1999) Phenotype of the tomato *high pigment-2* mutant is caused by a mutation in the tomato homolog of DEETIOLATED1. Plant Cell 11:145–157
Mutschler MA, Doerge RW, Liu SC, Kuai JP, Liedl BE, Shapiro JA (1996) QTL analysis of pest resistance in the wild tomato *Lycopersicon pennellii*: QTLs controlling acylsugar level and composition. Theor Appl Genet 92:709–718
Ohmori T, Murata M, Motoyoshi F (1996) Molecular characterization of RAPD and SCAR markers linked to the *Tm-1* locus in tomato. Theor Appl Genet 92:151–156
Paran I, Michelmore RW (1993) Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce. Theor Appl Genet 85:985–993
Paran I, Zamir D (1997) QTL analysis of morpholocial traits in a tomato recombinant inbred line population. Genome 40:242–248
Paran I, Goldman I, Tanksley SD, Zamir D (1995) Recombinant inbred lines for genetic mapping in tomato. Theor Appl Genet 90:542–548
Parrella G, Ruffel S, Moretti A, Morel C, Palloix A, Caranta C (2002) Recessive resistance genes against potyviruses are localized in colinear genomic regions of the tomato (*Lycopersicon* spp.) and pepper (*Capsicum* spp.) genomes. Theor Appl Genet 105:855–861
Paterson AH, Lander ES, Hewitt JD, Peterson S, Lincoln SE, Tanksley SD (1988) Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature 335:721–726
Paterson AH, Damon S, Hewitt JD, Zamir D, Rabinowitch HD, Lincoln SE, Lander ES, Tanksley SD (1991) Mendelian factors underlying quantitative traits in tomato: comparison across species, generations, and environments. Genetics 127:181–197
Peters JL, Széll M, Kendrick RE (1998) The expression of light-regulated genes in the high-pigment-1 muatnt of tomato. Plant Physiol 117:797–807
Pierce LC (1971) Linkage test with *Ph* conditioning resistance to race 0. Rep Tomato Genet Coop 21:30
Pillen K, Pineda O, Lewis C, Tanksley SD (1996) Status of genome mapping tools in the taxon *Solanaceae*. In: Paterson A (ed) Genome mapping in plants. RG Landes, Austin, TX, pp 281–308
Pnueli L, Carmel-Goren L, Hareven D, Gutfinger T, Alvarez J, Ganal J, Zamir D, Lifschitz E (1998) The *SELF-PRUNING* gene of tomato regulates vegetative to reproductive switching of sympodial meristems and is the ortholog of *CEN* and *TFl1*. Development 125:1979–1989
Rick CM (1973) Potential genetic resources in tomato species: clues from observation in native habitats. In: Srb AM (ed) Genes, enzymes, and populations. Plenum Press, New York, pp 255–269
Rick CM (1975) The tomato. In: King RC (ed) Handbook of genetics. Plenum Press, New York, pp 247–280
Rick CM (1976a) Natural variability in wild species of *Lycopersicon* and its bearing on tomato breeding. Genet Agraria 30:249–259
Rick CM (1976b) Tomato, *Lycopersicon esculentum* (Solanaceae). In: Simmonds NW (ed) Evolution of crop plants. Longman, London, pp 268–273
Rick CM (1978) The Tomato. Sci Am 23:76–87

Rick CM (1979a) Biosystematic studies in *Lycopersicon* and closely related species of *Solanum*. In: Hawkes JC, Lester RN, Skelding AD (eds) The biology and taxnomy of the solanaceae. Academic Press, New York, pp 667–678

Rick CM (1979b) Potential improvement of tomatoes by controlled introgression of genes from wild speies. Proceedings of the Conference of Broadening Genetic Base of Crops. Pudoc, Wageningen, pp 167–173

Rick CM (1980) Tomato. Hybridization of crop plants. Am Soc Agron/Crop Sci Soc Am, Madison, WI, pp 669–680

Rick CM (1991) Tomato paste: a concentrated review of genetic highlights from the beginnings to the advent of molecular genetics. Genetics 128:1–5

Rick CM, Fobes JF (1974) Association of an allozyme with nematode resistance. Rep Tomato Genetic Coop 24:25

Rick CM, Fobes JF (1975) Allozyme variation in the cultivated tomato and closely related species. Bul Torrey Bot Club 102:376–384

Rick CM, Gill BS (1973) Reproductive errors in aneuploids: generation of variant extra-chromosomal types by tomato primary trisomics. Can J Genet Cytol 15:299–308

Rivers BA, Bernatzky R, Robinson SJ, Jahnen-Dechent W (1993) Molecular diversity at the self-incompatibility locus is a salient feature in natural populations of wild tomato (*Lycopersicon peruvianum*). Mol Gen Genet 238:419–427

Robert VJM, West MAL, Inai S, Caines A, Arntzen L, Smith JK, Clair DA St (2001) Marker-assisted introgression of blackmold QTL alleles from wild *Lycopersicon cheesmanii* to cultivated tomato (*L. esculentum*) and evaluation of QTL phenotypic effects. Mol Breed 8:217–233

Ronen G, Cohen M, Zamir D, Hirschberg J (1999) Regulation of carotenoid biosyntheis during tomato fruit development: expression of the gene for lycopene epsilon-cyclase is down regulated during ripening and is elevated in the mutant *Delta*. Plant J 17:341–351

Ronen G, Carmel-Goren L, Zamir D, Hirschberg J (2000) An alternative pathway to beta-carotene formation in plant chloroplast discovered by map-based cloning of Beta and old-gold color muation in tomato. Proc Natl Acad Sci USA 97:11102–11107

Rossi M, Goggin FL, Milligan SB, Kaloshian I, Ullman DE, Williamson VM (1998) The nematode resistance gene *Mi* of tomato confers resistance against the potato aphid. Proc Natl Acad Sci USA 95:9750–9754

Rousseaux MC, Jones CM, Adams D, Chetelat RT, Bennett AB, Powell AA (2005) QTL analysis of fruit antioxidants in tomato using *Lycopersicon pennellii* introgression lines. Theor Appl Genet 111:1396–1408

Saliba-Colombani V, Causse M, Gervais L, Philouze J (2000) Efficiency of RFLP, RAPD, and AFLP markers for the construction of an intraspecific map of the tomato genome. Genome 43:29–40

Saliba-Colombani V, Causse M, Langlois D, Philouze J, Buret M (2001) Genetic analysis of organoleptic quality in fresh market tomatol. 1. Mapping QTLs for physical and chemical traits. Theor Appl Genet 102:259–272

Salmeron JM, Oldroyd GED, Rommens CMT, Scofield SR, Kim HS, Lavelle DT, Dahlbeck D, Staskawicz BJ (1996) Tomato *Prf* is a member of the leucine-rich repeat class of plant disease resistance genes and lies embedded withing the *Pto* kinase gene cluster. Cell 86:123–133

Sandbrink JM, van-Ooijen JW, Purimahua CC, Vrielink M, Verkerk R, Zabel P, Lindhout P (1995) Localization of genes for bacterial cancer resistance in *Lycopersicon peruvianum* using RFLPs. Theor Appl Genet 90:444–450

Sarfatti M, Abu-Abied M, Katan J, Zamir D (1991) RFLP mapping of *I1*, a new locus in tomato conferring resistance against *Fusarium oxysporum* f. sp. *lycopersici* race 1. Theor Appl Genet 80:22–26

Schornack S, Ballvora A, Gürlebeck D, Peart J, Ganal M, Baker B, Bonas U, Lahaye T (2004) The tomato resistance protein Bs4 is a predicted non-nuclear TIR-NB-LRR protein that mediates defense responses to severely truncated derivatives of AvrBs4 and overexpressed AvrBs3. Plant J 37:46–60

Scott JW, Agrama HA, Jones JP (2004) RFLP-based analysis of recombination among resistance genes to fusarium wilt races 1, 2 and 3 in tomato. J Am Soc Hort Sci 129:394–400

Simons G, Groenendijk J, Wijbrandi J, Reijans M (1998) Dissection of the Fusarium *I2* gene cluster in tomato reveals six homologs and one active gene copy. Plant Cell 10:1055–1066

Slater A, Maunders MJ, Edwards K, Schuch W, Grierson D (1985) Isolation and characterization of cDNA clones for tomato polygalacturonase and other ripening-related proteins. Plant Mol Biol 5:137–147

Spooner DM, Anderson GJ, Jansen RK (1993) Chloroplast DNA evidence for the interrelationships of tomatoes, potatoes and pepinos (Solanaceae). Am J Bot 80:676–688

Stamova BS, Chetelat RT (2000) Inheritance and genetic mapping of cucumber mosaic virus resistance introgressed from *Lycopersicon chilense* into tomato. Theor Appl Genet 101:527–537

Stevens MA, Rick CM (1986) Genetics and breeding. In: Atherton JG, Rudich J (eds) The tomato crop. Chapman and Hall, New York, pp 35–109

Stevens MR, Lamb EM, Rhoads DD (1995) Mapping the *Sw-5* locus for tomato spotted wild virus resistance in tomatoes using RAPD and RFLP analyses. Theor Appl Genet 90:451–456

Stommel JR, Zhang YP (1998) Molecular markers linked to quantitative trait loci for anthracnose resistance in tomato (Abstract). Hort Science 33:514

Suliman-pollatschek S, Kashkush K, Shats H, Hillel J, Lavi U (2002) Generation and mapping of AFLP, SSRs, and SNPs in *Lycopersicon esculentum*. Cell Mol Biol Letts 7:583–597

Tanksley SD (1993) Linkage map of the tomato (*Lycopersicon esculentum*) (2N = 24). In: O'Brian SJ (ed) Genetic maps: locus maps of complex genomes, 6th edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp 6.3–6.15

Tanksley SD (2004) The genetic, developmental, and molecular bases of fruit size and shape variation in tomato. Plant Cell 16:S181–S189

Tanksley SD, Loaiza-Figueroa F (1985) Gametophytic self-incompatibility is controlled by a single major locus on chromosome 1 in *Lycopersicon peruvianum*. Proc Natl Acad Sci USA 82:5093–5096

Tanksley SD, Medina-Filho H, Rick CM (1982) Use of naturally-occuring enzyme variation to detect and map genes controlling quantitative traits in an interspecific cross of tomato. Heredity 49:11–25

Tanksley SD, Rick CM (1980) Isozyme gene linkage map of tomato: applications in genetics and breeding. Theor Appl Genet 57:161–170

Tanksley SD, Rick CM, Vallejos CE (1984) Tight linkage between a nuclear male-sterile locus and an enzyme marker in tomato. Theor Appl Genet 68:109–113

Tanksley SD, Ganal MW, Prince JP, de-Vicente MC, Bonierbale MW, Broun P, Fulton TM, Giovannoni JJ, Grandillo S, Martin GB, Messeguer R, Miller JC, Miller L, Paterson AH, Pineda O, Roder MS, Wing RA, Wu W, Young ND (1992) High density molecular linkage maps of the tomato and potato genomes. Genetics 132:1141–1160

Tanksley SD, Grandillo S, Fulton TM, Zamir D, Eshed Y, Petiard V, Lopez J, Beck-Bunn T (1996) Advanced backcross QTL analysis in a cross between an elite processing line of tomato and its wild relative *L. pimpinellifolium*. Theor Appl Genet 92:213–224

Thompson AE, Tomes ML, Erickson HT, Wann EV, Armstrong RJ (1967) Inheritance of crimson fruit color in tomatoes. Proc Am Soc Hort Sci 91:495–504

Thoquet P, Olivier J, Sperisen C, Rogowsky P, Laterrot H, Grimsley NH (1996) Quantitative trati loci determining resistance to bacterial wilt in tomato cultivar Hawaii 7996. Mol Plant Microbe Interact 9:826–836

Tigchelaar EC (1986) Tomato breeding. In: Bassett MJ (ed) Breeding for vegetable crops. AVI Publishing Co., Westport, CT, pp 135–171

Tør M, Manning K, King GJ, Thompson AJ, Jones GH, Seymour GB, Armstrong SJ (2002) Genetic analysis and FISH mapping of the *Clourless non-ripening* locus of tomato. Theor Appl Genet 104:165–170

Truco MJ, Randall LB, Bloom AJ, Clair DA St (2000) Detection of QTL associated with shoot wilting and root ammonium uptake under chilling temperatures in an interspecific backcross population from *Lycopersicon esculentum* × *L. hirsutum* . Theor Appl Genet 101:1082–1092

USDA (2005) Agricultural statistics 2005. United State Department of Agtriculture, National Agricultural Staticstics Service

Vakalounakis DJ, Laterrot H, Moretti A, Ligoxigakis EK, Smardas K (1997) Linkage between *Fr1* (*Fusarium oxysporum* f. sp. *radicis-lycopersici* resistance) and *Tm-2* (tobacco mosaic virus resistance-2) loci in tomato (*Lycopersicon esculentum*). Ann Appl Biol 130:319–323

Vallejos CE, Tanksley SD (1983) Segregation of isozyme markers and cold tolerance in an interspecific backcross of tomato. Theor Appl Genet 66:241–247

van der Biezen EA, Glagotskaya T, Overduin B, Nijkamp HJJ, Hille J (1995) Inheritance and genetic mapping of resistance to *Alternaria alternata* f. sp. *lycopersici* in *Lycopersicon pennellii* . Mol Gen Genet 247:453–461

van der Hoeven R, Ronning C, Giovannoni J, Martin G, Tanksley S (2002) Deductions about the number, organization, and evolution of genes in the tomato genome based on analysis of a large expressed sequence tag collection and selective genomic sequencing. Plant Cell 14:1441–1456

van der Knaap E, Tanksley SD (2003) The making of a bell pepper-shaped tomato fruit: identification of loci controlling fruit morphology in yellow stuffer tomato. Theor Appl Genet 107:139–147

van der Knaap E, Lippman ZB, Tanksley SD (2000) Extremely elongated tomato fruit controlled by four quantitative trait loci with epistatic interactions. Theor Appl Genet 104:241–247

van der Knaap E, Sanyal A, Jackson SA, Tanksley SD (2004) High-resolution fine mapping and fluorescence in situ hybridization analysis of sun, a locus controlling tomato fruit shape, reveals a region of the tomato genome prone to DNA rearrangements. Genetics 168:2127–2140

van Heusden AW, Koornneef M, Voorrips RE, Brüggemann W, Pet G, Vrielink-van Ginkel R, Chen X, Lindhout P (1999) Three QTLs from *Lycopersicon peruvianum* confer a high level of resistance to *Clavibacter michiganensis* ssp. *michiganensis*. Theor Appl Genet 99:1068–1074

van Ooijen JW, Sandbrink JM, Vrielink M, Verkerk R, Zabel P, Lindhout P (1994) An RFLP linkage map of *Lycopersicon peruvianum*. Theor Appl Genet 89:1007–1013

van Tuinen A, Cordommier-Pratt M-M, Pratt LH, Verkerk R, Zabel P, Koornneef M (1997) The mapping of phytochrome genes and photomorphogenic mutants of tomato. Theor Appl Genet 94:115–122

Veremis JC, van Heusden AW, Roberts PA (1999) Mapping a novel heat-stable resistance to *Meloidogyne* in *Lycopersicon peruvianum*. Theor Appl Genet 98:274–280

Villalta I, Reina-Sånchez A, Cuartero J, Carbonell EA, Asins MJ (2005) Comparative microsatellite linkage analysis and genetic structure of two populations of F_6 lines derived from *Lycopersicon pimpinellifolium* and *L. cheesmanii* . Theor Appl Genet 110:881–894

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucl Acids Res 23:4407–4414

Vos P, Simons G, Jesse T, Wijbrandi J, Heinen L, Hogers R, Frijters A, Groenendijk J, Diergaarde P, Reijans M, Fierens-Onstenk J, de Both M, Peleman J, Liharska T, Hontelez J, Zabeau M (1998) The tomato *Mi-1* gene confers resistance to both root-knot nematodes and potato aphids. Nat Biotechnol 16:1365–1369

Vrebalov J, Ruezinsky DM, Padmanabhan V, White R, Medrano D, Drake R, Schuch W, Giovannoni J (2002) A MADS-box gene necessary for fruit ripening at the tomato *ripening-inhibitor* (*rin*) locus. Science 296:343

Wang Y, Van der Hoeven R, Nielsen R, Mueller LA, Giovannoni J, Tanksley SD (2005) Characteristics of the tomato nuclear genome as determined by sequencing unmethylated *Eco*RI digest fragment. Theor Appl Genet 112:72–84

Wann EV, Jourdain EL (1985) Effects of mutant genotypes *hp* og^c and *dg* og^c on tomato fruit quality. J Am Soc Hort Sci 110:212–215

Warnock SJ (1988) A review of taxonomy and phylogeny of the genus *Lycopersicon*. Hort Sci 23: 669–673

Warren GF (1998) Spectacular increases in crop yields in the twentieth century. Weed Technol 12:752–760

Wilkinson J, Lanahan MB, Yen H, Giovannoni J, Klee H (1995) An ethylene-induced component of signal transduction encoded by *Never-ripe*. Science 270:1807–1809

Williams JGK, Kubelik AE, Levak KJ, Rafalski JA, Tingey SC (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucl Acids Res 18:6531–6535

Wing RA, Zhang HB, Tanksley SD (1994) Map-based cloning in crop plants. Tomato as a model system: I. Genetic and physical mapping of *Jointless*. Mol Gen Genet 242:681–688

Yaghoobi J, Kaloshian I, Wen Y, Williamson VM (1995) Mapping of a new nematode resistance locus in *Lycopersicon peruvianum*. Theor Appl Genet 91:457–464

Yang W-Y, Francis DM (2005) Marker assisted selection for combining resistance to bacterial spot and bacterial speck in tomato. J Am Soc Hort Sci 130:716–721

Yang W-Y, Sacks EJ, Lewis Ivey ML, Miller SA, Francis DM (2005) Resistance in *Lycopersicon esculentum* intraspecific crosses to race T1 strains of *Xanthomonas campestris* pv. *vesicatoria* causing bacterial spot of tomato. Phytopathology 95:519–527

Yates HE, Frary A, Doganlar S, Frampton AJ, Eannetta NT, Uhlig J, Tanksley SD (2004) Comparative fine mapping of fruit quality QTLs on chromosome 4 introgression derived from two wild tomato species. Euphytica 135:283–296

Yen HC, Lee S, Tanksley SD, Lanahan MB, Klee HJ, Giovannoni JJ (1995) The tomato *Never-ripe* locus regulates ethylene-inducible gene experession and is linked to a homologue of the *Arabidopsis ETR1* gene. Plant Physiol 107:1343–1353

Yen HC, Shelton BA, Howard LR, Lee S, Vrebalov J, Giovannoni JJ (1997) The tomato *high-pigment* (*hp*) locus maps to chromosome 2 and influences plastome copy number and fruit quality. Theor Appl Genet 95:1069–1079

Yu ZH, Wang JF, Stall RE, Vallejos CE (1995) Genomic localization of tomato genes that control a hypersensitive reaction to *Xanthomonas campestris* pv. *vesicatoria* (Doidge) dye. Genetics 141:675–682

Zamir D, Ekstein-Michelson I, Zakay U, Navot N, Zeidan M, Sarfatti M, Eshed Y, Harel E, Pleban T, van-Oss H, Kedar N, Rabinowitch HD, Czosnek H (1994) Mapping and introgression of a tomato yellow leaf curl virus tolerance gene, *Ty-1*. Theor Appl Genet 88:141–146

Zamir D, Tal M (1987) Genetic analysis of sodium, potassium and chloride ion content in *Lycopersicon*. Euphytica 36:187–191

Zhang H-B, Budiman MA, Wing RA (2000) Genetic mapping of *jointless-2* to tomato chromosome 12 using RFLP and RAPD markers. Theor Appl Genet 100:1183–1189

Zhang H-B, Martin GB, Tanksley SD, Wing RA (1994) Map-based cloning in crop plants. Tomato as a model system: II. Isolation and characterization of a set of overlapping artificial chromosomes encompassing the *Jointless* locus. Mol Gen Genet 244:613–621

Zhang L, Lin GY, Niño-Liu DO, Foolad MR (2003) Mapping QTLs conferring early blight (*Alternaria solani*) resistance in a *Lycopersicon esculentum* × *L. hirsutum* cross by selective genotyping. Mol Breed 12:3–19

Zhang LP, Khan A, Niño-Liu D, Foolad MR (2002) A molecular linkage map of tomato displaying chromosomal locations of resistance gene analogs based on a *Lycopersicon esculentum* × *L. hirsutum* cross. Genome 45:133–146

Zhang Y, Stommel JR (2000) RAPD and AFLP tagging and mapping of Beta (*B*) and Beta modifier (Mo_B), two genes which influence beta-carotene accumulation in fruit of tomato (*Lycopersicon esculentum* Mill). Theor Appl Genet 100:368–375

Zhang Y-P, Stommel JR (2001) Development of SCAR and CAPS markers linked to the *Beta* gene in tomato. Crop Sci 41:1602–1608

CHAPTER 14

GENOMICS FOR IMPROVEMENT OF ROSACEAE TEMPERATE TREE FRUIT

PERE ARÚS[1,*] AND SUSAN GARDINER[2]

[1]*IRTA (Institut de Recerca i Tecnologia Agroalimentàries), Carretera de Cabrils Km.2, E-08348 Cabrils, Spain*

[2]*HortResearch, Private Bag 11030, Palmerston North, New Zealand*

Abstract: Genomic studies of Rosaceous fruit trees have concentrated on two species: peach (*Prunus persica*), which has served as a model for other species of the same genus, such as the stone fruits (apricot, cherry and plum) and almond; and apple (*Malus* x *domestica*), which itself is a model for other close species such as pear, quince and loquat. High density or saturated maps exist in both peach and apple, and sets of microsatellite markers spaced across the genome of both species are used for gene tagging and mapping in other populations. Efficient methods for mapping new markers and genes have been developed, such as "bin mapping" and the "genome scanning approach". Tens of major genes and QTLs have been located on the maps of both species, and some of them are close to markers routinely used for selection in plant breeding. Comparative mapping has shown that all members of the *Prunus* genus share the same genome structure and that apple and pear genomes have a highly similar genetic organization. There are chromosomal rearrangements between the genomes of apple and *Prunus*, but extensive regions of synteny and collinearity are maintained. Several genes of apple and peach have been cloned using map-based techniques or are in the process of being cloned. A physical map is in an advanced stage of construction for peach and one has recently been started in apple. Large EST collections have been developed, particularly in apple and *Prunus* providing tens of thousands of new markers and gene sequences useful for functional analysis and map construction. Microarrays are proving to be valuable tools for identifying candidate genes for characters of interest. This information is stored in several databases with varying degrees of public access.

1. INTRODUCTION

The Rosaceae are divided into four subfamilies. Two of them, Prunoideae and Maloideae, include the fruit tree crops that are the subject of this chapter. The other two subfamilies are the Rosoideae, with rose (*Rosa*), strawberry (*Fragaria*),

*Author for correspondence: Pere.Arus@irta.es

R.K. Varshney and R. Tuberosa (eds.), Genomics Assisted Crop Improvement: Vol. 2: Genomics Applications in Crops, 357–397.

blackberry and raspberry (*Rubus*) as the most important crops, and the Spiraeoideae, which has some ornamental trees and bushes of minor economic importance. The Pruniodeae are characterized by their fruit, the drupe, where the fleshy mesocarp surrounds a stony endocarp usually protecting a single large seed. All cultivated members of the Prunoideae subfamily belong to the genus *Prunus*, which encompasses almond (*P. dulcis*) and all stone fruits: peach (*P. persica*), apricot (*P. armeniaca*), European (*P. domestica*) and Japanese (*P. salicina*) plum, and sweet (*P. avium*) and sour (*P. cerasus*) cherry. The characteristic fruit of the Maloideae is the pome, formed by a fleshy hypanthyum tissue surrounding the smaller ovary core where the seeds are located. The main cultivated species of this subfamily are apple (*Malus* x *domestica*), pear (*Pyrus communis* and *P. pyrifolia*), quince (*Cydonia oblonga*) and loquat (*Eriobotrya japonica*).

All *Prunus* species have the same basic chromosome number ($2n = 2x = 8$) and are diploids, except *P. domestica* (6x) and *P. cerasus* (4x). Their genome size is small, of approximately 300 Mbp for the diploid species (Arumuganathan and Earle 1991). The Maloideae have a tetraploid genome ($2n = 34$), thought to contain either a genome of the Spiraeoideae ($x = 9$) and one of the Prunoideae ($x = 8$), or two different genomes of the Spiraeoideae with further loss of one chromosome pair (Evans and Campbell 2002). The size of the apple and pear genomes is 600–800 Mbp (Arumuganathan and Earle 1991), approximately twice that of *Prunus*, suggesting that the two constituent genomes are small and of similar size to that of *Prunus*. In this chapter, we will review the progress in genomics of this group of crops, which has rapidly advanced over the last decade and has already resulted in relevant breeding applications.

2. MAPPING IN ROSACEOUS FRUIT TREES

2.1. *Prunus*

One of the first maps to be constructed with RAPD (random amplified polymorphic DNA) markers in plants was that of peach (Chaparro et al. 1994). RAPDs were the first markers based on the polymerase chain reaction (PCR) and their use opened a new era of map construction, where maps could be obtained with moderate investment of both time and money. RAPDs were useful but proved to be difficult to reproduce, and more robust markers have been used to construct maps since then. The first saturated map constructed with transferable markers, mostly restriction fragment length polymorphisms (RFLPs) plus a few isozyme genes, was based on an F2 almond ('Texas') x peach ('Earlygold') (TxE) population (Joobeur et al. 1998). The TxE map has since been improved (Aranzana et al. 2003; Dirlewanger et al. 2004a; Howad et al. 2005). It now has 826 markers: 449 SSRs (simple sequence repeat or microsatellite markers), 361 RFLPs, 11 isozymes and 5 STSs (sequence tagged sites), detecting the expected eight linkage groups (G1 to G8), spanning a total distance of 524 cM (average density 0.63 cM/marker) and with a maximum gap of 7 cM. This map, and most maps involving *Prunus* species, can

be viewed and compared with other maps at the Genetic Database for Rosaceae (GDR; http://www.mainlab.clemson.edu/gdr/).

The TxE map has been used by *Prunus* geneticists as the reference map to establish linkage group terminology and orientation. Many of the markers used for map construction in other populations have also been selected from TxE to allow good coverage of the *Prunus* chromosomes. Sixteen of these maps, with at least 28 anchor markers with TxE, have already been published (Arús et al. 2005a). A "consensus" map, containing the markers and major genes that have been mapped in different segregating populations, with TxE as the backbone, has been created, allowing the position of major genes and quantitative trait loci (QTLs) to be established on a single map. Twenty-eight major genes (Dirlewanger et al. 2004a) and 28 QTLs (Arús et al. 2005b), responsible for flower, fruit and leaf characteristics, plant architecture, disease resistance, and other features, have been located on this consensus map, allowing the identification of many transferable molecular markers (RFLPs, SSRs, etc.) in the genomic regions where these genes or QTLs are found.

A recent development in *Prunus* is the use of the "bin mapping" strategy proposed by Vision et al. (2000) to place new markers on the TxE map, with reasonable precision, using only a few plants. A set of six plants of the TxE population were selected by Howad et al. (2005), which allowed mapping of any marker to 67 unique chromosome fragments (bins) of 7.8 cM average size, covering the entire map length. Using this approach, Howad et al. (2005) mapped 264 new SSRs with about 1/14 of the effort needed to map all the 82 plants of the population and with only a moderate loss of accuracy. The TxE high density reference map is based on a highly polymorphic interspecific F2 population, which makes bin mapping a very efficient strategy. An F2 population is easier to use and more suitable for mapping purposes than the F1 segregating populations, as it allows selection of a smaller set of plants than other population structures, e.g. a set of 14–20 plants of a backcross would be needed to achieve a similar accuracy to 6 plants of an F2. The implementation of bin mapping in apple, pear and *Prunus* F1 populations is clearly desirable, and possible, but it will be more complex as they include loci that segregate in all possible ratios, and require the construction of two maps, one for each parent of the cross.

2.2. *Malus*

Apple possesses a rich resource of simply inherited resistance genes with major effects. Molecular mapping in apple therefore initially focused on identifying genetic markers for pest and disease resistances (Durham et al. 1994; Manganaris et al. 1994) that could be used for marker assisted selection by breeders. The emphasis on pathogen resistance has continued, but is now coupled to construction of full genetic framework maps (Maliepaard et al. 1998; Liebhard et al. 2003b). Such maps with a high density of transferable markers are needed for construction of QTL maps of complex traits (Durel et al. 2003; Liebhard et al. 2003a; Liebhard et al. 2003c),

for whole genome scans using the Genome Scanning Approach (GSA) (Erdin et al. 2006; Patocchi et al. 2004; Patocchi et al. 2005) and for pedigree genotyping (van de Weg et al. 2004). The map positions of at least 59 major genes, including 24 major gene resistances and 25 genes potentially involved in fruit allergenicity, are summarized in a recent review (Gardiner et al. 2006b).

The earliest genetic maps in apple utilized non-transferable RAPD markers with a small number of transferable isoenzyme markers (Conner et al. 1997; Hemmat et al. 1994). The first apple reference map was developed in a 'Prima' x 'Fiesta' progeny and included a number of co-dominant transferable markers (microsatellites, isoenzymes and RFLPs as well as a small number of RAPDs and AFLPs (amplified fragment length polymorphisms) distributed over all 17 linkage groups (Maliepaard et al. 1998). This cross has been employed as the basis for maps of QTLs determining fruit texture (King et al. 2000; King et al. 2001) as well as resistance to apple scab (Durel et al. 2003) and fireblight (Calenge et al. 2005a).

The second reference genetic map, constructed in a 'Fiesta' x 'Discovery' progeny (Liebhard et al. 2003b) initially consisted of 129 SSRs plus 710 dominant AFLPs and RAPDs. This map has recently been enlarged to encompass a further 149 SSRs (Silfverberg-Dilworth et al. 2006), is the most detailed apple framework map available, and is the current reference for linkage group number (LG1-LG17) and orientation. The new markers enabled selection of an initial set of 86 highly polymorphic SSRs that will eventually allow genome wide genotyping of apple with 100 markers at an average distance of 15cM. The earlier version of this map was used as the framework to map both QTLs determining fireblight and scab resistance, and physiological traits (Calenge et al. 2005a; Calenge et al. 2005b; Liebhard et al. 2003a; Liebhard et al. 2003c), and 18 RGAs (resistance gene analogues) (Baldi et al. 2004) that are homologs of NBS/LRRs (nucleotide binding site/leucine rich repeat resistance genes). A map constructed in a 'Telamon' x 'Braeburn' progeny that shares some markers with the 'Fiesta' x 'Discovery' map has been used to map QTLs for fruit vitamin C content (Davey et al. 2006).

A third reference map, in 'Discovery' x TN10-8 has been used to locate 23 markers with homology to apple NBS/LRRs (Calenge et al. 2005b) in positions close to major scab and powdery mildew resistances, as well as QTLs for resistance to these diseases identified previously in the same progeny (Calenge et al. 2004a; Calenge et al. 2004b). It has also been used for more detailed mapping of QTLs for powdery mildew resistance (Calenge and Durel 2006) and to map markers for woolly apple aphid resistance *Er3* to LG8 (Gardiner et al. 2006b).

2.3. *Pyrus*

Genetic mapping in *Pyrus* was initiated later than in *Malus,* using a progeny resulting from a cross between a European (*P. communis*) and a Japanese pear (*P. pyrifolia*) (Yamamoto et al. 2004; Yamamoto et al. 2006) and the latest maps comprise 76 and 64 SSRs in 'Bartlett' (*P. communis*) and 'Hosui' (*P. pyrifolia*) respectively, as

well as several hundred AFLPs (Yamamoto et al. 2004). A recent conference report raises the number of SSRs mapped in these two varieties to 157 and 112 respectively (Yamamoto et al. 2006). A map of 'La France' (*P. communis*) has 176 SSRs, with the number of linkage groups consistent with the chromosome number of 17 and almost complete genome coverage (Yamamoto et al. 2006). The self incompatibility (SI) locus maps to LG17 in Japanese and European pears (Yamamoto et al. 2004; Yamamoto et al. 2002), and *Vnk* resistance to *Venturia nashicola* on LG1 (Terakami et al. 2006). Maps constructed in *P. communis* progenies of 'Passe Crassane' x 'Harrow Sweet' and 'Abbé Fetel' x 'Max Red Bartlett' are less complete, containing 41 and 31 SSRs respectively (Pierantoni et al. 2004). QTLs for resistance to fireblight have been identified in 'Harrow Sweet' (Dondini et al. 2004).

3. COMPARATIVE GENOMICS

3.1. Within the *Prunus* Genus

Prunus maps have been obtained with intra- and interspecific segregating populations, including individuals from seven species: peach, almond, apricot, cherry, *P. cerasifera*, *P. davidiana*, and *P. ferganenesis*. In all of them the maps obtained with common markers were essentially syntenic and collinear (Arús et al. 2005a). In the rare occasions where marker location was not that predicted, it is probable that they occurred in duplicated genome regions and only one of the two possible markers could be mapped. These results strongly suggest that the *Prunus* genome is in common over all the species of the genus, in agreement with the pattern of frequent inter-crossability and interspecific hybrid fertility known to occur between many species of this genus (Scorza and Sherman 1996).

Only two major chromosomal rearrangements, both reciprocal translocations, have been reported in *Prunus*. The first was found by Jáuregui et al. (2001) in the F2 of a cross between 'Garfi' almond and the red leaf peach rootstock 'Nemared'. The translocated chromosomal segments occurred in G6 and G8. This reciprocal translocation has been confirmed in the three-way cross *P. cerasifera* x GN22, where GN22 is a progeny of 'Garfi' x 'Nemared' (Dirlewanger et al. 2004b). The same translocation was found by Yamamoto et al. (2005) in the cross 'Akame' x 'Juseitou', where 'Akame' is also a red-leafed genotype. Based on the current information, it seems likely that this mutation occurs only within the group of red-leafed peaches. The fact that the breakpoint of the mutation was located at the same genomic region of the gene determining red leaf colour (*Gr*) suggests a relationship of cause and effect between these two characteristics. The other reciprocal translocation was reported by Lambert et al. (2004) in a genotype of myrobolan plum (*P. cerasifera*) and this affects G3 and G5. These results suggest that reciprocal translocations may occur and be maintained within a given *Prunus* species, characterizing monophyletic transects, although the majority of the individuals of the species are within the general *Prunus* genome configuration.

3.2. *Malus* and *Pyrus*

Comparative genome mapping between *Malus* and *Pyrus* has occurred simultaneously with the construction of *Pyrus* genetic linkage maps, as they have all included a proportion of microsatellite markers developed for apple map construction. The latest pear maps (Dondini et al. 2004; Pierantoni et al. 2004; Yamamoto et al. 2006) each contain between 31 and 79 apple SSRs, enabling alignment of all pear and apple linkage groups. This suggests that genome organization is conserved between apple and pear, and that more detailed comparative mapping could assist transfer of genomic information from resource-rich apple to the less well studied pear and other members of the Maloideae (Arús et al. 2005a). Since 2004, the linkage groups of pear have been assigned the same numbers as in apple and comparative studies have indicated that at least half the apple SSRs are polymorphic in pear (Hemmat et al. 2003; Pierantoni et al. 2004; Wang et al. 2005) and the reverse holds also true for pear SSRs (S. Gardiner and J-M Celton, unpublished).

3.3. *Prunus* and *Malus*

An incomplete comparison between *Malus* and *Prunus* is possible based on 30 common loci (24 RFLPs and 6 isozymes) between the TxE *Prunus* map and the 'Prima' x 'Fiesta' apple map (Dirlewanger et al. 2004a). Two *Prunus* linkage groups (G3 and G4) had three or more markers in common with the apple map, each one detecting a pair of homeologous apple groups with the same marker order. *Prunus* G1 had eight markers in common with apple. G1 is the longest and most populated with markers in TxE and most other *Prunus* maps, suggesting that it corresponds with chromosome 1, as this is clearly longer than the rest (Salesses and Mouras 1977). Such a long chromosome does not exist in apple (Bouvier et al. 2000) and four of the markers located in the upper part of G1 correspond to two homeologous groups of apple, and the four remaining markers map to another apple linkage group. These results indicate that the long *Prunus* chromosome may be split into two in apple or in the original species that formed the *Malus* amphydiploid. Unpublished results (E. van de Weg, pers. comm.) detect additional syntenic regions between apple and *Prunus*, which suggest an overall high similarity between these two genomes.

3.4. *Prunus* and *Pyrus*

Comparative studies between *Pyrus* and *Prunus* have been initiated. Seven out of 65 *Prunus* microsatellite markers screened segregated in *Pyrus* (Yamamoto et al. 2004; Yamamoto et al. 2002) compared with 14 out of 31 apple microsatellite markers, which is in line with the relative genetic distances of *Malus* and *Prunus* from *Pyrus*.

3.5. *Prunus* and *Arabidopsis*

The *Arabidopsis* sequence has been compared with the map positions of: a) 227 RFLP loci mapped in TxE, most of them detected with probes based on Rosaceae or *Arabidopsis* EST sequences (Dominguez et al. 2003), b) 475 peach ESTs anchored to the *Prunus* consensus map (Jung et al. 2006), c) 1097 peach ESTs physically located in 431 *Prunus* BAC contigs (Jung et al. 2006), d) the complete sequence of one, and the partial sequence of two, peach BAC clones (Georgi et al. 2003), and e) the complete sequence of one myrobolan plum (*P. cerasifera*) BAC clone (Claverie 2004). In all cases it was possible to find syntenic regions, but they were limited to small DNA fragments indicating an overall picture of fragmentary genome conservation. The largest *Prunus-Arabidopsis* conserved fragment was found by Dominguez et al. (2003), where a region of 25 cM in G2 of *Prunus* corresponded to one of 5.4 Mbp in chromosome 5 of *Arabidopsis*.

4. LARGE INSERT LIBRARIES AND PHYSICAL MAPS

BAC (bacterial artificial chromosome) libraries have been constructed in apple and *Prunus* as tools for physical mapping and map-based cloning. The first BAC library in peach was developed by Wang et al. (2001) using the Chinese cultivar Jingyu and consists of 20,736 clones of average size 95 kb and estimated coverage of 6.7x the peach haploid component. Georgi et al. (2002) produced a 'Nemared' library (44,160 clones; average size 50–70 kb; 8x genome coverage) that has been used for construction of the peach physical map (Zhebentyayeva et al. 2006), the physical map around the evergrowing gene (Bielenberg et al. 2004), and to find SSRs for specific genome regions (Georgi et al. 2002). Vilanova et al. (2003) developed the first BAC library in apricot (101,376 clones; average size 64 kb; 22x genome coverage), using the plum pox virus (sharka) resistant cultivar Goldrich, and employed it to find clones linked to the self-incompatibility locus. A library of myrobolan plum was constructed in the nematode resistant accession P.2175 by Claverie et al. (2004) for map-based cloning of the *Ma* gene for resistance to root-knot nematodes (*Meloidogyne* spp.). The library has 30,720 clones of average size 145 kb, with an estimated 14–15x coverage of the genome.

A *Prunus* physical map is under construction by a consortium of research groups led by Dr. Albert G. Abbott at Clemson University (South Carolina, USA), using BAC fingerprinting in two peach BAC libraries ('Nemared' and one more recently constructed in haploid 'Lovell'). The current physical map (April 2006) contains fingerprints of more than 18,387 BACs assembled into 1,367 contigs (including 11,193 BACs) and 7,194 singletons covering more than 60% of the *Prunus* genome (Zhebentyayeva et al. 2006). One of the most interesting features of this map is that it is anchored in the *Prunus* genetic map (Horn et al. 2005). The position of 290 genetically mapped markers was established on BACs used to construct the physical map, 150 of them using RFLP probes mapped in TxE covering the whole genome. Moreover, a transcript map was created by hybridizing more than 3,000

EST putative unigenes to the 'Nemared' BAC library. With this approach it has been possible to determine the genetic map position of ESTs that are either in the same BAC or in the same BAC contig as one of the markers of the genetic map. In total, 749 EST unigenes that connect the *Prunus* genetic and physical maps have been identified so far.

The first apple BAC library of 37,000 clones was developed in 'Florina', which carries the *Vf* resistance to scab. Genome coverage is 5 haploid genomes and average insert size is 120kb (Vinatzer et al. 1998). The original source of *Vf*, *M. floribunda* 821, was used to construct a library of 31,000 clones covering 5 haploid genomes with an average insert size of 125kb (Xu et al. 2001). 'Goldrush', also carrying *Vf*, was used to construct a 35,712 clone library of similar genome coverage and average insert size of 110kb with initial *Bam*HI digestion (Xu et al. 2002). For the purposes of developing a genome-wide physical map (Han et al. 2006a), a complementary 'Goldrush' BAC library has been developed using *Hin*dIII digestion. This library consists of 46,791 clones with an average genome size of 115kb and represents 8 haploid genome equivalents. Primers have been designed for 1,050 SSRs to develop markers for BAC fingerprinting. A 56,000 clone BAC library and a 168,000 clone cosmid library have been constructed in 'Pinkie', a selection containing *Vf* and the powdery mildew resistance gene *Pl2*. Each library should cover nearly seven haploid genome equivalents (E. Rikkerink, pers. comm.).

Limited physical maps around resistance genes that have been targets for map-based cloning have been published. These include contigs around *Vf* on LG1 (Patocchi et al. 1999b; Vinatzer et al. 2001b; Xu and Korban 2002; Xu et al. 2001) and *Sd1* on LG7 (Cevik and King 2002b) and *SBE* (starch branching enzyme) (Han et al. 2006b).

5. TRANSCRIPTOMICS

A range of complementary tools is available to determine functionality of candidate genes, ranging from annotation in EST (expressed sequence tag) databases to *in vivo* expression. Generally, a broad first cut is made using bioinformatics annotation of ESTs and microarray analysis, and subsequently the number of candidates for mapping to QTLs is narrowed down using RT-PCR and / or *in vivo* expression.

5.1. EST Collections

Several collections of ESTs have been developed by different research groups and in various fruit crops of this family. Some of them are public and others partially or completely private. The private EST collections are being progressively released to the public. The descriptions of the current EST collections have been divided into three parts, one concerning the publicly available data for the whole family, followed by the descriptions of EST developments of individual research groups in *Prunus* and in apple. The major EST sequencing projects usually involve the

development of databases for EST handling and analysis. These databases will be described in Section 6 of this chapter.

5.1.1. ESTs in public databases

More than 370,000 Rosaceae ESTs are currently available in public databases, most of them apple (~260,000) and *Prunus* (~85,000), mainly peach and apricot. The number of unigenes deduced from ~260,000 of these EST sequences, held in the Genome Database for Rosaceae (GDR) at a clustering threshold of 90%, is ~72,000 for the Rosaceae, ~68,000 in *Malus* and 13,000 in *Prunus*. ESTs are a source of high-quality molecular markers such as SSRs that are particularly abundant in these sequences (Morgante et al. 2002; Jung et al. 2005). In *Prunus*, 14,292 (19.8%) were detected after analysis of 72,114 EST sequences. In *Malus*, 19% of the HortResearch EST collection (see below) unigene sequences contain di- or tri-nucleotide repeats. Dinucleotide repeats were most frequently identified in the 5' untranslated region of the genes and AG repeats made up 88.3% of this class. In contrast, trinucleotide repeats are most common in the predicted coding regions and exhibited a lesser degree of sequence bias in their representation. Fifty-seven percent of di-, tri- and tetra-nucleotide repeats were 12 to 14 bases long, and only 17% were longer than 20 bases. This abundance of SSRs makes apple and *Prunus* ESTs a rich source of molecular markers for construction of framework maps (117 have already been mapped in the *Prunus* reference map and at least as many in *Malus*) and worth screening for when mapping candidate genes, as the markers are generally simple to develop and use.

EST collections are also a source of another important class of markers, single nucleotide polymorphisms (SNPs). Their frequency is usually high, so it is generally possible to identify SNPs between the parents of mapping populations when one is seeking to map candidate genes (S. Gardiner, H. Bassett, unpublished). SNP markers, although sometimes difficult to develop, are extremely useful for QTL mapping and eventually for MAS. Results available on SNPs in various Rosaceae species has been extracted from the GDR and are summarized in Table 1.

5.1.2. Prunus

With the objective of detecting and cloning genes coding for characters of tree and fruit development, Horn et al. (2005) created an EST collection from a cDNA

Table 1. Number of single nucleotide polymorphisms (SNPs) found in expressed sequence tag (EST) contigs of different Rosaceae species (source: Genome Database for Rosaceae)

	Nº of contigs	Nº SNPs	Nº SNPs/kb
Malus	22,435	10,426	0.5
Pyrus	35	1	0.0
Prunas	5,047	3,776	0.9
Fragaria	372	283	0.3
Rosa	705	316	0.7

library from developing fruit mesocarp of peach doubled haploid line P-21-5-2N. A total of 9,984 ESTs, 5'-sequenced and with an average size of 502 bp, were assembled in 3,842 putative unigenes (1,309 contigs and 2,533 singletons). Of these unigenes, 24.3% had no significant homology to the NCBI nr protein database. These unigenes were used for hybridisation with the 'Nemared' BAC library for the peach transcript map mentioned in Section 4.

The Chilean Functional Genomics Consortium (Meisel et al. 2006), involved in the study of the postharves behaviour of peach fruits, developed 50,625 cDNA sequences from four libraries of peach 'O'Henry' in four postharvest stages (mature fruits and mature fruits plus 15 days at 4ºC, and both treatments after a shelf life period). The ESTs had an average length of 673 bp and were sequenced from the 5' end. A total of 45,809 good quality ESTs revealed 10,830 unigenes (6,661 singletons and 4,169 contigs). Only 636 (5.8%) of these unigenes had no homology to *Arabidopsis* genes. Most ESTs of this collection (32,535) have recently been publicly released.

The IBMCP of València (Spain) and the University of California (Davis) are collaborating to analyse gene expression associated with chilling injury in peach fruit. In this project, CHILLPEACH, a collection of full-length enriched cDNAs from fruit mesocarp has been developed using two genotypes, one sensitive and one tolerant to chilling injury (Granell et al. 2006). Sequences of high quality (3,049) were clustered as 495 contigs and 1,595 singletons for a total of 2,090 putative unigenes. About half the unigenes were previously unreported in *Prunus* and 23.6% of the sequences were not present in *Arabidopsis*. A subtractive hybridisation library enriched with cDNAs regulated during cold storage has also been developed by these groups.

Three cDNA libraries of 'Bergeron' apricot pericarp at three different stages of development (89, 107 and 117 days after anthesis, corresponding to green, half-ripe and ripe fruit, respectively) were used by Grimplet et al. (2005) to obtain 13,006 good quality sequences in approximately equal amounts from each library, with an average length of 527 bp, the majority sequenced from the 3' end. These ESTs were assembled in 5,219 unigenes (3,426 singletons and 1,793 contigs). Twenty-four per cent of these unigenes did not show any similarity with the non-redundant protein public database. Fifteen transcripts were found to be strongly differentially expressed based on the redundancy of each unigene at each stage of fruit maturation. Twelve of them have been validated with quantitative PCR (Q-PCR) (Chevalier et al. 1999; Grimplet et al. 2005), but some of the genes classified as differentially expressed with lower degrees of stringency were not validated with Q-PCR.

5.1.3. Malus

The first large-scale EST sequencing of apple resulted in the collection and analysis of 151,687 cDNA sequences by HortResearch (Newcomb et al. 2006). The sequences are predominantly from 'Royal Gala' (almost 120,000), with lesser numbers from 'Pinkie', 'Pacific Rose', 'M.9', 'Aotea 1', 'Braeburn' and 'Northern Spy'. The 43 source libraries represent a range of tissue types, with a bias towards

fruit (53,620 ESTs). The fruit libraries include a series from developing and then ripening fruit with samples from flower, whole fruit, fruit cortex and seeds. This series was developed to provide a resource of genes for research on processes involved in fruit development (e.g. early cell proliferation, cell expansion and ripening). Other tissues sampled include buds, shoots, leaves, roots, phloem and xylem (76,472 ESTs). Other ESTs were expressed in tissues exposed to biotic and abiotic stresses, for example leaves infected with *V. inaequalis*, or exposed to temperature stress, fruit stored at low temperature and altered atmospheric conditions after harvest, and fruit cell lines exposed to boron (21,595 ESTs). cDNA sequencing was predominantly from the 5' end, and average edited length of the sequences was 468 bases. Sequence clustering yielded 43,938 putative unigene sequences at a clustering threshold of 95% (a more stringent threshold than used by GDR above), broken down into 17,460 contigs and 25,478 singletons. The number of unigene sequences is expected to be an overestimate of the number of protein-coding transcripts represented in apple and the authors recommended further sequencing to reduce the number of unigene sequences. Comparison of *Arabidopsis* unigenes from all ESTs with protein coding genes suggests that the apple unigene set in this study represents approximately half the genes in apple.

Functional annotation is an important first step in identification of candidate genes for specific traits of economic importance. Predictive bioinformatics showed that only 5% of the apple unigene sequences failed to match a sequence in the genome of *Arabidopsis* (Newcomb et al. 2006). Candidates for pest and disease resistances constitute 0.25–3% of the ESTs and 0.55% of contigs in the dataset (Crowhurst et al. 2005; Rikkerink et al. 2003). Structural gene families that are potentially involved in the biosynthesis and maintenance of flavour and health promoting compounds in apple fruit are well represented. Over 1,000 transcription factors have been identified among the ESTs as a resource for candidates for regulation of biosynthetic pathways (Crowhurst et al. 2005; Newcomb et al. 2006).

The other major EST sequencing programme has delivered 109,824 sequences to the Rosaceae research community, with a further 50,000 scheduled (Gasic et al. 2006). The 34 source libraries were deliberately constructed from a diverse range of germplasm, to facilitate *in silico* SNP mining. Sources include: 'Goldrush', 'Jonagold', 'Granny Smith', 'Fuji', 'Braeburn', 'Suncrisp', 'M.9', 'M.111', 'Geneva 3041', and 'Geneva 3041' x *M. sieversii*. Both normalized and primary libraries have been constructed. Different developmental stages of 'Goldrush' are represented by libraries from leaf, bud, flower and shoot. Sequencing was from the 5' end, and sequence clustering yielded 13,538 contigs and 15,896 singletons. Bioinformatic analysis enabled assignment of putative functional category for 82.5% of sequences. When classified according to gene ontology, 38.3% of sequences were assigned to biological processes, 42.7% to cellular components and 33.3% to molecular function.

SSH (Suppression Subtractive Hybridization) is used to generate libraries enriched with transcripts that are differentially expressed and increases the likelihood of identifying weakly expressed genes (Diatchenko et al. 1999). It has

been used to identify 480 ESTs differentially expressed between uninfected leaves of the scab resistant cultivar 'Remo' and 'Elstar', which is susceptible to scab infection (Degenhardt et al. 2005). Levels of transcripts coding for proteins already known to be related to plant defence, such as B-1,3-glucanase, ribonuclease-like PR10, cysteine proteinase inhibitor, endochitinase, ferrochetalase and ADP-ribosylation factor were higher in 'Remo' than in 'Elstar'.

A recent study utilized EST frequency analysis of publicly available ESTs (150,000 sequences at the time), with the goal of identifying genes most likely to be highly expressed in apple fruit, expressed uniquely or preferentially in fruit, and/or temporarily or spatially regulated during fruit growth and development (Park et al. 2006). A subset of genes that may be related to the generation of flavour and aroma components in mature apple fruit was identified. It was concluded that although this technique offers an entry point into fruit molecular biology, there are a number of limitations and caveats. One is the sensitivity of identification of the most abundant ESTs to artefacts, including differential amplification of cDNAs during library preparation, and another is the contamination of libraries with abundant organellar DNAs, highly repetitive DNA in the nuclear genome and microbial nucleic acids.

cDNA-AFLP provides an expression profile of bands that can be sequenced for identification of candidate genes. It has been used in apple, but it tends to be cumbersome in application and is expected to be superseded by microarray technology. A study by Jensen et al. (2003) compared expression of genes in 'Gala' grafted onto apple rootstocks 'M.7' and 'M.9' and challenged with fireblight (*Erwinia amylovora*). Another study was performed to identify genes with expression changes during fruit maturation and ripening (Lin 2005). One hundred and fifty-five of the 204 differentially expressed fragments were confirmed by microarray analysis after sequencing, and 127 assigned to functional classes. A study with similar goals is yielding candidate genes (Dreesen et al. 2006) that will be mapped on a QTL map for fruit maturation and ripening (Kenis and Keulemans 2005).

Another form of transcript profiling, mRNA differential display, coupled with RT-PCR to eliminate false positives, has been employed to detect 14 cDNAs representing transcripts up-regulated during ripening of 'Modial Gala' fruit ripening (Goulão and Oliviera 2006).

5.2. Microarrays

Microarray analysis is becoming a significant tool for characterizing transcription profiles of collections of thousands of genes simultaneously, providing an initial selection of several hundred potential candidate genes that could be associated with a specific physiological process.

A peach microarray containing about 4,800 oligonucleotide probes obtained from EST sequences corresponding to different unigenes selected for their expression in the last stages of fruit development (μPEACH1.0) was developed by Trainotti

et al. (2006) and used to study gene expression in the transition from the pre-climacteric to the post-climacteric phase in peach fruits. Transcriptome analysis of this microarray detected 267 up-regulated and 109 down-regulated genes. These include genes for ethylene biosynthesis and perception directly involved in the maturation process, genes responsible for quality factors such as flesh firmness and carotenoid content, and transcription factors involved in different aspects of fruit development and maturation. Microarray development and analysis is also being carried out by the groups dealing with peach fruit postharvest behaviour (Orellana et al. 2006; Granell et al. 2006) and by those targeting apricot fruit development (Grimplet et al. 2005).

The power of microarray technology has been demonstrated in apple by two detailed studies of fruit development (Janssen et al. 2006) and of ripening (Schaffer et al. 2006) that utilize a 15,760 50mer array developed by HortResearch (New Zealand). In the first study, nearly 2,000 of 13,000 unigenes demonstrated a significant change during development over 150 days from pollination to full tree ripeness (Janssen et al. 2006). Cluster analysis enabled grouping of genes associated with specific stages, such as cell division during early development, starch accumulation during mid development and starch decline as fruit begin to ripen. Quantitative PCR has been used to examine expression patterns for over 50 genes, confirming the expression patterns found by microarray for 70% of these, and identifying specific genes for further study. Key stages in ripening of fruit have been studied using transgenic apples transformed with an antisense copy of ACC oxidase. Fruit from these trees do not ripen unless they are treated with exogenous ethylene, allowing ethylene-induced ripening to be studied. Genes identified in this study are involved in the increase in volatiles related to flavour perception, increase in sweetness, softening of the flesh, and reduction of acidity (Schaffer et al. 2006). Candidate genes responding slowly to ethylene stimulus included those involved in flavour biosysnthesis and ripening. By combining data from both of the above studies, it has been possible to identify the 500 genes that respond to ethylene, and distinguish between genes involved specifically in ethylene-induced ripening and genes induced by ethylene that are not specifically related to ripening. In addition it is possible to identify ripening associated genes that are not ethylene-induced (Schaffer et al. 2006). This type of approach that combines data from multiple experiments in the Rosaceae is likely to yield further significant results, particularly in the area of fruit development. The HortResearch microarray is available to the research community on a commercial and collaborative basis.

A 60-mer oligo array representing 55,000 apple contigs from the NCBI database and developed in collaboration with NimbleGen (http://www.ncbi.nlm.nih.gov/projects/geo/query/acc.cgi?acc=GPL3715) is being used in a study of gene expression patterns of ‘Gala’ scions grafted on eight different apple rootstocks to identify a set of rootstock regulated genes that may be associated with resistance to disease and drought (Jensen et al. 2006). This microarray is commercially available. Another oligo-based array is being planned at INRA, Angers, France (Wattebled et al. 2006). Variations in gene expression related to apple texture will be confirmed

by quantitative PCR, enzymatic assays, or by transgenic validation in tomato. Candidate genes will then be mapped to QTLs in populations segregating for these traits and the data obtained will be applied in apple breeding programmes.

A specialized 5,000 clone cDNA microarray (MSU MDAZ5k) has been developed for the purpose of characterizing gene expression in natural and chemical-mediated apple fruit abscission, and developing improved strategies for regulating flower and fruit abscission, and hence crop load (van Nocker et al. 2006).

The European project HiDRAS (High-Quality Disease Resistant Apples for a Sustainable Agriculture) (Gianfranceschi 2006) has developed a small (1,600 clone), specialized cDNA microarray that targets genes controlling fruit development and maturation. Subtractive DNA libraries were used to enrich for fruit specific genes. Genes showing modulated development during fruit development (about 10% of those on the slides) have been sequenced and annotated prior to marker development and mapping to QTLs controlling fruit quality traits on the 'Fiesta' x 'Discovery' apple reference map (Soglio et al. 2006). A Japanese group has developed a similarly sized cDNA microarray (1,431 clones) from shoot and fruit libraries, and used it to examine gene expression changes following treatment with an *Alternaria alternata* host specific toxin (Hatsuyama et al. 2003).

In apple, there has been a focus on development of *in vivo* expression systems for verification of candidate genes identified from annotation and microarray experiments. The precise strategy will depend on the predicted function of the gene product, and can involve expression in yeast or *E. coli* followed by headspace analysis (for volatile products), transient expression in tobacco following *Agrobacterium* infiltration (for transcription factors that modify the expression levels of enzymes), stable expression in *Arabidopsis* or apple cell lines / whole plants (for other genes affecting complex phenotype) (Souleyre et al. 2005; Lee and Lee 2005; Hellens et al. 2005; Chagné et al. 2006a; Chagné et al. 2006b). The elucidation of the function of a gene product using a combination of these systems is clearly illustrated for the study of *M. pumila* alcohol acyl transferase (MpAAT1) (Souleyre et al. 2005), which used headspace analysis of a range of substrates by an *E. coli* expression system and a transient expression system in tobacco. It was demonstrated that this gene is responsible for the final step in the synthesis of a range of the esters that characterize the flavour of 'Royal Gala' fruit. In another case, the identification of a gene controlling red flesh in apple was assisted by expression of apple candidate genes in tobacco and apple (Chagné et al. 2006a; Chagné et al. 2006b; Espley et al. 2006).

Gene knock-down RNA interference (RNAi) technology (Wesley et al. 2001) utilizes the expression of gene transcripts in the whole plant in a sense or antisense fashion, and observation of the effect on phenotype to infer gene function. However, this requires an efficient transformation system and regulatory conditions that permit growth of transgenic plants, preferably under field conditions. A good transformation system exists for apple, but for *Prunus* the efficiency is low (Petri and Burgos 2005). For *Prunus* transformation is more easily performed in juvenile materials and is extremely dependent on species and genotypes, meaning that these

experiments often have to be done in other organisms. RNAi technology has been used in a number of studies on the role of ethylene in controlling development of apple fruit quality attributes by suppressing ethylene production in transgenic plants to study its effects on e.g. sugar, acid and ester level in fruit (Defilippi et al. 2004), levels of sorbitol, sucrose and starch in leaves (Cheng et al. 2005), flavour, aroma and texture in fruit (Dandekar et al. 2004; Defilippi et al. 2005), and is being used in a HortResearch study involving microarray analysis of genes involved in fruit ripening (Schaffer et al. 2006).

6. DATABASES AND ASSOCIATED BIOINFORMATICS SYSTEMS

There are a number of databases that have been developed by members of the Rosaceae research community, with varying degrees of public access. Only one of them, the GDR, is public, covers diverse topics (mainly structural and comparative genomics and transcriptomics) and includes all species of the family. The remaining databases are usually oriented towards one species, correspond to projects of research organizations, consortiums or individual research groups, and are mainly focused on the area of transcriptomics. This diversity of databases has led to a duplication of effort as well as to lack of availability of particular features developed in-house in individual systems, to the community as a whole.

The GDR is a curated and integrated web-based relational database that includes genomics and genetics data for members of the genus Rosaceae (Jung et al. 2004). The GDR contains comprehensive data of the genetically anchored peach physical map, Rosaceae maps and markers, and an annotated EST database of peach, almond, strawberry, and all publicly available Rosaceae sequences. The integrated map viewer provides a graphical interface to the genetic, transcriptome and physical mapping information. ESTs, BACs and markers can be queried by various categories and the search result sites are linked to the integrated map viewer or to the WebFPC physical map sites. CMap, the comparative map viewer, allows users to compare various Rosaceae genetic maps and the transcriptome map. In addition to browsing and querying the database, users can compare their sequences with the annotated GDR sequences via a dedicated sequence similarity server running either the BLAST or FASTA algorithm. GDR can be accessed at http://www.rosaceae.org and provides free EST processing services to the Rosaceae community.

ESTree is a public database (http://www.itb.cnr.it/estree/), developed by Lazzari et al. (2005), with a collection of 18,630 ESTs (6,155 from four *P. persica* mesocarp cDNA libraries obtained by this group and the rest from public databases) and is intended as a resource for peach functional genomics. These sequences can be processed through an automated pipeline including public software and in-house PERL scripts that allow these ESTs to be processed and provide information on contig assembly, SNP detection, gene annotation with Blastx and assignment to Gene Ontology classifications. Contig assembly has allowed 9,219 putative unigenes from these sequences to be detected. A total of 166 of these unigenes containing

putative SNPs have been identified and are currently being validated and bin mapped.

Two databases, CHILLPEACH (Granell et al. 2006) developed by the US-Spanish project and JUICE (Meisel et al. 2006) developed by the Chilean consortium mentioned in Section 4, have been created for the processing and assembly of the cDNAs obtained in these projects and are for the moment only for internal use. CHILLPEACH also includes devices for functional annotation, such as GO terms, BLAST results, EC numbers, protein family (PFAM) domains and sequence analysis to identify SNPs or SSRs.

The HiDRAS AppleBreed database provides access for members of the HiDRAS programme to genotypic and phenotypic data derived from plant populations with different pedigrees (progenies, cultivars used as progenitors, breeding selections) (Antofie et al. 2006). Its goal is to link molecular information to phenotypic data, and to ensure linkage between information on cultivars and on crosses (families). The core of the database model is based on individual trees and individual DNA samples. This structure allows multiple annual observations to be stored individually by genotype, even when the nature of these observations is very different (e.g. molecular data, physical or chemical measurements of fruit quality traits, evaluation of disease susceptibility, etc.) Phenotypic data, molecular marker data and links to the individual trees and DNA samples from which these data were derived, pedigrees, marker descriptions, primer sequences, molecular marker linkage maps and synonyms of the cultivar names are included, as well as validation procedures for phenotypic and marker data. Basic statistical overviews on the data are provided. The information on apple microsatellites is available online at http://www.hidras.unimi.it, including facilities to encourage interaction between research groups in this area (Silfverberg-Dilworth et al. 2006).

Tree Fruit Technology (TFT), a database developed at Michigan State University, which contains sequences from public databases for Rosaceous fruit crops grown in temperate regions of the USA, provides another resource for analysis of DNA sequence information. TFT is available online at http://genomics.msu.edu/fruitdb.

The HortResearch enterprise databases contain at least 270,000 EST, PCR, and full cDNA sequences from *Malus*. The database includes all HortResearch sequences merged with *Malus* sequences extracted from GenBank (Crowhurst et al. 2006). HortResearch has developed an automated analysis and visualization programme for the annotation of non-genomic DNA (BioView™ Sequence Analysis and Annotation System). The system is primarily written in PERL and employs MySQL as a relational database management system for persisting data and job scheduling. The automated pipeline assembles sequences into tentative consensus sequences, and implements modules for performing BLAST comparisons, identifying SSRs and motifs, predicting SNPs and microRNA targets, designing microarray oligonucleotides, compiling keyword dictionaries, and defining gene families. Visualizing relationships captured within the database is achieved via a web viewer. Administering the system, tracking and sharing mining efforts, initiating follow-on sequencing and vector construction, performing virtual subtraction,

and augmenting automated annotations by capturing end user knowledge is also undertaken via the web viewer (Crowhurst et al. 2006). The BioView system will incorporate a database for markers and support GMOD/CMap modules (http://www.gmod.org/?q=node/77) for comparative map display in its next release. This in-house system is acknowledged to be comprehensive and user-friendly.

7. CURRENT APPLICATIONS

Genomics provides a 'tool box' of technologies that have started to be used for development of new, improved varieties of *Prunus* and *Malus.* Genetic markers are used directly by breeders for germplasm assessment, breeders' rights protection and marker-assisted selection from breeding populations. Other 'tools', e.g. bioinformatics, microarrays, quantitative PCR and *in vivo* expression, are providing insight into the function of gene products that determine characters of interest to breeders.

7.1. *Prunus*

7.1.1. Marker assisted selection (MAS)

Most of the major genes that have been placed on the *Prunus* consensus map (Dirlewanger et al. 2004a) are important in breeding programmes. They affect fruit quality (peach v. nectarine; flat v. round fruit; clingstone v. freestone; subacid v. acid flavour; yellow v. white or melting v. non-melting flesh), kernel (bitter v. sweet) or shell (hard v. soft) traits in almond; tree characteristics (columnar v. normal), disease resistance (root-knot nematode, plum pox virus, powdery mildew) or other characters (self-incompatibility, blooming time). Given the marker density of the current *Prunus* map, markers (usually SSRs) are available sufficiently close to most of these genes for use in marker assisted selection.

However, markers tightly linked or derived from the sequence of three genes are the only ones routinely used for selection in breeding programmes. The two independent genes (*Mi* and *Ma*) that confer root-knot nematode resistance are currently being applied for pyramiding these genes into a new wave of rootstocks (Dirlewanger et al. 2005). Markers obtained from the sequence of the stylar RNAse involved in the self-incompatible reaction are being used to assign genotypes to self-incompatibility classes or to select for self-compatible individuals in almond, apricot and cherry progeny (López et al. 2005; Vilanova et al. 2005; Sonneveld et al. 2006). Other genes with potential for MAS are those of the major factors involved in plum pox virus resistance on G1 (Vilanova et al. 2003), and two more affecting fruit quality: the acid/subacid (*D*) gene and the melting/non-melting (*M*) gene. In addition, microsatellite markers are extensively used for cultivar fingerprinting, mainly for the enforcement of breeders' rights and for applications such as parentage tests, or the management of germplasm collections (Arús et al. 2005c).

7.1.2. Research on characters of value to breeders

One of the targets of plant breeding is the incorporation of disease resistance genes in elite varieties. The ever-increasing amount of information on the sequence and function of genes involved in plant resistance to pests and diseases in many organisms is being used to complement the data on the known positions of major genes and QTLs to plum pox virus (Vilanova et al. 2003; Decroocq et al. 2005), powdery mildew (Foulongne et al. 2003; Verde et al. 2005), root-knot nematodes (Dirlewanger et al. 2004b), leaf curl (Viruel et al. 1998) and green leaf aphid (Foulongne 2002) in the search for candidate genes or the locations of disease resistance gene clusters. The strategy of resistance gene analogs (RGA) to map resistance gene clusters developed by Leister et al. (1996) in potato has been applied by Bliss et al. (2002) in peach and by Soriano et al. (2005) in apricot to map various RGAs in the genomes of these species. More extensive work has been done by Lalli et al. (2005) using the physical and genetic *Prunus* maps, the EST sequences kept at the GDR, and sequence data from RGAs obtained from degenerate primers in various *Prunus* species. Fifty-eight DNA fragments representing putative RGAs and/or genes involved in host defence or resistance were hybridised to the 'Nemared' BAC library used for physical map construction. Thirty of them identified 42 positions in the TxE map distributed in all linkage groups but one (G3), without the use of segregating populations. The positions detected by the RGAs often coincided with the locations of resistance genes or QTLs mapped previously, providing a new source of markers for MAS of seedlings with favourable resistance genotypes.

One of the key characters for the selection of new stone fruit cultivars is fruit quality, evolving from traits related to the external appearance (size, shape and skin colour) to aspects related to their organoleptic qualities, such as flavour, flesh consistency and aroma. Some of these characters were analysed in peach (Dirlewanger et al. 1999; Etienne et al. 2002) and peach x *P. davidiana* (Quilot et al. 2004; Quilot et al. 2005) progenies, and major genes or QTLs for fruit size, fruit shape, contents of different organic acids (citric, malic, quinic and shikimic) and sugars (sucrose, fructose, glucose and sorbitol) were identified. One of the candidate genes studied by Etienne et al. (2002) corresponded to a proton pump-encoding gene and co-located with the position of a QTL for sucrose and soluble solid content. Interestingly, some of these major genes or QTLs are located within the same genomic regions across different progenies and species, suggesting that gene position and gene action are conserved across species. For example, QTLs for fruit weight and dry matter weight mapped in both of these populations to the end of G5, close to the gene *G* that determines the skin hairiness of the fruit (peach v. nectarine).

The postharvest behaviour of fruit has become a major aspect of fruit breeding programmes. The stone fruit sold out of season in the main consumption centres of the Northern hemisphere is produced in the Southern hemisphere (mainly New Zealand, Chile and South Africa), implying a requirement for long periods (more than 30 days) between harvest and consumption. Research is being directed towards finding alleles of genes involved in fruit maturation, which may allow, alone or

in combination, the development of new varieties that maintain their high quality flavour after reaching the consumer. One of these is the previously mentioned *M* gene that determines melting (soft) v. non-melting (hard) flesh consistency (Peace et al. 2005). Another interesting gene (*Hd*) confers the stonyhard characteristic, where the fruit keeps its texture for a long time but, unlike fruit with the *M* gene, does not produce ethylene, and a postharvest ethylene treatment is required to soften the fruit (Haji et al. 2005). Tatsuki et al. (2006) showed that the stonyhard genotype occurs as a consequence of the suppressed expression during the ripening stage of a gene encoding one of the enzymes of ethylene biosynthesis, an ACC syntase (*Pp-ACS1*) gene that is normally expressed in tissues subjected to wounding. These results suggest that the *Hd* gene is part of the regulatory mechanism of *Pp-ACS1*. Mapping the position of the *Hd* gene with respect to that of candidate genes may help to clarify its nature. The fruit softness of peaches is currently being analysed with the candidate gene approach by Peace et al. (2006a) and a transcriptomics-based search of genes involved in fruit post-harvest life is being used by Granell et al. (2006) and Orellana et al. (2006).

Blooming and fruit maturity times are also important characters in *Prunus* breeding programmes (Scorza and Sherman 1996). Early blooming may expose flowers or recently formed fruits to late frosts, and late blooming cultivars are sought in almond and stone fruits to increase production and crop stability. The chilling requirement to overcome dormancy is associated with blooming time and is an important factor for genotype adaptation of most species, where the number of chill hours has to be sufficient to allow for regular flowering and fruit development in the region where the cultivar is grown. The production in regions with mild climates can be extended for early or late maturing cultivars, usually leading to higher market prices. Both bloom and maturity times have been described as polygenic and highly heritable characters (de Souza et al. 1998). QTLs for blooming time have been identified in three regions of the genetic map (Joobeur 1998; Dirlewanger et al. 1999; Quilot et al. 2004; Verde et al. 2005): in the central region of G4 that includes also a major gene (*Lb*) for blooming time of almond (Ballester et al. 2001), in the middle of G7, and in the southern region of G1, close to the evergrowing (*Evg*) gene that determines a continuous terminal meristematic growth in peach in winter dormancy-inducing conditions (Wang et al. 2002). Twelve candidate genes with sequences highly homologous to those involved in vernalization and flowering of *Arabidopsis* were mapped in TxE by Silva et al. (2005), and one of them, a MADS box transcription factor, was located near the *Evg* locus.

7.1.3. Map-based cloning

The root-knot nematode resistance gene *Ma* from myrobolan plum has been fine mapped and a BAC including two flanking markers identified (Claverie et al. 2004). This BAC has been fully sequenced and candidate genes for resistance identified by Claverie (2004). The *Evg* gene on G1 has been fine-mapped and a contig of three BACs spanning a region of 132 kb containing the gene was obtained by Bielenberg et al. (2004). Seven candidate genes have been identified with the wild

type sequence of this genomic region (Bielenberg et al. 2006), six of which are highly similar copies of a MICK structural class of MADS-box genes, and the seventh is a calcium-binding protein. The mutant genotype has a deletion in this region of 40-45 kb, which affects up to four of the MADS-box candidate genes. The gene *D* that determines subacid v. acid taste is in the process of being map-based cloned (E. Dirlewanger, pers. comm.). Other important genes have also been genetically characterized in depth. The two major components (stylar and pollen) of the self-incompatibility locus have been sequenced (Ushijima et al. 2003) and their variation has been associated with the known self-incompatibility alleles in different *Prunus* species. Strong evidence exists that an endopolygalacturonase gene corresponds to the gene *M* that determines the melting v. non-melting flesh consistency of the peach fruit (Lester et al. 1996; Peace et al. 2005).

7.2. *Malus*

7.2.1. Marker assisted selection (MAS)

Genetic markers have been utilized to varying degrees for at least 8 years to increase the efficiency of apple breeding programmes (Bus et al. 2000; Bus et al. 2002; Cheng et al. 1998; Kellerhals et al. 2000). Applications include screening of germplasm with markers for *Vm* (Cheng et al. 1998; Mattison and Nybom 2005) as an alternative to cumbersome allelism tests (MacHardy 1996); selection of parents with desirable alleles for a range of characters, including desired combinations of different resistance genes against the same pathogen (Bus et al. 2002); early selection from large breeding populations for adult traits at the seedling stage, including traits that are expensive to phenotype and for gene pyramids for resistances that cannot be selected by phenotype (Bus et al. 2002; Costa et al. 2005; Itai et al. 2003). MAS is being used internationally for at least nine major gene resistances against apple scab, five major resistances against powdery mildew, two against woolly apple aphid, and one against rosy apple aphid (Gardiner et al. 2006a). Selecting for desirable QTL traits with markers is less advanced although it has been initiated at INRA, France for scab resistance (C.E. Durel, pers. comm.). The value of MAS in a breeding programme is greater than the cost of replacing phenotypic screens, and includes the potential value of durable resistances in horticultural production systems (Luby and Shaw 2001). The introduction of automated systems for DNA extraction, setting up reactions and gel loading, as well as the use of marker multiplexes (Cook and Gardiner 2004; Frey et al. 2004) is reducing the costs of labour and consumables for MAS to breeding programmes.

High density saturated genetic framework maps are essential for the determination of marker-trait associations in apple using the Genome Scanning Approach (Patocchi et al. 2004; Patocchi et al. 2005; Silfverberg-Dilworth et al. 2006). Because of the highly polymorphic and co-dominant nature of microsatellite markers, framework maps are also a practical source of alternate markers for genes in introgression programmes incorporating new ‘quality parents’ in successive

generations (Gardiner et al. 2006a). Genetic framework maps that utilize multi-locus markers, such as microsatellites, RFLPs or SNPs developed from paralogous genes, are also invaluable for identifying regions that are homeologous between linkage groups (Liebhard et al. 2002; Maliepaard et al. 1998; Gardiner et al. 2006b). Genetic markers can be used to elucidate the segregation of traits subject to segregation distortions, such as the *Er3* resistance to woolly apple aphid (Gardiner et al. 2006b), as well as the number of genes involved in determining a trait such as the *Pl-1* powdery mildew resistance (Dunemann et al. 2005).

7.2.2. Research on characters of value to breeders

Analysis of the function of a gene product using application of a combination of genomic technologies is shown by the case of *MpAAT* (Souleyre et al. 2005). Gene mining in the apple EST database had identified 20 acyl transferase-encoding genes that were candidates for determining the ester profile behind the fruit flavour of 'Royal Gala' apples. One of these was chosen for study, because it was derived from fruit libraries, is closely related in sequence to acyl transferases that use alcohol as an acceptor, microarray analysis showed it is up-regulated in developing 'Royal Gala' fruit and ethylene treated fruit and it is expressed in apple flowers and fruit. The protein expressed by recombinant *E. coli* included the predicted fusion protein containing the expected MpAAT1 polypeptide, confirmed by LC-MS (Liquid Chromatography Mass Spectrometry). Headspace analysis by GC-MS (Gas Chromatography Mass Spectrometry) of *E. coli* cultures expressing *MpAAT1* that had been provided with a range of alcohol substrates and CoA donors defined the range of activities for MpAAT1. A transient expression system utilizing *Nicotiana benthamiana* (tobacco) produced a smaller range of esters compared with those from *E. coli,* not all found in apple fruit. It was concluded that the substrate pool in apple fruit is likely to determine the range of esters produced. The final step will be to map *MpAAT1* on a genetic map to determine whether this candidate gene co-segregates with a QTL for apple 'fruit flavour'.

The ultimate goal of the candidate gene approach is to identify markers with a gene directly controlling the trait of interest. An example of how the application of genomic technologies makes this possible is the identification and mapping of a gene controlling red colour in apple flesh and foliage (Chagné et al. 2006a; Chagné et al. 2006b; Espley et al. 2006). Candidate genes were identified in the apple EST databases on the basis of homology to genes controlling red colour in model plants, either genes involved in anthocyanin biosynthetic pathways or transcription factors (TFs) regulating those pathways. The candidates that represented biosynthetic genes were excluded on the basis of expression data from Q-PCR and microarray experiments that demonstrated that biosynthetic genes are all switched on or off in red and white fruits, respectively. A further candidate TF sequence was obtained, following PCR amplification with degenerate primers from the parents of a mapping population segregating for red flesh. The gene candidates were then reduced to the 13 that were differentially expressed between extreme phenotypes in Q-PCR and microarray analyses. One of these, a MYB transcription factor (MdMYB10)

expressed red pigment patches following transient transformation in tobacco, particularly when co-transformed with another apple transcription factor (Espley et al. 2006). Molecular markers developed from SNPs and microsatellites identified in the sequence alignments were screened over mapping populations to identify co-segregation with phenotype (Chagné et al. 2006a; Chagné et al. 2006b). One of the candidates, MdMYB10, exhibited an allele closely associated with red core colour and was mapped on a 'M.9' x R5 reference map (Celton et al. 2006a; Celton et al. 2006b). The linkage between MdMYB10 and the red core trait was confirmed in an association study by screening several markers developed from different SNPs in the gene sequence over a set of unrelated red and white fleshed individuals from international germplasm collections (S. Gardiner, D. Chagné, unpublished). Overexpression of MdMYB10 resulted in generation of highly pigmented apple plants that contained enhanced levels of anthocyanin. This suggests that differences in the activity of MdMYB10 determine the red core colour trait, but confirmation awaits fruiting of the transformants.

Eighteen resistance gene analogues (RGAs) with homology to NBS/LRR genes have been mapped as cleaved amplified polymorphism sequences (CAPS) or single-strand conformation polymorphism (SSCP) markers (Baldi et al. 2004) and 23 more were mapped by NBS profiling (Calenge et al. 2005b). Work in progress using candidates derived from EST sequences has resulted in mapping of over 40 more candidates, either as microsatellites or SNPs (Gardiner et al. 2003; S. Gardiner, unpublished). Clustering of RGAs and association with major gene and QTL loci conferring pathogen resistance has been noted in all three studies.

7.2.3. Map-based gene cloning

Map-based cloning in apple focused initially on major genes for resistance, with a more recent project targeting a gene involved in starch metabolism. Most progress has been made in identifying the *Vf* gene. The first BAC contig containing the *Vf* gene covered a 550 kb minimal tiling path and was developed by Patocchi et al. (1999b) on the basis of fine mapping of markers around the *Vf* gene and on new markers generated from BAC clones. Identification of candidate genes for *Vf*, named *HcrVf1* to *HcrVf4* (Vinatzer et al. 2001a) involved the use of BAC inserts to probe a large cDNA library and built on further work by Vinatzer et al. (1998) and Patocchi et al. (1999a). The candidate genes showed homology to the *Cf* resistance genes cloned from tomato. When one of the candidate genes (*HcrVf2*) was expressed in 'Gala' under a 35S promoter, it conferred resistance to apple scab (Belfanti et al. 2004), which was overcome by a scab race that specifically overcomes *Vf* resistance (Silfverberg-Dilworth et al. 2005). Screening of BAC subclones with labelled total cDNA (Xu and Korban 2002b) yielded an almost identical set of candidate genes (*Vfa1* to *Vfa4*). As transformation of cultivars 'Galaxy' and 'McIntosh' has demonstrated that *Vfa1* and *Vfa2* each independently confer resistance to apple scab (Malnoy et al. 2006b), it is not clear which candidates (or combination of candidates) are responsible for the *Vf* resistance. Work is at an advanced stage on map-based cloning of both *Sd-1* which confers resistance

to the aphid *Dysaphis devecta* (Cevik and King 2002), and the powdery mildew *(Podosphaera leucotricha)* resistance *Pl2* (E. Rikkerink, pers. comm.). There is a brief report on the map-based cloning and characterization of a gene encoding starch branching enzyme (*SBE1*) from apple (Han et al. 2006b).

7.2.4. Genetic transformation

Transformation has been employed in apple to enhance resistance to pathogens, particularly where endogenous monogenic resistances have not been available in the germplasm (Malnoy and Aldwinckle 2006). Introduction of heterologous antibacterial or antifungal genes enhanced resistance to fireblight in apple (Ko et al. 2002; Liu et al. 2001; Norelli et al. 1994; Schneider et al. 2006) and pear (Lebedev et al. 2002; Malnoy et al. 2003), as well as to apple scab (Bolar et al. 2000; Bolar et al. 2001; Chevreau et al. 2004). Resistance to light brown apple moth was enhanced by introduction of heterologous genes coding for biotin binding proteins (Markwick et al. 2003). More recently, transformation with the endogenous NPR1 (Non-expressor of Pathogenesis Related 1) gene was used to increase fireblight resistance under some conditions (Malnoy et al. 2004; Malnoy et al. 2006a). Silencing of DspE (Disease Specific E) interacting proteins (DIPMs) (Borejsza-Wysocka et al. 2004; Borejsza-Wysocka et al. 2006) increased resistance to fireblight in some transgenic lines of 'Galaxy'. The development of promoters that are activated specifically by exposure to pathogens at the site of infection will reduce accumulation of transgenic proteins in the plant and has been demonstrated in apple for the *Pgst1* promoter from potato (Malnoy et al. 2006c).

8. FUTURE PROSPECTS

8.1. Resource Integration

Rapid progress in Rosaceae tree fruit research depends on the integration of research resources internationally within and across genera, as adequate funding is not available in any one country for any of these second or third tier crops to make significant progress alone in genomics. At the International Rosaceae Genomics Conferences, the third of which was held in Napier, New Zealand in 2006, such interaction was initiated on a broad scale (http://www.mainlab.clemson.edu/gdr/), and enabled the setting up of the International Rosaceae Genomics Initiative (RosIGI). A national programme has been initiated in the USA (Norelli et al. 2006; Peace et al. 2006b).

Links between databases and integration of bioinformatics systems are critical, and there should be a common portal for accessing databases. The most comprehensive of these databases is currently the GDR, which has already proven helpful to many research groups and could become a central resource for the whole Rosaceae scientific community. There needs to be a way that other databases that have developed special features can be used to support the wider Rosaceae community, while protecting any special features relating to intellectual property

for the developer. Substantial new resources should be developed in collaboration, e.g. microarray analysis systems, genome comparison analyses and whole genome sequences.

Sharing of significant mapping populations at the plant and/or DNA level according to individual countries' regulations has been initiated on a regional scale in Europe, and if extended across the world, would aid in correlation of mapping datasets and reduce duplication of effort. An international consortium of research groups (the RosPOP consortium), supported by the RosIGI, is currently being launched and will have its base in the GDR. The aim is for all members to have access and share populations, DNA and data. This is likely to facilitate interactions between different research groups in diverse activities related to major gene, QTL and candidate gene analysis as well as expression profiling, enabling studies to be conducted with the same genes or markers over several populations in common, but grown in different environments. Sharing of DNA samples from the parents of the mapping populations will enable polymorphism assessment of markers developed by different groups within RosPOP. Additional phenotyped populations will be linked into this network by the mapping of common markers.

The very large (>260,000) set of apple EST sequences available in public databases provides the resource needed for construction of the first internationally available microarray containing all available apple features. Areas under discussion internationally include the type of platform, length of the oligos and whether chips should be designed for individual genera, or a 'family chip' could cover all the Rosaceae.

The comparison of the apple and *Prunus* genomes is only beginning. It requires a more detailed analysis to establish with precision the major rearrangements that have occurred between these genomes. This has the additional interest of comparing two close genomes, one of which (apple) has recently undergone a process of polyploidization, and may help to understand how this event took place and some of its genetic consequences. Research on the comparison between the genome of other Rosaceous crops such as *Prunus* and *Fragaria* or *Fragaria* and *Rosa* is underway (Rousseau et al. 2006; P. Arús, unpublished). One of the challenges ahead is to work with the Rosaceae genome as a whole, particularly in the development of certain resources (microarrays, whole genome sequence etc.) and for this, knowledge of both the gross and finer differences between genomes is crucial.

8.2. Mapping and Sequencing

There are already almost complete sets of *Prunus* and *Malus* microsatellite markers spaced across the genome (Howad et al. 2005; Silfverberg-Dilworth et al. 2006) that are highly useful tools for breeding. The *Prunus* set is extensive (an average density of 1.05 cM/SSR), and highly transferable across the genus, but additional coverage of some regions is still needed. For apple, a set of alternatives needs to be developed at each approximate map position. A similar set of markers is needed for pear, and a spaced set of universal apple/pear primers would be invaluable

for comparative mapping. Information on alleles across a cultivar set would aid in choice of framework markers for new mapping projects.

New approaches to mapping have increased the efficiency of locating new markers and mapping traits. Bin mapping, developed in *Prunus*, will be an efficient strategy for mapping large numbers of new markers, e.g. microsatellites or candidate gene markers quickly and cheaply (Howad et al. 2005). The Genome Scanning Approach using sets of spaced markers provides an efficient method of mapping major genes to existing framework maps (Erdin et al. 2006; Patocchi et al. 2004; Patocchi et al. 2005) and apple mapping teams internationally are using this technology as a replacement for bulked segregant analysis using RAPD markers.

Pedigree genotyping (Bink et al. 2002; van de Weg et al. 2004) offers a new approach to QTL mapping in fruit trees by tracing the QTLs across multiple generations of related breeding material derived from a set of common ancestors. This means that the majority of alleles in a breeding programme can be identified and exploited, and that QTLs are identified that otherwise would remain undetected in a single cross because of epistasis.

The extent of linkage disequilibrium (LD), the non-random association of alleles, in Rosaceae fruit crops is unknown. Most Rosaceae are cross-pollinating because of their self-incompatibility system and, as for other allogamous species such as maize (Rafalski and Morgante 2004) and loblolly pine (Brown et al. 2004), they are expected to have a low level of LD conservation. But there are exceptions: peach is self-compatible, and other species, such as apricot, almond and cherry have self-compatibility alleles that are extensively selected for in breeding programmes. Self-pollinating species genomes become nearly fixed after a few generations making most recombination events ineffective, resulting in a slower LD decay. Higher levels of LD have been found in selfers such as *Arabidopsis* (Nordborg et al. 2002) and barley (Caldwell et al. 2006). Another factor that increases LD is the occurrence of a recent bottleneck, as happened in most of the peach cultivars currently growing in North America and Europe (Scorza et al. 1985). Vegetative reproduction, the usual way to propagate fruit trees, may also favour the maintenance of LD, because the genotype of each cultivar is fixed and the number of generations from a common ancestor may be fewer than in species with annual sexual reproduction. In fact, the highest level of LD described so far was in sugarcane (Jannoo et al. 1999), a clonally reproducing species. The analysis of LD in the Rosaceae is likely to uncover different levels of LD in different species, and within different origins of the same species, depending on their reproductive characteristics and their history and evolution as cultivated species.

Detailed knowledge of the genetic structure of fruit tree variability and the LD will be useful for many applications in genetics and plant breeding. Association analysis in selected unrelated genotype collections of morphological variability and markers will detect genome positions containing genes of interest and will be useful for fine mapping of genes involved in quantitative traits and for identification of multiple alleles and allele combinations. Association between markers obtained

with candidate gene sequences and phenotypic data in genotype collections with low LD is a test that allows candidate genes that are in the region of interest but do not affect the character to be discarded, providing further evidence of cause and effect relationships in those that remain associated. Variability analysis of whole genome scans with markers in selected unrelated populations may detect the presence of regions of the genome that exhibit low variability, which may correspond to selective "sweeps" caused by directional selection. In some cases it is possible to make hypotheses on selective sweeps in certain genome locations, as for the self-incompatibility locus in peach, which has been fixed for self-compatibility alleles, or for the *M* locus in non-melting peaches because of fixation of the recessive allele that determines the non-melting behaviour. Unexpected results may question the current knowledge about the recent evolution of these characters. Additional selective sweeps may detect interesting regions, possibly corresponding to positions of genes selected during the process of fruit tree domestication. The selection of populations with different levels of LD based on their different histories may also allow a two-tiered strategy for gene analysis (Nordborg et al. 2002; Caldwell et al. 2006), where high LD collections are used for whole genome scans to search for the positions of major genes or QTLs, and lower LD collections allow the validation of specific candidate genes or the study of intragenic variability in connection with the observed phenotypic variability. Given the strong similarity between genomes within Rosaceae subfamilies, this approach can be attempted using genotype collections from different species (i.e. apple-pear or peach-almond).

Association analysis, particularly when performing whole genome scans, requires high throughput technology for SNP genotyping, as well as a large number of independent SNP markers. Now that there are hundreds of thousands of ESTs available from a broader base of cultivars, the way is open either for joint projects to develop and map thousands of SNP markers, or at a minimum, registration of interest in marker development at a central website to minimize duplication of effort.

The TILLING (targeting local lesions in genomes) method for genotyping (McCallum et al. 2000) was originally developed for detection of mutations induced by chemical agents such as ethylmethanesulphonate (EMS), but it may also be employed to genotype SNPs in natural populations ('ecoTILLING'), and may be used for association analysis in plants (Comai et al. 2004; Gilchrist and Haughn 2005).

Sequencing at least one of the Rosaceae genomes is of highest priority and given the high degree of similarity between species of the same subfamily (probably also between Prunoideae and Maloideae, although results so far are only partial), sequence information from one species should be useful for studies in others. Integrated genetic and physical maps will serve as sources for map-based cloning of genes controlling important traits. In-depth comparative mapping will reveal highly and poorly conserved genomic regions that respectively underlie similarities and differences between different species or genera. These will point to key areas for

targeted re-sequencing to reveal key associations involving important traits and to enable development of markers for MAS. New high capacity methods for targeted, in depth re-sequencing are becoming available and will enable microsynteny studies at the sequence level.

8.3. Regulation of Gene Expression

Much of the functionality of any genome is defined by the complex interactions among transcription factors (TFs) and genomic DNA (Gong et al. 2004). An initial study on the MYB family of TFs has indicated the value of investigating the functions of individual TFs. Full length sequencing and *in vivo* and *in vitro* expression analysis of 37 apple MYB transcription factor genes enabled speedy identification of a marker for red core in apple (see Section 7 above and Espley et al. 2006; Chagné et al. 2006a; Chagné et al. 2006b) and also indicated involvement in response to abiotic stress, pathogen response, temperature tolerance and branching (Allan et al. 2006).

Genomic approaches have also begun to advance the study of the classical plant growth regulators. Cloning of hormone receptors and identification of their mechanism of action has started to identify how these hormones regulate gene expression. For example, the auxin receptor has been shown to lead directly to degradation of transcriptional repressors by the SCF-mediated protein degradation pathway (Dharmasiri et al. 2005; Kepinski and Leyser 2005), indicating that plant hormones are directly involved in the regulation of gene expression. Recent work on the gibberellin receptor reviewed by Hartweck and Olszewski (2006) indicates degradation of DELLA proteins that in turn regulate gene expression. These close links between plant growth hormone and regulators of gene expression show that the classical plant hormones are an integral part of the regulatory networks. Genomic approaches that integrate manipulation of the hormone receptors and their downstream transcriptional regulators are likely to lead to control of key developmental traits in Rosaceae.

It has become clear that RNA is a key regulatory molecule in eukaryote cells and may contribute to the integration of expression of sets of genes (Baulcombe 2004). Small RNAs regulate target genes through sequence-specific mechanisms that may involve either gene silencing or the microRNA pathway. MicroRNA action may be through inhibition of transcription by methylation, or via gene silencing (Baulcombe 2004; MacDiarmid 2005). An understanding of which genes are regulated by small RNAs and the types of mechanism involved will be important for the ultimate development of systems to alter expression patterns in intact plants. An indication of the significance of the microRNA pathway is provided by the range of targets, which includes both transcription factors and protein degradation regulators, as well as genes involved in response to the environment, such as disease resistance genes (Baulcombe 2004). Bioinformatic analysis of the HortResearch EST database has enabled identification of several conserved primary small RNAs and their homologous target mRNAs. Direct isolation and cloning

has identified novel 19–25mers with many having putative target mRNAs in the database (C. A-Dwamena pers. comm.). Seven small RNA-related genes have been mapped on an apple framework map (S. Gardiner, D. Chagné, unpublished).

8.4. Microarrays

New microarray-based marker systems, such as single feature polymorphisms (SFPs) (Borevitz et al. 2003) and gene expression markers (GEMs) (West et al. 2006) that enable simultaneous genotyping and genome wide mapping of expression QTLs (eQTLs) in several systems, including higher plants, could provide a new tool for identifying candidate genes and analysing novel genetic interactions (Schadt et al. 2003; Kirst et al. 2005) as well as revolutionizing the scale of genetic mapping and MAS. The effectiveness of microarray technology has been demonstrated in principle for monitoring decrease in postharvest quality attributes in apple during storage (van Wordragen et al. 2003), and monitoring of gene expression appears to offer real possibilities for supporting decisions that need to be made during fruit storage and transport. Similar systems that could assist in making decisions during production, such as application of sprays to induce thinning, could be envisaged. A further application of microarray technology for border control purposes and for nurserymen, would be for the simultaneous detection of hundreds of viruses, viroids and phytoplasmas (Hadidi et al. 2004).

8.5. Introduction of Technologies into Breeding Programmes

The markers and technologies to introduce large scale MAS in fruit trees for some characters, are now available. An example is MAS for resistance genes into apple breeding programmes in order to reduce grower dependence on chemical control of pests and diseases, i.e. screening of breeding populations of tens of thousands of seedlings (see Sections 2 and 7 above). DNA can be extracted from 1,180 apple leaf samples overnight (Cook and Gardiner 2004; Cook et al. 2002) and the same number of PCR reactions processed and products analysed within the next 24 hours in a laboratory working with the HortResearch apple breeding programme (S. Gardiner, M. Cook, unpublished). As more markers are identified for QTLs and used for screening by breeders, targeted breeding for complex traits determining fruit quality will also become a reality.

The logical progression from simple MAS is “whole genome selection” (WGS), i.e. the use of markers to select the seedlings in a breeding population that possess a genome most like the quality grandparent(s), as well as the desired new character (Gardiner et al. 2006b; Pradhan et al. 2003). Fine mapping of resistance genes will provide markers to enable the simultaneous reduction of linkage drag in a very precise area around the introgressed genes. WGS is being investigated at HortResearch in combination with technology to decrease the time to first flowering. Preliminary trials have demonstrated that a seed-to-seed period of 18 months is possible in some genetic backgrounds when seedlings are grown under

a controlled environment regime with the timed application of a growth inhibitor (Austin et al. 2006). A transgenic approach to reducing the juvenile phase (Kotoda et al. 2006) may also be useful if applied to breeding parents.

8.6. Genetically Modified Food Products

In many countries the use of transgenic plants in food production is not currently accepted by regulatory authorities, and in any case is mistrusted by the consumer, with the result that transformation cannot be used for new variety development. For this reason, efforts have been made to introduce technologies that 'avoid or minimize the inclusion of superfluous transgenes or sequences' and marker-free transgenic apple plants have been obtained using a model vector that employed excision by recombination (Krens et al. 2004). In another study in apple, acceptable rates of transformation with a model construct (12–25% in preliminary experiments) have been achieved in the absence of a selectable marker (Malnoy et al. 2006d), and this technology may offer a route to future transgenic variety development utilizing endogenous apple genes. In addition, it has been proposed that cisgenic plants should be treated like traditionally bred plants by regulatory authorities (Schouten et al. 2006).

Some believe that transgenic rootstocks may be more acceptable to the consumer than transgenic scions and would provide a quicker route to new rootstock varieties than traditional breeding. Transformation with the *rolB* gene from *Agrobacterium rhizogenes* has been reported to enhance the rooting of pear and apple rootstocks (Zhu et al. 2001a; Zhu et al. 2001b; Zhu et al. 2003; Welander et al. 2004). However, the effect of such rootstocks on scion growth and fruit characters is yet to be reported. Other rootstock traits that could be manipulated with benefit to growers would be enhanced resistance to woolly apple aphid, *Phytophthora* and fireblight. An interesting possibility for developing apple cultivars with reduced allergenicity by using RNAi to silence allergen genes has been demonstrated in principle (Gilissen et al. 2005).

9. CONCLUSIONS

We have discussed ways in which genomics has started to support breeding of new varieties of Rosaceous tree crops, and how the applications of genomic technologies could be extended. Plans are being drawn up internationally , e.g. the Strategic Planning Document of the United States Rosaceae Genomics, Genetics and Breeding Initiative (http://www.mainlab.clemson.edu/gdr/community/rosexec/RosWP_March_2006.doc), for collective research strategies that will extend the impact of genomics on breeding for enhanced quality, shelf life and health enhancing attributes, lowered chemical pesticide inputs during production, as well as decreased labour and energy costs, and reduction of losses due to environmental stressors. In addition, the comparative studies across the wider Rosaceae family may prove a useful model system for answering fundamental questions for fruit crops, e.g. how

does the genome define the formation of a tree rather than a herbaceous plant?, and what determines the formation of a fleshy rather than a dehiscent fruit, or an aggregate versus a single fruit?

ACKNOWLEDGEMENTS

Writing of this review was partially supported by project AGL2003-04691/AGR funded by the Spanish Ministry of Education and by the New Zealand Foundation for Research, Science and Technology, programme C06X0207. The authors thank Bart Janssen and Anne Gunson for their critical reading of the manuscript and Sarah Fryer for her assistance in manuscript preparation.

REFERENCES

Allan AC, Bolitho K, Espley RV, Grafton K, Hellens RP, Lin-Wang K, Karunairetnam S, Gleave AP, Laing W (2006) The MYB transcription factors of apple: a family of genes involved in controlling a wide range of plant responses. 3rd international rosaceae genomics conference, 19–22 March, Napier. http://www.mainlab.clemson.edu/gdr/community/conferences/RG3_abstracts.pdf

Antofie A, Lateur M, Oger R, Patocchi A, Durel CE, Van de Weg WE (2006) Creation of a new versatile database for linking molecular and phenotypic information of apple (Malus × domestica Borkh): the HiDRAS 'AppleBreed Database'. 3rd international rosaceae genomics conference, 19–22 March, Napier. http://www.mainlab.clemson.edu/gdr/community/conferences/RG3_abstracts.pdf

Aranzana MJ, Pineda A, Cosson P, Dirlewanger E, Ascasibar J, Cipriani G, Ryder CD, Testolin R, Abbott A, King GJ, Iezzoni AF, Arús P (2003) A set of simple-sequence repeat (SSR) markers covering the *Prunus* genome. Theor Appl Genet 106:819–825

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

Arús P, Yamamoto T, Dirlewanger E, Abbott AG (2005a) Synteny in the Rosaceae. Plant Breed Rev Vol. 27. In: Janick J (ed) Plant breeding reviews, vol. 27. Wiley, New York, pp 175–211

Arús P, Howad W, Mnejja, M (2005b) Marker development and marker-assisted selection in temperate fruit trees. In: Tuberosa R, Phillips RL, Gale M (eds) Proceedings of the international congress. In the wake of the double helix: from the green revolution to the gene revolution. Avenue Media, Bologna, Italy, pp 309–325

Arús P, Ballester J, Villarroel A, Howad W (2005c) Marcadores moleculares en identificación varietal y mejora del melocotonero y otras especies *Prunus*: aplicaciones y potenciales. Fruticultura Profesional 152:47–52

Austin P, Norling C, Volz R, Bus V, Gardiner S (2006) Using controlled environments to accelerate flowering of Malus seedlings. 3rd international rosaceae genomics conference, 19–22 March 2006, Napier. http://www.mainlab.clemson.edu/gdr/community/conferences/RG3_abstracts.pdf

Baldi P, Patocchi A, Zini E, Toller C, Velasco R, Komjanc M (2004) Cloning and linkage mapping of resistance gene homologues in apple. Theor Appl Genet 109:231–239

Ballester J, Socias i Company R, Arús P, de Vicente MC (2001) Genetic mapping of a major gene delaying blooming time in almond. Plant Breed 120:268–270

Baulcombe D (2004) RNA silencing in plants. Nature 431:356–363

Belfanti E, Silfverberg-Dilworth E, Tartarini S, Patocchi A, Barbieri M, Zhu J, Vinatzer BA, Gianfranceschi L, Gessler C, Sansavini S (2004) The *HcrVf2* gene from a wild apple confers scab resistance to a transgenic cultivated variety. Proc Nat Acad Sci USA 101:886–890

Bielenberg DG, Wang Y, Fan S, Reighard GL, Scorza R, Abbott AG (2004) A deletion affecting several gene candidates is present in the evergrowing gene mutation. J Hered 95:436–444

Bielenberg DG, Fan S, Reighard GL, Abbott AG (2006) Sequencing and annotation of the evergrowing locus from wild type and mutant genomes reveals several candidate genes for the control of terminal bud formation in response to dormancy inducing conditions, Acta Hort 738:559–565

Bink MCAM, Uimari P, Sillanpää MJ, Janss LLG, Jansen RC (2002) Multiple QTL mapping in related plant populations via a pedigree-analysis approach. Theor Appl Genet 104:751–762

Bliss FA, Arulsekar S, Foolad MR, Becerra V, Gillen AM, Warburton ML, Dandekar AM, Kocsisne GM, Mydin KK (2002) An expanded genetic linkage map of *Prunus* based on an interspecific cross between almond and peach. Genome 45:520–529

Bolar JP, Norelli JL, Wong K-W, Hayes CK, Harman GE, Aldwinckle HS, Wong KW (2000) Expression of endochitinase from *Trichoderma harzianum* in transgenic apple increases resistance to apple scab and reduces vigor. Phytopathology 90:72–77

Bolar JP, Norelli JL, Harman GE, Brown SK, Aldwinckle HS (2001) Synergistic activity of endochitinase and exochitinase from *Trichoderma atroviride* (*T. harzianum*) against the pathogenic fungus (*Venturia inaequalis*) in transgenic apple plants. Transgenic Res 10:533–543

Borejsza-Wysocka EE, Malnoy M, Meng X, Bonasera JM, Nissinen RM, Kim JF, Beer SV, Aldwinckle HS (2004) Silencing of apple proteins that interact with DspE, a pathogenicity effector from *Erwinia amylovora*, as a strategy to increase resistance to fire blight. Acta Hort 1:469–473

Borejsza-Wysocka E, Malnoy M, Meng X, Bonasera JM, Beer SV, Aldwinckle HS (2006) Increasing resistance to Erwinia amylovora in apple by silencing apple DIPM genes. 3rd international rosaceae genomics conference, 19–22 March, Napier. http://www.mainlab.clemson.edu/gdr/community/conferences/RG3_abstracts.pdf

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

Bouvier L, Lespinasse Y, Schuster M (2000) Karyotype analysis of a haploid plant of apple (*Malus domestica*). Acta Hort 538:321–324

Brown GR, Gill GP, Kuntz RJ, Langley CH, Neale DB (2004) Nucleotide diversity and linkage disequilibrium in loblolly pine. Proc Natl Acad Sci USA 101:15255–15260

Bus V, Ranatunga C, Gardiner S, Bassett H, Rikkerink E (2000) Marker assisted selection for pest and disease resistance in the New Zealand apple breeding programme. Acta Hort 538 2: 541–547

Bus V, White A, Gardiner S, Weskett R, Ranatunga C, Samy A, Cook M, Rikkerink E (2002) An update on apple scab resistance breeding in New Zealand. Acta Hort 595:43–47

Caldwell KS, Russell J, Langridge P, Powell W, (2006) Extreme population-dependent linkage disequilibrium detected in an inbreeding plant species, Hordeum vulgare. Genetics 172:557–567

Calenge F, Durel C-E (2006) Both stable and unstable QTLs for resistance to powdery mildew are detected in apple after four years of field assessments. Mol Breed 17:329–339

Calenge F, Faure A, Drouet D, Parisi L, Brisset MN, Paulin JP, Van der Linden CG, Van de Weg WE, Schouten H, Lespinasse Y, Durel CE (2004a) Genomic organization of resistance factors against scab (*Venturia inaequalis*), powdery mildew (*Podosphaera leucotricha*) and fire blight (*Erwinia amylovora*) in apple. Biol Plant Microbe Interact 4:35–39

Calenge F, Faure A, Goerre M, Gebhardt C, Van de Weg WE, Parisi L, Durel C-E (2004b) Quantitative trait loci (QTL) analysis reveals both broad-spectrum and isolate-specific QTL for scab resistance in an apple progeny challenged with eight isolates of Venturia inaequalis. Phytopathology 94: 370–379

Calenge F, Drouet D, Denancé C, Van de Weg WE, Brisset M-N, Paulin J-P, Durel C-E (2005a) Identification of a major QTL together with several minor additive or epistatic QTLs for resistance to fire blight in apple in two related progenies. Theor Appl Genet 111:128–135

Calenge F, Van der Linden CG, Van de Weg E, Schouten HJ, Van Arkel G, Denance C, Durel C-E (2005b) Resistance gene analogues identified through the NBS-profiling method map close to major genes and QTL for disease resistance in apple. Theor Appl Genet 110:660–668

Celton J-M, Gardiner S, Rusholme R, Tustin S, Ambrose B (2006a) Pedigree analysis of apple rootstocks in relation to dwarfing. In: Mercer C (ed) Proceedings 13th Australasian plant breeding conference Christchurch, New Zealand, pp 645–650

Celton J-M, Rusholme R, Tustin S, Ward S, Ambrose B, Ferguson I, Gardiner S (2006b) Genetic mapping of *Dw1*, a locus required for dwarfing of apple scions by 'M.9' rootstock. 3rd international rosaceae genomics conference, 19–22 March, Napier. http://www.mainlab.clemson.edu/gdr/community/conferences/RG3_abstracts.pdf

Cevik V, King GJ (2002) Resolving the aphid resistance locus Sd-1 on a BAC contig within a subtelomeric region of *Malus* linkage group 7. Genome 45:939–945

Chagné D, Carlisle C, Volz R, Allan A, Espley R, Hellens R, Crowhurst R, Gardiner S (2006a) Mapping genes linked to red flesh in apple. In: Mercer C (ed) Proceedings 13th Australasian plant breeding conference Christchurch, New Zealand, pp 847–851

Chagné D, Carlisle C, Volz R, Allan A, Espley R, Hellens R, Crowhurst R, Gardiner S (2006b) SNP discovery in apple genes: application for red color. 3rd international rosaceae genomics conference, 19–22 March, Napier. http://www.mainlab.clemson.edu/gdr/community/conferences/RG3_abstracts.pdf

Chaparro JX, Werner DJ, O'Malley D, Sederoff RR (1994) Targeted mapping and linkage analysis of morphological isozyme, and RAPD markers in peach. Theor Appl Genet 87:805–815

Cheng FS, Weeden NF, Brown SK, Aldwinckle HS, Gardiner SE, Bus VG (1998) Development of a DNA marker for Vm, a gene conferring resistance to apple scab. Genome 41:208–214

Cheng L, Zhou R, Reidel EJ, Sharkey TD, Dandekar AM (2005) Antisense inhibition of sorbitol synthesis leads to up-regulation of starch synthesis without altering CO_2 assimilation in apple leaves. Planta 220:767–776

Chevalier T, de Rigal D, Mbéguié-Mbéguié D, Gauillard F, Richard-Forget F, Fils-Licaon BR (1999) Molecular cloning and characterization of apricot fruit polyphenol oxydase. Plant Physiol 119: 1261–1270

Chevreau E, Faize M, Dupuis F, Sourice S, Parisi L (2004) Combination of a transgene-mediated defense mechanism with a natural resistance gene increases apple scab resistance. Acta Hort 1:447–452

Claverie M (2004) Stratégie de clonage positionnel du gène *Ma* conférant la résistence aux nématodes du genre Meloidogyne chez le prunier myrobolan. Ph.D. Thesis. Ecole Superieure Agronomique de Montpellier (France)

Claverie M, Dirlewanger E, Cosson P, Bosselut N, Lecouls AC, Voisin R, Kleinhentz M, Lafargue B, Caboche M, Chalhoub B, Esmenjaud D (2004) High-resolution mapping and chromosome landing at the root-know nematode resistance locus Ma from Myrobalan plum using a large-insert BAC DNA library. Theor Appl Genet 109:1318–27

Comai L, Young K, Till BJ, Reynolds SH, Greene EA, Codomo CA, Enns LC, Johnson JE, Burtner C, Odden AR, Henikoff S (2004) Efficient discovery of DNA polymorphisms in natural populations by Ecotilling. Plant J 37:778–786

Conner PJ, Brown SK, Weeden NF (1997) Randomly amplified polymorphic DNA-based genetic linkage maps of three apple cultivars. J Am Soc Hort Sci 122:350–359

Cook M, Gardiner S (2004) Development of a fully automated system to extract DNA from difficult plant tissues for genomics research. Plant & animal genome XII conference, San Diego, CA, http://www.intl-pag.org/12/abstracts/

Cook MR, Xu P, Gardiner SE (2002) Development of an automated system for DNA extraction from leaf tissue. Projects 7:1–8

Costa F, Stella S, van de Weg WE, Guerra W, Cecchinel M, Dallavia J, Koller B, Sansavini S (2005) Role of the genes Md-ACO1 and Md-ACS1 in ethylene production and shelf life of apple (*Malus domestica* Borkh). Euphytica 141:181–190

Crowhurst RN, Allan AC, Atkinson RG, Beuning LL, Davey M, Friel E, Gardiner SE, Gleave AP, Greenwood DR, Hellens RP, Janssen BJ, Kutty-Amma S, Laing WA, MacRae EA, Newcomb RD, Plummer KM, Schaffer R, Simpson RM, Snowden KC, Templeton MD, Walton EF, Rikkerink EHA (2005) The HortResearch apple EST database – a resource for apple genetics and functional genomics. Plant & animal genome XIII conference, San Diego, CA, http://www.intl-pag.org/13/abstracts/

Crowhurst RN, Deng C, Davy M (2006) BioView – an enterprise bioinformatics system for automated analysis and annotation of non-genomic DNA sequence. 3rd international rosaceae genomics conference, 19–22 March, Napier. http://www.mainlab.clemson.edu/gdr/community/conferences/RG3_abstracts.pdf

Dandekar AM, Teo G, Defilippi BG, Uratsu SL, Passey AJ, Kader AA, Stow JR, Colgan RJ, James DJ (2004) Effect of down-regulation of ethylene biosynthesis on fruit flavor complex in apple fruit. Transgenic Res 13:373–384

Davey MW, Razavi F, Keulemans W (2006) Breeding functional apples; identification of QTL's for mean vitamin C contents of fruit skin and flesh. 3rd international rosaceae genomics conference, 19–22 March, Napier. http://www.mainlab.clemson.edu/gdr/community/conferences/RG3_abstracts.pdf

de Souza VAB, Byrne DH (1998) Heritability, genetic and phenotypic correlations, and predicted selection response of quantitative traits in peach: II. An analysis of several fruit traits. J Am Soc Hort Sci 123:604–611

Decroocq V, Foulongne M, Lambert P, Gall OL, Mantin C, Pascal T, Schurdi-Levraud V, Kervella J (2005) Analogues of virus resistance genes map to QTLs for resistance to sharka disease in *Prunus davidiana*. Mol Genet Genomics 272:680–689

Defilippi BG, Dandekar AM, Kader AA (2004) Impact of suppression of ethylene action or biosynthesis on flavor metabolites in apple (*Malus domestica* Borkh) fruits. J Agric Food Chem 52:5694–5701

Defilippi BG, Dandekar AM, Kader AA (2005) Relationship of ethylene biosynthesis to volatile production, related enzymes, and precursor availability in apple peel and flesh tissues. J Agric Food Chem 53:3133–3141

Degenhardt J, Al-Masri AN, Kürkcüoglu S, Szankowski I, Gau AE (2005) Characterization by suppression subtractive hybridization of transcripts that are differentially expressed in leaves of apple scab-resistant and susceptible cultivars of *Malus domestica*. Mol Genet Genomics 273: 326–335

Dharmasiri N, Dharmasiri S, Estelle M (2005) The F-box protein TIR1 is an auxin receptor. Nature 435:441–445

Diatchenko L, Lukyanov S, Lau Y-FC, Siebert PD (1999) Suppression subtractive hybridization: a versatile method for identifying differentially expressed genes. Methods Enzymol 303:349–380

Dirlewanger E, Moing A, Rothan C, Svanella L, Pronier V, Guye A, Plomion C, Monet R (1999) Mapping QTLs controlling fruit quality in peach (*Prunus persica* (L.) Batch). Theor Appl Genet 98:18–31

Dirlewanger E, Graziano E, Joobeur T, Garriga-Calderé F, Cosson P, Howad W, Arús P (2004a) Comparative mapping and marker-assisted selection in Rosaceae fruit crops. Proc Natl Acad Sci USA 101:9891–9896

Dirlewanger E, Cosson P, Howad W, Capdeville G, Bosselut N, Claverie M, Voisin R, Poizat C, Lafargue B, Baron O, Laigret F, Kleinhentz M, Arús P, Esmenjaud D (2004b) Microsatellite genetic linkage maps of myrobalan plum and an almond-peach hybrid – location of root-knot nematode resistance genes. Theor Appl Genet 109:827–838

Dirlewanger E, Kleinhentz M, Laigret F, Gómez-Aparisi J, Rubio-Cabetas MJ, Claverie M, Bosselut N, Voisin R, Esmenjaud D, Xyloyannis C, Dichio B, Poëssel JL, Di Vito M, Arús P, Howad W (2005) Breeding for a new generation of *Prunus* rootstocks based on marker-assisted selection: A European initiative. Acta Hort 663:829–833

Dominguez I, Graziano E, Gebhardt C, Barakat A, Berry S, Arús P, Delseny M, Barnes S (2003) Plant genome archeology: evidence for conserved ancestral chromosome segmentsin dicotyledonous plant species. Plant Biotech J 1:91–99

Dondini L, Pierantoni L, Gaiotti F, Chiondini R, Tartarini S, Bazzi C, Sansavini S (2004) Identifying QTLs for fire-blight resistance via a European pear (*Pyrus communis* L.) genetic linkage map. Mol Breed 14:407–418

Dreesen R, Vanholme B, Keulemans J (2006) Transcriptomics of ripening in apple as a tool to improve apple quality traits. 3rd international rosaceae genomics conference, 19–22 March, Napier. http://www.mainlab.clemson.edu/gdr/community/conferences/RG3_abstracts.pdf

Dunemann F, Urbanietz A, Gardiner S, Bassett H, Legg W, Rusholme R, Bus V, Ranatunga C (2005) Marker assisted selection for *Pl-1* powdery mildew resistance in apple – old markers for a new resistance gene? Acta Hort 663

Durel CE, Parisi L, Laurens F, Van de Weg WE, Liebhard R, Jourjon MF (2003) Genetic dissection of partial resistance to race 6 of *Venturia inaequalis* in apple. Genome 46:224–234

Durham RE, Korban SS (1994) Evidence of gene introgression in apple using RAPD markers. Euphytica 79:109–114

Erdin N, Tartarini S, Broggini GAL, Gennari F, Sansavini S, Gessler C, Patocchi A (2006) Mapping of the apple scab-resistance gene *Vb*. Genome 49:1238–1245

Espley RV, Hellens RP, Putterill J, Stevenson DE, Kutty-Amma S, Allan AC (2006) Red colouration in apple fruit is due to the activity of the MYB transcription factor, *MdMYB10*. Plant J (in press)

Etienne C, Rothan C, Moing A, Plomion C, Bodenes C, Dumas LS, Cosson P, Pronier V, Monet R, Dirlewanger E (2002) Candidate genes and QTLs for sugar and organic acid content in peach (*Prunus persica* (L.) Batsch). Theor Appl Genet 105:145–159

Evans RC, Campbell CS (2002) The origin of the apple subfamily (Maloideae; Rosaceae) is clarified by DNA sequence data from duplicated GBSSI genes. Am J Bot 89:1478–1484

Foulongne M (2002) Introduction d'une résistence polygénique à l'oïdium chez le pêcher *Prunus persica* à partir d'une espèce sauvage *Prunus davidiana*. PhD Thesis. Université de la Mediterranée-Faculté de Sciences de Marseille-Luminy

Foulongne M, Pascal T, Pfeiffer F, Kervella J (2003) QTLs for powdery mildew resistance in peach × *Prunus davidiana* crosses: consistency across generations and environments. Mol Breed 12:33–50

Frey JE, Frey B, Sauer C, Kellerhals M (2004) Efficient low-cost DNA extraction and multiplex fluorescent PCR method for marker-assisted selection in breeding. Plant Breed 123:554–557

Gardiner S, Murdoch J, Meech S, Rusholme R, Bassett H, Cook M, Bus V, Rikkerink E, Gleave A, Crowhurst R, Ross G, Warrington I (2003) Candidate resistance genes from an EST database prove a rich source of markers for major genes conferring resistance to important apple pests and diseases. Acta Hort 622:141–151

Gardiner SE, Bus V, Volz, R, Bassett, H (2006a) Marker assisted selection in apple breeding internationally. In: Mercer C (ed) Proceedings 13th Australasian Plant Breeding Conference Christchurch, New Zealand, pp. 681–686

Gardiner SE, Bus VGM, Rusholme RL, Chagné D, Rikkerink EHA (2006b) Apple. In: Kole C (ed) Genome mapping and molecular breeding in plants, Vol. 4, Fruits and Nuts, Springer, Berlin pp.1–62

Gasic K, Gonzales DO, Malnoy M, Thimmapuram J, Vodkin LO, Liu L, Aldwinckle HS, Carroll N, Orvis K, Goldsbrough P, Clifton S, Clifton L, Dante M, Hou S, Courtney W, Korban SS (2006) Analysis and functional annotation of an expressed sequence tag (EST) collection of apple (*Malus* × *domestica*). 3rd international rosaceae genomics conference, 19–22 March, Napier. http://www.mainlab.clemson.edu/gdr/community/conferences/RG3_abstracts.pdf

Georgi LL, Wang Y, Yvergniaux D, Ormsbee T, Iñigo M, Reighard G, Abbott AG (2002) Construction of a BAC library and its application to the identification of simple sequence repeats in peach (*Prunus persica* (L.) Batsch). Theor Appl Genet 105:1151–1158

Georgi LL, Wang Y, Reighard GL, Mao L, Wing RA, Abbott AG (2003) Comparison of peach and *Arabidopsis* genomic sequences: fragmentary conservation of gene neighborhoods. Genome 46:268–276

Gianfranceschi L (2006) HiDRAS: an innovative multidisciplinary EU-funded research project to breed high-quality disease resistant apples. Plant & animal genomes XIV conference, 14–18 January, San Diego, CA, http://www.intl-pag.org/14/abstracts/PAG14_W132.html

Gilchrist EJ, Haughn GW (2005) TILLING without a plough: a new method with applications for reverse genetics. Curr Opin Plant Biol 8:211–215

Gilissen LJWJ, Bolhaar STHP, Matos CI, Rouwendal GJA, Boone MJ, Krens FA, Zuidmeer L, Van Leeuwen A, Akkerdaas J, Hoffmann-Sommergruber K, Knulst AC, Bosch D, van de Weg E, van Ree R (2005) Silencing the major apple allergen Mal d 1 by using the RNA interference approach. J Allergy Clin Immunol 115:364–369

Gong W, Shen Y-P, Ma L-G, Pan Y, Du Y-L, Wang D-H, Yang J-Y, Hu L-D, Liu X-F, Dong C-X, Ma L, Chen Y-H, Yang X-Y, Gao Y, Zhu D, Tan X, Mu J-Y, Zhang D-B, Liu Y-L, Dinesh-Kumar SP, Li Y, Wang X-P, Gu H-Y, Qu L-J, Bai S-N, Lu Y-T, Li J-Y, Zhao J-D, Zuo J, Huang H, Deng XW, Zhu Y-X (2004) Genome-wide ORFeome cloning and analysis of *Arabidopsis* transcription factor genes1(w). Plant Physiol 135:773–782

Goulão L, Oliveira CM (2006) Molecular identification of novel differentially expressed mRNAs up-regulated during ripening of apples. Plant Sci 72:306–318

Granell A, Crisosto CH, Martí-Ibáñez C, Gradziel TM, Forment J, Peace C (2006) "CHILLPEACH" a functional database to understand peach chilling injury. 3rd international rosaceae genomics conference, 19–22 March, Napier. http://www.mainlab.clemson.edu/gdr/community/conferences/RG3_abstracts.pdf

Grimplet J, Romieu C, Audergon J-M, Marty I, Albagnac G, Lambert P, Bouchet J-P, Terrier N (2005) Transcriptomic study of apricot fruit (*Prunus armeniaca*) ripening among 13,006 expressed sequence tags. Physiol Plant 125:281–292

Hadidi A, Czosnek H, Barba M (2004) DNA microarrays and their potential applications for the detection of plant viruses, viroids, and phytoplasmas. J Plant Pathol 86:97–104

Haji T, Yaegaki H, Yamaguchi M (2005) Inheritance and expression of fruit texture melting, non-melting and stony hard in peach. Scientia Hort 105:241–248

Han Y, Gasic K, Marron B, Beever JE, Korban SS (2006a) Development of a genome-wide physical map of apple genome. 3rd international rosaceae genomics conference, 19–22 March, Napier. http://www.mainlab.clemson.edu/gdr/community/conferences/RG3_abstracts.pdf

Han Y, Gasic K, Xu M, Korban SS (2006b) Characterization of an SBE1 gene encoding starch branching enzyme I in apple. Plant & animal genomes XIV conference, 14–18 January. http://www.intl-pag.org/14/abstracts/PAG14_P498.html

Hartweck LM, Olszewski NE (2006) Rice GIBBERELLIN INSENSITIVE DWARF1 is a gibberellin receptor that illuminates and raises questions about GA signalling. Plant Cell 18:278–282

Hatsuyama Y, Igarashi M, Fukasawa-Akada T, Hashimoto A, Ohta T, Sato Y, Honda S, Kishimoto N, Kikuchi S, Suzuki M (2003) Monitoring effects of a host specific toxin on gene expression in apple leaves by DNA microarray. Plant & animal genomes XI conference, 11–15 January, San Diego, CA, http://www.intl-pag.org/11/abstracts/P7a_P776_XI.html

Hellens RP, Allan AC, Friel EN, Bolitho K, Grafton K, Templeton MD, Karunairetnam S, Gleave AP, Laing WA (2005) Transient expression vectors for functional genomics, quantification of promoter activity and RNA silencing in plants. Plant Methods 1:13

Hemmat M, Weeden NF, Manganaris AG, Lawson DM (1994) Molecular marker linkage map for apple. J Hered 85:4–11

Hemmat M, Weeden NF, Brown SK (2003) Mapping and evaluation of *Malus* × *domestica* microsatellites in apple and pear. J Am Soc Hort Sci 128:515–520

Horn R, Lecouls A-C, Callahan A, Dandekar A, Garay L, McCord P, Howad W, Chan H, Verde I, Main D, Jung S, Georgi L, Forrest S, Mook J, Zhebentyayeva T, Yu Y, Kim HR, Jesudurai C, Sosinski B, Arús P, Baird V, Parfitt D, Reighard G, Scorza R, Tomkins J, Wing R, Abbott AG (2005) Candidate gene database and transcript map for peach, a model species for fruit trees. Theor Appl Genet 110:1419–1428

Howad W, Yamamoto T, Dirlewanger E, Testolin R, Cosson P, Cipriani G, Monforte AJ, Georgi L, Abbott AG, Arús P (2005) Mapping with a few plants: using selective mapping for microsatellite saturation of the *Prunus* reference map. Genetics 171:1305–1309

Itai A, Kotaki T, Tanabe K, Tamura F, Kawaguchi D, Fukuda M (2003) Rapid identification of 1-aminocyclopropane-1-carboxylate (ACC) synthase genotypes in cultivars of Japanese pear (*Pyrus pyrifolia* Nakai) using CAPS markers. Theor Appl Genet 106:1266–1272

Jannoo N, Grivet L, Dookun A, D'Hont A, Glaszmann JC (1999) Linkage desequilibrium among modern sugarcane cultivars. Theor Appl Genet 99:1053–1060

Janssen B, Schaffer R, Thodey K, Bishop R, Bajaj S, Snowden K, Crowhurst R, Bowen J, Ledger S, Davy M, Dayatilake D, Ward S, McCartney S, Wunsche J (2006) Microarray analysis of fruit

development in apple. 3rd international rosaceae genomics conference, 19–22 March, Napier. http://www.mainlab.clemson.edu/gdr/community/conferences/RG3_abstracts.pdf

Jáuregui B, de Vicente MC, Messeguer R, Felipe A, Bonnet A, Salesses G, Arús P (2001) A reciprocal translocation between 'Garfi' almond and 'Nemared' peach. Theor Appl Genet 102:1169–1176

Jensen PJ, Rytter J, Detwiler EA, Travis JW, McNellis TW (2003) Rootstock effects on gene expression patterns in apple tree scions. Plant Mol Biol 53:493–511

Jensen PJ, Altman N, Crassweller R, Makalowska I, Maximova S, Praul C, Travis JW, McNellis TW (2006) Apple tree functional genomics: getting to the rootstock of it. Plant & animal genomes XIV conference, 14–18 January, San Diego, CA, http://www.intl-pag.org/14/abstracts/PAG14_W135.html

Joobeur T (1998) Construcción de un mapa de marcadores moleculares y análisis genético de caracteres agronómicos en *Prunus*. PhD thesis, Universtat de Lleida

Joobeur T, Viruel MA, de Vicente MC, Jáuregui B, Ballester J, Dettori MT, Verde I, Truco MJ, Messeguer R, Batlle I, Quarta R, Dirlewanger E, Arús P (1998) Construction of a saturated linkage map for Prunus using an almond x peach F2 progeny. Theor Appl Genet 97:1034–1041

Jung S, Jesudurai C, Staton M, Du Z, Ficklin S, Cho I, Abbot A, Tomking J, Main D (2004) GDR (Genome Database for Rosaceae): integrated web resources for Rosaceae genomics and genetics research. BioMed Central:1–8

Jung S, Abbott AG, Jesudurai C, Tomkins J, Main D (2005) Frequency, type, distribution and annotation of simple sequence repeats in Rosaceae ESTs. Funct Integr Genomics 5:136–143

Jung S, Main D, Staton M, Cho I, Zhebentyayeva T, Arús P, Abbott AG (2006) Synteny conservation between the *Prunus* genome and both the present and ancestral *Arabidopsis* genomes. BMC Genomics 7:81

Kellerhals M, Gianfranceschi L, Seglias N, Gessler C (2000) Marker-assisted selection in apple breeding. Acta Hort 521:255–265

Kenis K, Keulemans J (2005) Genetic linkage maps of two apple cultivars (*Malus* × *domestica* Borkh) based on AFLP and microsatellite markers. Mol Breed 15:205–219

Kepinski S, Leyser O (2005) The *Arabidopsis* F-box protein TIR1 is an auxin receptor. Nature 435:446–451

King GJ, Maliepaard C, Lynn JR, Alston FH, Durel CE, Evans KM, Griffon B, Laurens F, Manganaris AG, Schrevens E, Tartarini S, Verhaegh J (2000) Quantitative genetic analysis and comparison of physical and sensory descriptors relating to fruit flesh firmness in apple (*Malus pumila* Mill). Theor Appl Genet 100:1074–1084

King GJ, Lynn JR, Dover CJ, Evans KM, Seymour GB (2001) Resolution of quantitative trait loci for mechanical measures accounting for genetic variation in fruit texture of apple (*Malus pumila* Mill). Theor Appl Genet 102:1227–1235

Kirst M, Basten CJ, Myburg AA, Zeng Z-B, Sederoff RR (2005) Genetic architecture of transcript-level variation in differentiating xylem of a *Eucalyptus* hybrid. Genetics 169:2295–2303

Ko K, Norelli JL, Reynoird J-P, Aldwinckle HS, Brown SK (2002) T4 lyozyme and attacin genes enhance resistance of transgenic 'Galaxy' apple against *Erwinia amylovora*. J Am Soc Hort Sci 127:515–519

Kotoda N, Iwanami H, Takahashi S, Abe K (2006) Antisense expression of MdTFL1, a TFL1-like gene, reduces the juveline phase in apple. J Am Soc Hort Sci 131:74–81

Krens FA, Pelgrom KTB, Schaart JG, den Nijs APM, Rouwendal GJA (2004) Clean vector technology for marker-free transgenic fruit crops. Acta Hort 663:431–435

Lalli DA, Decroocq V, Blenda AVS, Levraud V, Garay L, Gall OL, Damsteegt V, Reighard GL, Abbott, AG (2005) Identification and mapping of resistance gene analogs (RGAs) in *Prunus*: a resistance map for *Prunus*. Theor Appl Genet 111:1504–1513

Lambert P, Faurobert M, Pelpoir E, Moreau K, Poëssel JL, Audergon JM (2004) Comparative mapping of *Prunus armeniaca*, *P. cerasifera* × *P. armeniaca* and *Prunus* reference map. Acta Hort 663: 91–94

Lazzari B, Caprera A, Vecchietti A, Stella A, Milanesi L, Pozzi C (2005) ESTree db: a tool for peach functional genomics. BMC Bioinformatics 6:516

Lebedev VG, Dolgov SV, Skryabin KG (2002) Transgenic pear clonal rootstocks resistant to herbicide "Basta". Acta Hort 596:193–197

Lee S-Y, Lee D-H (2005) Expression of *MbR4*, a TIR-NBS type of apple *R* gene, confers resistance to bacterial spot disease in *Arabidopsis*. J Plant Biol 48:220–228

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

Lester DR, Sherman WB, Atwell BJ (1996) Endopolygalacturonase and the melting flesh (*M*) locus in peach. J Am Soc Hort Sci 121:231–235

Liebhard R, Gianfranceschi L, Koller B, Ryder CD, Tarchini R, Van de Weg E, Gessler C (2002) Development and characterisation of 140 new microsatellites in apple (Malus × domestica Borkh). Mol Breed 10:217–241

Liebhard R, Kellerhals M, Pfammatter W, Jertmini M, Gessler C (2003a) Mapping quantitative physiological traits in apple (*Malus* × *domestica* Borkh). Plant Mol Biol 52:511–526

Liebhard R, Koller B, Gianfranceschi L, Gessler C (2003b) Creating a saturated reference map for the apple (Malus × domestica Borkh) genome. Theor Appl Genet 106:1497–1508

Liebhard R, Koller B, Patocchi A, Kellerhals M, Pfammatter W, Jermini M, Gessler C (2003c) Mapping quantitative field resistance against apple scab in a 'Fiesta' × 'Discovery' progeny. Phytopathology 93:493–501

Lin S (2005) Transcript profiling as a method to study fruit maturation, tree-ripening, and the role of "Tree Factor in 'Gala' and 'Fuji' apples" Faculty of the Graduate School University of Maryland, College Park, PhD Thesis

Liu Q, Ingersoll J, Owens L, Salih S, Meng R, Hammerschlag F (2001) Response of transgenic Royal Gala apple (Malus × domestica Borkh.) shoots carrying a modified cecropin MB39 gene, to *Erwinia amylovora*. Plant Cell Rep 20:306–312

López M, Mnejja M, Romero MA, Vargas FJ, Arús P, Batlle I (2005) Use of Sf-specific PCR for early selection of self-compatible seedlings in almond breeding. Options Méditerranéenes 63:269–274

Luby JJ, Shaw DV (2001) Does marker-assisted selecton make dollars and sense in a fruit breeding program? Hort Sci 36:872–879

MacDiarmid R (2005) RNA silencing in productive virus infections. Ann Rev of Phytopath 43:523–544

MacHardy WE (1996) Apple scab: biology, epidemiology, and management. APS Press, St Paul MN

Maliepaard C, Alston FH, van Arkel G, Brown LM, Chevreau E, Dunemann F, Evans KM, Gardiner S, Guilford P, van Heusden AW, Janse J, Laurens F, Lynn JR, Manganaris AG, den Nijs APM, Periam N, Rikkerink E, Roche P, Ryder C, Sansavini S, Schmidt H, Tartarini S, Verhaegh JJ, Vrielink-van Ginkel M, King GJ (1998) Aligning male and female linkage maps of apple (*Malus pumila* Mill.) using multi-allelic markers. Theor Appl Genet 97:60–73

Malnoy MA, Aldwinckle HS (2006) Development of fire blight resistance by recombinant DNA technology. In: Janick J (ed) Plant Breeding Reviews 29, Wiley, New York, USA, pp. 315–358

Malnoy M, Venisse JS, Brisset MN, Chevreau E (2003) Expression of bovin lactoferrin cDNA confers resistance to *Erwinia amylovora* in transgenic pear in through iron chelation. Mol Breed 12: 231–244

Malnoy M, Boresjza-Wysocka EE, Jin Q-L, He SY, Aldwinckle HS (2004) Over-expression of the apple gene MpNPR1 causes increased disease resistance in Malus × domestica. Acta Hort 663:463–467

Malnoy M, Boresjza-Wysocka E, Aldwinckle HS, Jin Q-L, He SY (2006a) Transgenic apple lines over-expressing the apple gene MpNPR1 have increased resistance to fire blight. Acta Hort 704:521–526

Malnoy M, Xu M, Borejsza-Wysocka EE, Korban SS, Aldwinckle HS (2006b) The role of Vfa RGA's at the Vf locus in resistance to *Venturia inaequalis*. 3rd international rosaceae genomics conference, 19–22 March, Napier. http://www.mainlab.clemson.edu/gdr/community/conferences/RG3_abstracts.pdf

Malnoy M, Reynoird JP, Borejsza-Wysocka EE, Aldwinckle HS (2006c) Activation of the pathogen-inducible *Gst1* promoter of potato after elicitation by *Venturia inaequalis* and *Erwinia amylovora* in transgenic apple (*Malus* × *domestica*). Transgenic Res 15:83–93

Malnoy M, Boresjza-Wysocka EE, Abbott P, Lewis S, Norelli JL, Flaishman M, Gidoni D, Aldwinckle HS (2006d) Genetic transformation of apple without use of a selectable marker. 3rd

international rosaceae genomics conference, 19–22 March, Napier. http://www.mainlab.clemson.edu/gdr/community/conferences/RG3_abstracts.pdf

Manganaris AG, Alston FH, Weeden NF, Aldwinckle HS, Gustafson HL, Brown SK (1994) Isozyme locus Pgm-1 is tightly linked to a gene (*Vf*) for scab resistance in apple. J Am Soc Hort Sci 119:1286–1288

Markwick NP, Docherty LC, Phung MM, Lester MT, Murray C, Yao JL, Mitra DS, Cohen D, Beuning LL, Kutty-Amma S, Christeller JT (2003) Transgenic tobacco and apple plants expressing biotin-binding proteins are resistant to two cosmopolitan insect pests, potato tuber moth and light-brown apple moth, respectively. Transgenic Res 12:671–681

Mattison H, Nybom H (2005) Application of DNA markers for detection of scab resistant apple cultivars and selections. Int J Hort Sci 11:59–61

McCallum CM, Comai L, Greene EA, Henikoff S (2000) Targeting induced local lesions in genomes (TILLING) for plant functional genomics. Plant Physiol 123:439–442

Meisel L, Vizoso P, Latorre M, Saba J, Loira N, Tittarelli A, Martínez V, Vargas C, Maldonado J, Caroca R, Bugueño M, Segovia S, Morales A, Silva H (2006) Bioinformatic advances of the Chilean nectarine functional genomics consortium. 3rd international rosaceae genomics conference, 19–22 March, Napier. http://www.mainlab.clemson.edu/gdr/community/conferences/RG3_abstracts.pdf

Morgante M, Hanafey M, Powell W (2002) Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. Nat Genet 30:194–200

Newcomb RD, Crowhurst RN, Gleave AP, Rikkerink EHA, Allan AC, Beuning LL, Bowen JH, Gera E, Jamieson KR, Janssen BJ, Laing WA, McArtney S, Bhawana N, Ross GS, Snowden KC, Souleyre EJF, Walton EF, Yauk Y (2006) Analyses of expressed sequence tags from apple. Plant Physiol 141:147–166

Nordborg M, Borevitz JO, Bergelson J, Berry CC, Chory J, Hagenblad J, Kreitman M, Maloof JN, Noyes T, Oefner PJ, Stahl EA, Weigel D (2002) The extent of linkage disequilibrium in *Arabidopsis thaliana*. Nat Genet 30:190–193

Norelli JL, Aldwinckle HS, Destéfano-Beltrán L, Jaynes JM (1994) Transgenic 'Malling 26' apple expressing the attacin *E* gene has increased resistance to *Erwinia amylovora*. Euphytica 77:123–128

Norelli JL, Korban SS, Volk GM, Zeng Z-B, Aldwinckle HS, Bassett CL, Beever J, Farrell RE, Gasic K Jr, Han Y, Kertbundit S, Marron B, Richards CM (2006) USDA-CSREES-NRI projects developing genomic resources for the Rosaceae (Malus). Plant & animal genome XIV conference. http://www.intl-pag.org/14/abstracts/PAG14_W133.html

Orellana A, Baeza R, Cambiazo V, Campos R, Defilippi B, González M, Meisel L, Retamales J, Silva H (2006) The Chilean peach functional genomics initiative, a progress report. 3rd international rosaceae genomics conference, 19–22 March, Napier. http://www.mainlab.clemson.edu/gdr/community/conferences/RG3_abstracts.pdf

Park S, Sugimoto N, Larson MD, Beaudry R, van Nocker S (2006) Identification of genes with potential roles in apple fruit development and biochemistry through large-scale statistical analysis of expressed sequence tags1. Plant Physiol 141:811–824

Patocchi A, Gianfranceschi L, Gessler C (1999a) Towards the map-based cloning of *Vf*: fine and physical mapping of the *Vf* Region. Theor Appl Genet 99:1012–1017

Patocchi A, Vinatzer BA, Gianfranceschi L, Tartarini S, Zhang HB, Sansavini S, Gessler C (1999b) Construction of a 550 kb BAC contig spanning the genomic region containing the apple scab resistance gene Vf. Mol Gen Genet 262:884–891

Patocchi A, Bigler B, Koller B, Kellerhals M, Gessler C (2004) *Vr2*: a new apple scab resistance gene. Theoretical and Applied Genetics 109:1087–1092

Patocchi A, Walser M, Tartarini S, Broggini GAL, Gennari F, Sansavini S, Gessler C (2005) Identification by genome scanning approach (GSA) of a microsatellite tightly associated to the apple scab resistance gene Vm. Genome 48:630–636

Peace CP, Crisosto CH, Gradziel TM (2005) Endopolygalacturonase: a candidate gene for Freestone and Melting flesh in peach. Mol Breed 161:21–31

Peace CP, Ogundiwin EA, Gradziel TM, Potter D, Weeks C, Badenes ML, Iezzoni, AF Bliss, FA, Crisosto, CH (2006a) Fruit softening in *Prunus*: progress and prospects of the

candidate gene approach. 3rd international rosaceae genomics conference, 19–22 March, Napier. http://www.mainlab.clemson.edu/gdr/community/conferences/RG3_abstracts.pdf

Peace CP, Abbott AG, Dai W, Iezzoni AF, Arus P, Baird WV, Callahan AM, Crisosto CH, Gradziel TM, Loescher W, Main D, Reighard G, Sosinski B, Tomkins J, van der Knaap E, Walla JA, Wang D (2006b) *Prunus* projects of the USDA CSREES national research initiative: synergies and progress. Plant & animal genome XIV conference, 14–18 January, San Diego, CA

Petri C, Burgos L (2005) Transformation of fruit trees. Useful breeding tool or continued future prospect. Transgenic Res 14:15–26

Pierantoni L, Cho K-H, Shin I-S, Chiodini R, Tartarini S, Dondini L, Kang S-J, Sansavini S (2004) Characterisation and transferability of apple SSRs to two European pear F1 populations. Theor Appl Genet 109:1519–1524

Pradhan AK, Gupta V, Mukhopadhyay A, Arumugam A, Sodhi YS, Pental D (2003) A high-density linkage map in *Brassica juncea* (Indian mustard) using AFLP and RFLP markers. Theor Appl Genet 106:607–614

Quilot B, Wu BH, Kervella J, Génard M, Foulongne M, Moreau K (2004) QTL analysis of quality traits in an advanced backcross between *Prunus persica* cultivars and the wild relative species *P. davidiana*. Theor Appl Genet 109:884–897

Quilot B, Kervella J, Génard M, Lescourret F (2005) Analysisng the genetic control of peach fruit quality through an ecophysiological model combined with a QTL approach. J Exp Bot 56:3083–3092

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

Rikkerink E, Hilario E, Rusholme R, Gardiner S, Bus V, Gleave A, Crowhurst R (2003) Mining the HortResearch apple EST database – in silico tissue expression analysis of resistance gene candidates and resistance gene classes. http://www.intl-pag.org/11/abstracts/P01_P14_XI.html

Rousseau M, Saint Oyant LH, Foucher F, Barrot L, Lalanne D, Sargent D, Simpson D, Laigret F, Denoyes-Rothan B (2006) 3rd international rosaceae genomics conference, 19–22 March, Napier. http://www.mainlab.clemson.edu/gdr/community/conferences/RG3_abstracts.pdf

Salesses G, Mouras A (1977) Tentative d'utilisation des protoplastes pour l'étude des chromosomes chez les *Prunus*. Ann Amelior Plantes 27:363–368

Schadt EE, Monks SA, Drake TA, Lusis AJ, Che N, Colinayo V, Ruff TG, Milligan SB, Lamb JR, Cavet G, Linsley PS, Mao M, Stoughton RB, Friend SH (2003) Genetics of gene expression surveyed in maize, mouse and man. Nature 422:297–302

Schaffer R, Friel E, Souleyre E, Janssen B, Thodey K, Bishop R, Davy M, Yao J-L, Cohen D, Newcomb R (2006) Microarray analysis of ripening in apple (cultivar Royal Gala). 3rd international rosaceae genomics conference, 19–22 March, Napier. http://www.mainlab.clemson.edu/gdr/community/conferences/RG3_abstracts.pdf

Schneider B, Berwarth C, Jelkmann W (2006) Engineering improved resistance against the fire blight pathogen in apple cultivars 'Elstar' and 'Royal Gala' by expression of human lactoferrin. Acta Hort 704:541–544

Schouten HJ, Krens FA, Jacobsen E (2006) Do cisgenic plants warrant less stringent oversight? Nat Biotechnol 24:753

Scorza R, Sherman WB (1996) Peaches. In: Janick J, Moore JN (eds) Fruit breeding, Wiley, New York, pp. 325–440

Scorza R, Mehlenbacher SA, Lightner GW (1985) Inbreeding and coancestry of freestone peach cultivars of the eastern United States and implications for peach germplasm improvement. J Am Soc Hort Sci 110:547–552

Silfverberg-Dilworth E, Patocchi A, Belfanti E, Tartarini S, Sansavini S, Gessler C (2005) *HcrVf2* introduced into Gala confers race-specific scab resistance. Plant and animal genome XIII conference, January, San Diego, CA, http://www.intl-pag.org/13/abstracts/

Silfverberg-Dilworth E, Matasci CL, Van de Weg WE, Van Kaauwen MPW, Walser M, Kodde LP, Soglio V, Gianfranceschi L, Durel CE, Costa F, Yamamoto T, Koller B, Gessler C, Patocchi A (2006) Microsatellite markers spanning the apple (*Malus* × *domestica* Borkh.) genome. Tree Genet Genomes 2:202–224

Silva C, Garcia-Mas J, Sánchez AM, Arús P, Oliveira MM (2005) Looking into flowering time in almond (*Prunus dulcis* (Mill) DA Webb): the candidate gene approach. Theor Appl Genet 110:959–968

Soglio V, Schouten H, Costa F, Gianfranceschi L (2006) Identification of genes with modulated expression during fruit development in *Malus* × *domestica* Borkh. 3rd international rosaceae genomics conference, 19–22 March, Napier. http://www.mainlab.clemson.edu/gdr/community/conferences/RG3_abstracts.pdf

Sonneveld T, Robbins TP, Tobutt KR (2006) Improved discrimination of self-incompatibility S-RNase alleles in cherry and high throughput genotyping by automated sizing of first intron polymerase chain reaction products. Plant Breed 125:305–307

Soriano JM, Vilanova S, Romero C, Llácer G, Badenes ML (2005) Characterization and mapping of NBS-LRR resistance gene analogs in apricot (*Prunus armeniaca* L). Theor Appl Genet 110: 980–989

Souleyre EJF, Greenwood DR, Friel EN, Karunairetnam S, Newcomb RD (2005) An alcohol acyl transferase from apple (cv Royal Gala), MpAAT1, produces esters involved in apple fruit flavor. FEBS J 272:3132–3144

Tatsuki M, Haji T, Yamaguchi M (2006) The involvement of 1-aminocyclopropane-1-carboxylic acid syntase isogene, *Pp-ACS1*, in peach fruit softening. J Exp Bot 57:1281–1289

Terakami S, Shoda M, Adachi Y, Gonai T, Kasumi M, Sawamura Y, Iketani H, Kotobuki K, Patocchi A, Gessler C, Hayashi T, Yamamoto T (2006) Genetic mapping of the pear scab resistance gene Vnk of Japanese pear cultivar Kinchaku. Theor Appl Genet 113:743–752

Trainotti L, Bonghi C, Ziliotto F, Zanin D, Rasori A, Casadoro G, Ramina A, Tonutti P (2006) The use of microarray microPEACH1.0 to investigate transcriptome changes during transition from pre-climacteric toclimacteric phase in peach fruit. Plant Sci 170:606–613

Ushijima K, Sassa H, Dandekar AM, Gradziel TM, Tao R, Hirano H (2003) Structural and transcriptional analysis of the self-incompatibility locus of almond: identification of a pollen-expressed F-Box gene with haplotype-specific polymorphism. Plant Cell 15:771–781

van de Weg WE, Voorrips RE, Finkers R, Kodde LP, Jansen J, Bink MCAM (2004) Pedigree genotyping: a new pedigree-based approach of QTL identification and allele mining. Acta Hort 663:45–50

van Nocker S, Sun L, Bukovac MJ, Fernandez-Lopez H (2006) Development of a genetic and molecular toolbox for the study of fruit abscission. Plant & animal genome conference, 14–18 January, San Diego, CA

van Wordragen M, Balk P, Hall R, Nijenhuis M, van den Broeck H, Vorst O, Poelman A (2003) Applied genomics – an innovative tool to improve quality in chains: predicting mealiness in apples – a case study. Acta Hort 604:387–394

Verde I, Lauria M, Dettori MT, Vendramin E, Balconi C, Micali S, Wang Y, Marrazzo MT, Cipriani G, Hartings H, Testolin R, Abbott AG, Motto M, Quarta R (2005) Microsatellite and AFLP markers in the [*Prunus persica* (L.) Batsch] × *P. ferganensis* BC1 linkage map: saturation and coverage improvement. Theor Appl Genet 111:1013–1021

Vilanova S, Romero C, Abbott AG, Llacer G, Badenes ML (2003) An apricot (*Prunus armeniaca* L.) F2 progeny linkage map based on SRR and AFLP markers, mapping plum pox virus resistance and self-incompatibility. Theor Appl Genet 107:239–247

Vilanova S, Romero C, Llacer G, Badenes ML, Burgos L (2005) Identification of self-(in)compatibility alleles in apricot by PCR and sequence analysis. J Am Soc Hort Sci 130:893–898

Vinatzer BA, Zhang HB, Sansavini S (1998) Construction and characterization of a bacterial artificial chromosome library of apple. Theor Appl Genet 97:1183–1190

Vinatzer BA, Patocchi A, Gianfranceschi L, Tartarini S, Zhang HB, Gessler C, Sansavini S (2001a) Apple contrains receptor-like genes homologous to the *Cladosporium fulvum* resistance gene family of tomato with a cluster of genes cosegregating with *Vf* apple scab resistance. Mol Plant Microbe Interact 14:508–515

Vinatzer BA, Patocchi A, Gianfranceschi L, Tartarini S, Zhang H-B, Gessler C, Sansavini S (2001b) Apple contains receptor-like genes homologous to the *Cladosporium fulvum* resistance gene family of tomato with a cluster of genes cosegregating with *Vf* apple scab resistance. Mol Plant Microbe Interact 14:508–515

Viruel MA, Madur D, Dirlewanger E, Pascal T, Kervela J (1998) Mapping quantitative trait loci controlling peach leaf curl resistance. Acta Hort 465:79–87

Vision TJ, Brown DG, Shmoys DB, Durrett RT, Tanksley SD (2000) Selective mapping: a strategy for optimizing the construction of high-density linkage maps. Genetics 155:407–420

Wang Q, Zhang K, Qu X, Jia J, Shi J, Jin D, Wang B (2001) Construction and characterization of a bacterial artificial chromosome library of peach. Theor Appl Genet 103:1174–1179

Wang Y, Georgi LL, Reighard GL, Scorza R, Abbott AG (2002) Genetic mapping of the evergrowing gene in peach (*Prunus persica* (L.) Batsch). J Hered 93:352–358

Wang C, Tian Y, Zhao J (2005) General application analysis of SSRs derived from apple (*Malus pumila*) on other species in Rosaceae. Acta Hort Sin 32:500–502

Wattebled F, Chevreau E, Durel CE, Laurens F (2006) Improving the knowledge of apple quality by functional genomics approaches. Perspectives at INRA Angers. 3rd international rosaceae genomics conference, 19–22 March, Napier. http://www.mainlab.clemson.edu/gdr/community/conferences/RG3_abstracts.pdf

Welander M, Zhu LH, Li XY (2004) Transformation of dwarfing apple and pear rootstocks with the rolB gene and its influence on rooting and growth. Acta Hort 1:437–442

Wesley SV, Helliwell CA, Smith NA, Wang M, Rouse DT, Liu Q, Gooding PS, Singh SP, Abbott D, Stoutjesdijk PA, Robinson SP, Gleave AP, Green AG, Waterhouse PM (2001) Construct design for efficient, effective and high-throughput gene silencing in plants. Plant J 27:581–590

West MAL, van Leeuwen H, Kozik A, Kliebenstein DJ, Doerge RW, Clair DA St, Michelmore RW (2006) High-density haplotyping with microarray-based expression and single feature polymorphism markers in *Arabidopsis*. Genome Res 16:787–795

Xu M, Korban SS (2002) A cluster of four receptor-like genes resides in the *Vf* locus that confers resistance to apple scab disease. Genetics 162:1995–2006

Xu M, Song J, Cheng Z, Jiang J, Korban SS (2001) A bacterial artificial chromosome (BAC) library of *Malus floribunda* 821 and contig construction for positional cloning of the apple scab resistance gene *Vf*. Genome 44:1104–1113

Xu M, Korban SS, Song J, Jiang J (2002) Constructing a bacterial artificial chromosome library of the apple cultivar goldrush. Acta Hort 595:103–112

Yamamoto T, Kimura T, Shoda M, Imai T, Saito T, Sawamura Y, Kotobuki K, Hayashi T, Matsuta N (2002) Genetic linkage maps constructed by using an interspecific cross between Japanese and European pears. Theor Appl Genet 106:9–18

Yamamoto T, Kimura T, Saito T, Kotobuki K, Matsuta N, Liebhard R, Gessler C, van de Weg WE, Hayashi T (2004) Genetic linkage maps of Japanese and European pears aligned to the apple consensus map. Acta Hort 663:51–56

Yamamoto Y, Yamaguchi M, Hayashi T (2005) An integrated genetic linkage map of peach by SSR, STS, AFLP and RAPD. J Jpn Soc Hort Sci 74:204–213

Yamamoto T, Terakami S, Nishitani C, Kimura T, Sawamura Y, Hirabayashi T, Hayashi T (2006) Genome mapping in pear. 3rd international rosaceae genomics conference, 19–22 March, Napier. http://www.mainlab.clemson.edu/gdr/community/conferences/RG3_abstracts.pdf

Zhebentyayeva T, Georgi L, Forrest S, Swire-Clark G, Mook J, Horn R, Jung S, Main D, Baird WV, Reighard G, Tomkins J, Abbott AG (2006) The peach physical/genetic map database: a tool for Rosaceae genomics. 3rd international rosaceae genomics conference, 19–22 March, Napier. http://www.mainlab.clemson.edu/gdr/community/conferences/RG3_abstracts.pdf

Zhu L, Ahlman A, Li X, Welander M (2001a) Integration of the *rol*A gene into the genome of the vigorous apple rootstock A2 reduced plant height and shortened internodes. J Hort Sci Biotechnol 76:758–763

Zhu L-H, Holefors A, Ahlman A, Xue Z-T, Welander M (2001b) Transformation of the apple rootstock M.9/29 with the *rol*B gene and its influence on rooting and growth. Plant Sci 160:433–439

Zhu L-H, Li X-Y, Ahlman A, Welander M (2003) The rooting ability of the dwarfing pear rootstock BP10030 (*Pyrus communis*) was significantly increased by introduction of the *rol*B gene. Plant Sci 165:829–835

CHAPTER 15

DNA MARKERS: DEVELOPMENT AND APPLICATION FOR GENETIC IMPROVEMENT OF COFFEE

PRASAD S. HENDRE AND RAMESH K. AGGARWAL*

Centre for Cellular and Molecular Biology (CCMB), Uppal Road, Hyderabad-500 007, India

Abstract: Coffee tree belongs to the genus *Coffea*, comprising of two main cultivated species *C. arabica* L. (the only tetraploid species with 2n = 4x = 44) and *C. canephora* Pierre ex A. Forehner (diploid, 2n = 2x = 22), yielding arabica and robusta types of coffee, respectively. In addition, there are ~100 diploid, wild, coffee species many of which hybridize readily with each other. These provide a rich and valuable source of desirable genetic variability for improvement of cultivated coffee germplasm. Arabica coffee is known for excellent cup quality but suffers from a narrow genetic base due to its domestication history and susceptibility for diseases and pests. In contrast, robusta coffee though poor in quality has better adaptability to various biotic/abiotic stresses. To meet the ever-changing demands of the environment and also of the sensibilities of market, there is a continuous need for genetic improvement of coffee, which unfortunately is severely constrained owing to inherently slow pace of tree breeding using conventional methods compounded with a general lack of genetic markers, screening and selection tools. The overall experience in India and worldwide from onerous conventional breeding efforts to develop new evolved coffee varieties has been rather frustrating with only few successes. The situation warrants recourse to newer, easy and efficient practical alternatives/technologies that can surmount the above problems and provide acceleration, reliability and directionality to the breeding efforts. In this context, development/utilization of: genomic variations based DNA markers, molecular linkage maps and markers-assisted breeding approaches that provide high-genetic resolution and new hopes and possibilities for genetic improvement of difficult species like coffee has become essential, a realization dawned on coffee research community rather recently. This chapter gives a brief overview of the worldwide efforts undertaken in recent years to develop, use and integrate DNA marker tools/technologies in coffee genetics research. It also provides few thoughts about future needs and perspectives to fully harness the potential of DNA marker based applications in managing and utilizing the available germplasm resources, construction of linkage maps, QTL mapping and genetic improvement of coffee.

*Author for correspondence: rameshka@ccmb.res.in

R.K. Varshney and R. Tuberosa (eds.), Genomics Assisted Crop Improvement: Vol. 2: Genomics Applications in Crops, 399–434.

1. INTRODUCTION

Coffee drink has become exceedingly popular since its age-old discovery as a drink in Arabia (ICO, London, 2006, http://dev.ico.org/). On a rough estimate, ~1.4 billion cups of this exotic beverage are consumed globally everyday for its distinctive flavour and stimulatory qualities. Though the discovery of coffee as a drink before 15th century is attributed to Arabs, coffee tree is supposed to have originated in the highlands of Ethiopia (Lashermes et al., 1996c). The Arabs, who got acquainted with coffee plant through their African slaves, cultivated coffee and safeguarded its cultivation by exporting only processed and nonviable coffee seeds thereby controlling the world coffee trade until the mid of 17th century (ICO, London, 2006). In order to break the Arab monopoly over coffee trade, in 1616, some Dutch people smuggled few coffee beans from Yemen and cultivated them in green houses in Holland. Mainly from this source, the Dutch and British people started cultivating coffee in their Asian, African and Caribbean/Latin American colonies. Thus coffee cultivation got established in its present places giving rise to different ecotypes, locally adopted forms, mutants and hybrids. As expected from this history and facts, today's cultivated coffee owes its origin to a handful of parental material introduced in the beginning to the present places of coffee cultivation thus exhibiting a very narrow genetic base (Ferwanda, 1976; Lashermes et al., 1996d). Today coffee is cultivated in about 80 countries in South and Central America, the Caribbean, Africa and Asia, in the areas generally lying between the tropics of Cancer and Capricorn with 70% of the coffee production coming from arabica and remaining from robusta coffee. By far, two biggest producers of coffee are Brazil and Colombia followed by Vietnam, Indonesia, Ethiopia, India and Mexico (ICO, London, 2006).

1.1. The Coffee Plant

Coffee tree belongs to the genus *Coffea*, comprising of two main cultivated species *C. arabica* L. (the only tetraploid species with 2n = 4x = 44) and *C. canephora* Pierre ex A. Forehner (diploid species having 2n = 2x = 22), and a third sparsely grown species *C. liberica* (or tree coffee) in some parts of Africa. In addition, there are >100 diploid, wild, coffee species and many species belonging to the related genus *Psilanthus* in the tribe *Coffeae* of *Rubiaceae* family that differ greatly in their size, morphology, adaptation and geographical distribution. Importantly, many of these diverse species hybridize readily with each other and thus are a rich and valuable source of desirable genetic variability for improvement of cultivated coffee germplasm.

C. arabica L., is supposed to have originated in the highlands of South-Western Ethiopia (Lashermes et al., 1996c) and *C. canephora* P. being originated in Central and West Africa in lowland forests of Eastern Liberia to Southern Kenya in Congo basin (Berthaud and Charrier, 1988). *C. arabica* is well adapted to highlands, whereas *C. canephora* and *C. liberica* thrives in lowland tropical areas. General plant features of two cultivated species of coffee are listed in Table 1.

Table 1. Plant features of *C. arabica* and *C. canephora*

	C. arabica	*C. canephora*
Growth habit	Shrubby, moderate vigour	Small tree, vigorous
Root system	Deep	Shallow
Ploidy level	Tetraploid (2n = 44)	Diploid (2n = 22)
Flowering time	Beginning of rainy season	Sporadic
Pollination	Autogamous (Carvalho, 1988)	Allogamous
Self incompatibility	Absent	Present (gametophytic)
Fruit maturation time	6–9 months	9–11 months
Fruit abscission at maturity	Yes	No
Yield potential	Lower	Higher
Optimal temperature	15–25°C	25–30°C
Optimal rainfall	1200–1500 mm/year	1500–3000 mm/year
Altitude grown	1000–2000 m	0–800 m
Disease and pest resistance	Low	High
Coffee flavour	Rich, acid	Bland, lacking acidity
Caffeine content	0.8 to 1.4%	1.7 to 4.0%
Price in global market	Higher	Lower

1.2. Economic Importance of Coffee and the World Coffee Crisis

Total global export of coffee exceeds US$ 9 billion and the sector employs more than 25 million people globally at different stages of production, processing and marketing of coffee (Kaplinsky, 2004). Although, the final consumption of coffee accounts for around US$55 billion annually, the coffee producing countries receive only ~US$8 billion (only around 15 percent of the total coffee trade). Coffee economy, nevertheless, remains one major export earner for undeveloped and developing countries (Cardenas, 2001) and is rather backbone of economy for many of them. It accounts for over half of the export earnings in countries like Burundi and Uganda and for some of the other coffee-producing third world countries, major share of revenue is derived from coffee exports (Wasserman, 2002; Table 2). Coffee was second largest commodity export after oil till late 1990s, when coffee prices started falling and reaching its 30 year lowest in 2001–2002 and now stands after oil, aluminium, wheat and coal (Ponte, 2001) and dragged the coffee world into the '*World Coffee Crisis*' (Cardenas, 2001; Osorio, 2004), which was a great economical, social and environmental loss and disaster for the coffee producing countries. The causes of crisis are deeply rooted into contemporary policies of world economy which involves globalisation/ liberalisation of trade coinciding with overproduction and imbalance in demand-supply (Cardenas, 2001) and of course in the laboratories and fields of coffee researchers including economists & policy makers, agronomists and geneticists & breeders as well.

Table 2. Importance of coffee in the economy of poor and underdeveloped countries (modified from Wasserman, 2002)

Country Name	Coffee as % of total export	Export % of GDP	Global position for coffee production (2004)
Burundi	56	8	28
Uganda	52	12	11
Rwanda	27	6	26
Ethiopia	26	16	5
Honduras	16	42	10
Nicaragua	15	37	19
Guatemala	15	20	9
Ecuador	11	42	18
Papua New Guinea	10	47	17
El Salvador	9	28	14

2. COFFEE GENETIC IMPROVEMENT

Domestication of the species has been characterized by successive reduction in genetic diversity (Etienne et al., 2002), which is further constrained especially in arabica by its primarily self-fertilizing reproduction behaviour leading to genetic homogenization. However, by using traditional breeding tools like introductions/explorations, selections, hybridisations, back crossing and hybrids, breeders have been able to create new and improved varieties for both arabica and robusta coffee.

Two botanical types are recognised each in *C. arabica* (Typica and Bourbon) and *C. canephora* (Robusta and Ngnda). Typica was the original arabica coffee taken to Asia and then to Latin America/Caribbean by the Dutch people. Bourbon coffee was identified as a coffee plant with smaller bean size found on the Bourbon island (now renamed as Reunion) in the Indian Ocean, and was later taken to South Americas and from there to East Africa to be cultivated as Bourbon variety. Bourbon is high yielding, slender variety lacking the bronze cast of the new foliage with smaller bean size and better resistance to berry disease than Typica. On the other hand, two botanical types of *C. canephora* are: 'Robusta or Erecta' (up right form) and 'Nganda' (spreading form).

Breeding efforts in early to mid nineties led to development of many improved arabica selections. Some prominent local selections (Sl) for arabica coffee are Blue Mountain in Jamaica, Vila Sarchi, Columnaris, Coorg, Kent and S 795, French Mission, K- 7, Sl 6, Sl 14, Sl 28, Sl 34 etc. Apart from this, many explorations (FAO, 1968; Guillaumet and Hallé, 1978) undertaken in Africa to search for novel diversity have added sources of disease, pest, nematode and abiotic resistance. Inter-varietal crosses have yielded hybrids such as Mundo novo, Catuai etc. Similarly, several spontaneous natural mutants showing desirable characteristics like Caturra, Purpurascens, San Ramon, Pache comum, Maragogype are identified and now being cultivated as well as used for varietal improvement. *C. arabica* being self pollinated, these varieties tend to be genetically more stable. Although, now a number

of high yielding arabica varieties are available, but in general these are highly susceptible to various diseases like leaf rust (*Hemileia vastatrix* Berk & Br.), Coffee Berry Disease (CBD caused by *Colletotrichum kahawae*), stem borer (*Xylotrechus quadripes* Chevrolat) and nematodes (*Meloidogyne* spp. and *Pratylenchus* spp.).

In comparison to arabica coffee, there have been limited improvement efforts for robusta coffee. As this coffee plant possesses gametophytic self-incompatibility, it produces many forms and varieties in wild. Modern robusta cultivars are BR series in India, SA and BP selections in Indonesia, IF clones in Ivory Coast and IAC lines and Apoata (IAC 2258) in Brazil (Vossen, 2004). Recently, some programmes have been initiated for genetic improvement of robusta coffee (Montagnon et al., 2004; Anim-Kwapong and Adu-Ampomah, 2004). Apart from an example of the C×R hybrid developed in India between *C. congensis* and *C. canephora*, diploid interspecific hybrids were not a great success (Vossen, 2004). But none the less, robustas remain important as they provide the most needed genetic diversity including high yield potential, plant stature, hardy nature and resistance to various diseases, pests and abiotic stresses.

For incorporating disease resistance, sturdy stature and high yielding quality of robusta into arabicas with excellent cup quality, several arabica x robusta hybrids were also attempted (Carvalho, 1988; Fazuoli at al., 2001). A natural tetraploid hybrid of *C. arabica* and *C. canephora* termed as Hybrido-de-Timor (HdeT) identified from the erstwhile Portuguese colony of East Timor Island (Bettencourt, 1973), till todate is probably the only best known source of resistance to leaf rust, some species of nematodes and coffee berry disease in arabica. Finding natural hybrids is a common phenomenon in many coffee plantations as well as in research stations (Cramer, 1957) and similar to HdeT, S-26 is another source of rust resistance that was found in India. Such natural hybrids provide important sources for desirable gene(s) and accordingly have been employed in breeding with varying success. Recently a sufficiently fertile intergeneric hybrid was produced between *Psilanthus ebracteolatus* Hiern and *C. arabica* L. suggesting that generic barriers could also be overcome in coffee for introgression purpose (Coutoron et al, 1998).

2.1. Present Status of Conventional Coffee Breeding Efforts and Need for Newer Molecular Tools

In today's scenario, thrust areas for coffee improvement programmes include enlarging the genetic base of arabica coffee, incorporation of disease, pest and nematode resistance (due to growing incidences of new diseases, pathotypes and lesser known diseases in new areas), abiotic stress tolerance (due to global climatic changes, deteriorating soil conditions and changing patterns of rainfall), encouragement for development of specialty, certified and branded coffees (like organic coffees, fair trade coffee, estate coffees etc.) and diversification of coffee plantations including quality, environmental stability and sustainable yield improvement (Herrera et al., 2001; Vossen, 2004). To meet these ever-changing demands of the environment and also of the sensibilities of market, a number of breeding and

selection programs have been undertaken worldwide including India since early 20th century. Despite few successes, overall experience from these long years of conventional breeding efforts worldwide has been rather frustrating, mainly because the pace of progress in coffee improvement has been abysmally slow. This dismal situation have been the outcome of inherently slow conventional breeding methods that are further compounded in case of coffee being a tree species that has its own problems. Some of the major constrains in coffee breeding, thus, have been:

1. Very long generation cycle; 4–6 years for seed to seed generation.
2. Narrow genetic base for arabicas.
3. Unavailability of true-to-breed type inbreds or homozygous lines for the diploid genotypes due to difficulty in selfing of plants.
4. Ploidy level variation (arabicas are tetraploids and robustas are diploids) and incompatibility barriers (robustas possess gametophytic self incompatibility system).
5. Reproductive barriers for interspecific crossing due to infertility, low frequency of intergenome crossing over and gametic or zygotic counter selection (Herrera et al., 2001, 2002a).
6. Poor knowledge base of genetics of coffee traits, unavailability of genetic markers/genetic maps, dense DNA marker based linkage maps with transferable landmarks across laboratories.
7. Unavailability of suitable screening tools/diagnostic markers for diseases and pathogens in correlation to continuous evolution of new pathotypes.
8. Relatively poor research base/resources in most of the producing countries.
9. Treatment as orphan species compared to other crop plants, leading to a delayed (relatively recent) efforts to integrate modern biology genomic tools/technologies for coffee research and improvement.

The situation called for newer, easy and efficient practical alternatives that could surmount the above problems faced in coffee improvement, a realization dawned on coffee research community rather recently. The advent of DNA variations based genetic markers and breeding approaches like MAS provided new hopes and possibilities for genetic improvement programs of difficult species like coffee. The great potential of DNA marker based technologies is now well demonstrated, and these are being utilized for crop improvement programmes through out the world for genotyping, varietal identification (and claiming intellectual property rights), germplasm finger printing, construction of linkage maps, Quantitative Trait Loci (QTL) identification and mapping in concert with Marker Assisted Selection (MAS) breeding to finally develop genetically improved crops with desirable traits. The DNA marker technologies were found equally useful in helping decipher many of the evolutionary puzzles about origin, spread, taxonomic relationships of many crop species. In recent years, therefore, conscious efforts have began globally to integrate such DNA/molecular markers based technologies, which can provide impetus, dependability and directionality to the efforts on genetic improvement of coffee. Accordingly, the coffee research is now experiencing many of these technologies evolve, flourish and to be utilized in future, but yet to go a long way before visible

gains become a reality. There are now large coffee genomics programs underway in many countries, few prominent being, Brazil, France, Italy, Columbia, and more recently India (where a large Coffee Genomic initiative has been started in 1999 under a National Network mode supported by Department of Biotechnology, Govt. of India).

In the following sections of this paper, an essence of global efforts put on for exploring and development of DNA markers for coffee, applications of DNA markers for deciphering genetic make up of coffee plant, deciphering origin of coffee plant, coffee germplasm assessment, construction of linkage maps, and QTL mapping are discussed.

3. DEVELOPMENT OF DNA MARKERS FOR GENETIC STUDIES IN COFFEE

With the advancement of molecular biology technologies, initial efforts with biochemical markers such as isozymes were soon replaced by DNA based approaches due to their apparent advantages for genetic studies in plants and animals. In the beginning, southern-based RFLP markers (Wyman and White, 1980) were widely used but were found to be difficult to adopt for large-scale studies that are hallmark of any worthwhile genetic study on genetic improvement. Therefore, soon with the advent of PCR technology in 1980s, a suite of PCR based marker approaches such as RAPD (Williams et al., 1990), AFLP (Vos et al., 1995), SSR (Weber and May, 1989; Litt and Luty, 1989; Tautz, 1989) and more recently SNP (Sachidanandam et al., 2001) based markers, are being developed and adopted for genetic studies worldwide. In case of coffee research, a similar (but modest) trend is seen in using the molecular markers to complement the conventional low-efficiency biometrical tools for genetic studies. Some of the initial studies in this regard include coffee germplasm assessment studies using markers like isozymes (Berthou and Trouslot, 1977; Moreno, 1990), RFLPs (Lashermes et al., 1999, 2000b; Hererra et al., 2002a) followed by multilocus RAPD (Lashermes et al., 1993, 1996a, 1996d; Agwanda et al., 1997; Anthony et al., 2001) and AFLP (Steiger et al., 2002; Anthony et al., 2002; Lashermes et al., 2000a). In the last decade, few efforts have been initiated to develop coffee specific genomic SSR markers that have resulted in a small repertoire of such markers comprising 183 publicly available SSRs (Table 3). In addition, there are 306 SSRs having limited access only through registration (www.coffeedna.org), and 170 SSRs developed by us (Aggarwal et al., 2004a) under the first Indian initiative on coffee genomics and are available on request. Here it is noteworthy to mention that our studies done to evaluate the occurrence and relative abundance of different SSRs have shown that coffee genome (across its different species), in general, is relatively very poor in SSRs with abundance of AT rich repeats (Figure 1, our unpublished data), thus suggesting it to be a difficult system for development of species-specific microsatellite markers. Recently, attempts have also been initiated to develop EST-SSRs (genic/functional SSR markers)

Table 3. Summary of publicly accessible coffee specific SSR primer pairs

Source reference	No. of working sets of primers published	Source species	Comments
Combes et al. (2000)	11	*C. arabica* (Caturra)	From library enriched for TG motif
Rovelli et al. (2000)	13	*C. arabica* (Caturra)	From two libraries each enriched with ATC and TG motif
Coulibaly et al. (2003b)	20	*C. arabica*	From available SSR containing sequences
Baruah et al. (2003)	9	*C. arabica* (HdeT)	From normal small-insert partial genomic library
Moncada and McCouch (2004)	34	*C. arabica* (Caturra)	From library enriched for GA motif
Poncet et al. (2004)	78	*C. arabica*	From available database sequences
Bhat et al. (2005)	9	CxR hybrid	From *in-house* EST library
Aggarwal et al. (2007)	16 (+ 200 designed primer sets)	*C. canephora* varieties (CxR hybrid and S-12)	From *in-house* EST library and available EST database
Total publicly available SSR primer pairs till date			183* (+ 200 designed primer sets)
Total SSRs mapped on different linkage maps *(Many of these markers are not in public domain but are available through mutual agreement)*			18 (Lashermes et al., 2001) 29 (Coulibaly et al., 2003b) 71 (Hendre, 2006) Total = 138

* In addition we have developed 170 coffee SSRs that are available on request (Aggarwal et al., 2004a)

utilizing coffee EST sequences (Bhat et al., 2005; Aggarwal et al., 2007). These new generation EST-SSR markers require availability of large EST databases as a prerequisite, which were severely limited for coffee till late 2005 (only ~400 ESTs were in public domain NCBI database). Utilizing the limited EST resources at hand, we have now designed 224 EST-SSR primer pairs, and have demonstrated

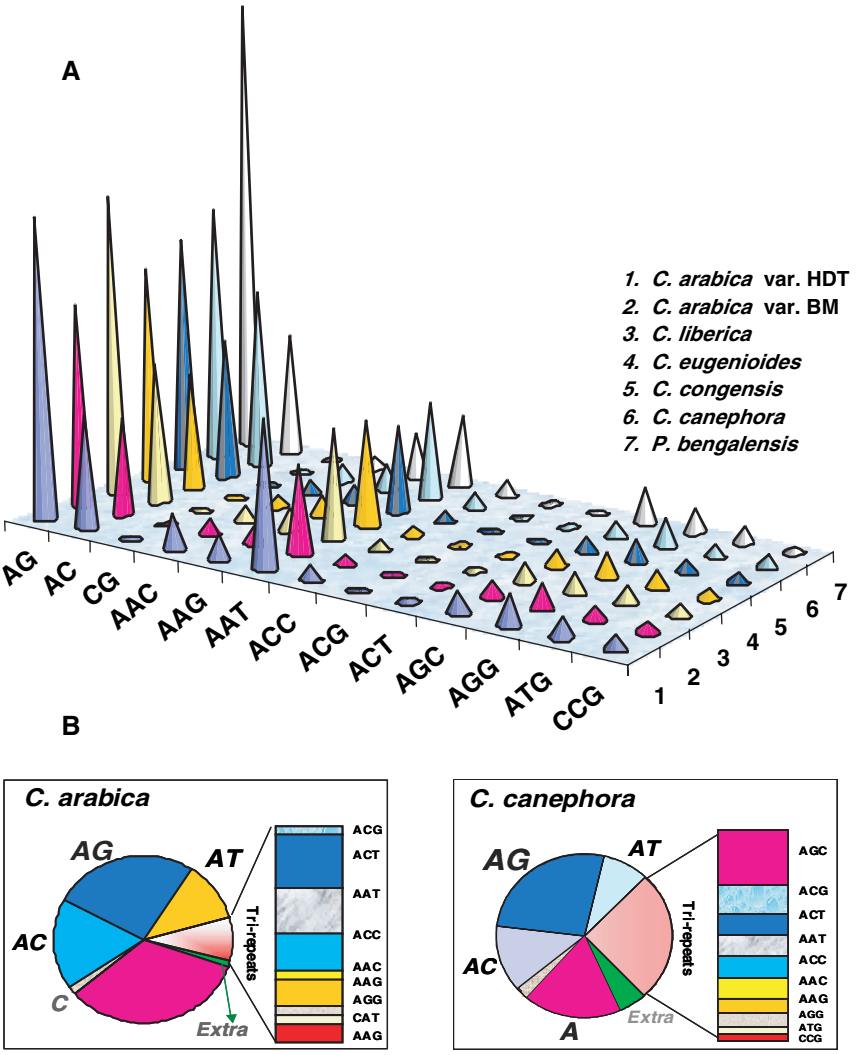


Figure 1. Relative abundance of different SSR motifs in genomes of different coffee species as revealed by: A) semi-quantitative Southern hybridization based slot-blot analysis, and B) *in-silico* sequence analysis of >1000 SSR positive clones from *C. arabica* and *C. canephora* (robusta) small-insert genomic libraries. Note almost similar comparable pattern/relative frequencies of different SSRs across coffee species and also between the two approaches of evaluation (our unpublished data)

their scope and utility in characterization of *C. arabica*, *C. canephora* germplasm and broad cross-species transferability (Aggarwal et al., 2007).

Despite these random successes with marker development, it will be important rather essential to expand the repertoire of coffee specific PCR based efficient SSR/microsatellite markers before we can really hope to integrate and realize the great potential of marker based approaches in research on coffee genetics and plant breeding/improvement programs. In this context, the recent addition of ~47,000 coffee ESTs with more than 13,000 unigenes (Lin et al., 2005) to NCBI database is a very fortunate development, which promises possibilities of many more coffee SSR markers. Similarly, the concept proposal to undertake low coverage genome sequencing of arabica coffee as a potential source to develop large number of SSR markers under the preview of recently enacted 'International Coffee Genomic Network' (2nd meeting at Trieste, Italy; www.coffeegenome.org) is a welcome step that if successful, would immensely help to hasten the realization of markers based molecular breeding in coffee.

4. APPLICATIONS OF DNA MARKERS FOR GENETIC DIVERSITY STUDIES IN COFFEE GENEPOOL

The coffee genepool consists of primary germplasm comprising wild and cultivated varieties of *C. arabica* and *C. canephora* and secondary genepool comprising >100 related diploid species of *Coffea* and *Psilanthus*. Cultivated arabica varieties have excellent cup quality but possess a narrow genetic base and susceptibility for diseases and pests. In contrast *C. canephora* varieties (robustas), generally are poor in quality. The secondary gene pool including robustas provides an important source for incorporating disease resistance, sturdy stature and high yielding quality. The diploid *Coffea* varieties are known to interbreed freely with each other and produce relatively fertile progeny (Anthony, 1992; Louarn, 1993) as well as fertile intergeneric hybrids can be produced (Coutoron et al., 1998). These attributes thus, demonstrate the importance of primary/secondary genepool in coffee improvement but also suggest the need for thorough exploration of the available germplasm resources for useful variability using high resolution genetic approaches for any meaningful utilization. Some of the major studies carried out using different types of DNA markers to ascertain the genetic diversity in the available coffee germplasm are briefly summarized in the Tables 4 and 5, and are discussed ahead.

4.1. Genetic Diversity in Cultivated Genepool of Coffee

The cultivated genepool of arabica is believed to have been derived from few seeds, which survived the efforts to bring coffee out of Yemen to its present places of cultivation as become apparent from the historical evidences and facts (Vossen, 1985; Anthony et al., 2001, 2002). The genetic variability in arabica populations is expected to have reduced further owing to its autogamous behaviour, leading

Table 4. DNA markers based molecular studies performed for germplasm assessment and species status of *Coffea* and related taxa

Study source/ Reference	Techniques used	Germplasm studied	Important results	Phylogeny grouping and relationship
Lashermes et al. (1993)	RAPD (23 decamers)	Arabica (6 acc.), HdeT robusta (5 acc.), *C. eugenioides*, *C. congensis*, *C. liberica* (2 acc.), *C. pseudozanguebariae*, *C. resinosa*, *C.* sp. and *C. stenophylla* (1 acc. each) *Total = 19*	Very low variation in arabicas, high intraspecific variation in robusta and liberica, extensive genetic variation between *Coffea* spp.	I = *C. canephora*, *C. congensis*, *C. liberica* and *C. stenophylla* II = *C. arabica*, *C. eugenioides* III = *C. resinosa* IV = *C. pseudozanguebariae* Clusters according to geographical locations.
Orozco-Castillo et al. (1994)	RAPD (30 decamers)	Cultivated arabica (10 acc.), wild arabica (7 acc.), robusta (3 acc.), HdeT and its derivatives (6 acc.), *C. liberica* (1 acc.) *Total = 27*	High diversity exists in the assessed germplasm, species-specific markers observed.	I = Bourbon and its derivatives II = Typica and its derivatives III = Ethiopian collections IV = robustas V = *C. liberica*
Orozco-Castillo et al. (1996)	RAPD (15 decamers)	Arabica (4 acc.), robusta (3 acc.), *C. eugenioides* (2 acc.), *C. liberica* (2 acc.), *C. stenophylla*, *C. racemosa*, *C. pseudozanguebariae*, *C. humilis*, *C. congensis*, *C. sessiflora*, *C. breviceps* (1 acc. each) *Total = 18*	Intraspecific variation in arabica, robusta, liberica. Extensive variation among *Coffea* spp.	A = *C. pseudozanguebariae, C. sessiflora, C. racemosa* B = *C. canephora*, *C. liberica*, *C. breviceps*, *C. congensis* C = *C. humilis*, *C. arabica*, *C. eugenioides*, *C. stenophylla*
Lashermes et al. (1996d)	RAPD (140 decamers)	Cultivated arabica (7 acc.), Wild arabicas (13 acc.) *Total = 20*	Very low variation (Only 15 polymorphic bands from 12 primers).	I = cutlivated arabicas. Kenyan wild arabicas II = wild arabicas, high diversity among wild than cultivated Clusters according to geographical location.

(Continued)

Table 4. (Continued)

Study source/ Reference	Techniques used	Germplasm studied	Important results	Phylogeny grouping and relationship
Agwanda et al. (1997)	RAPD (272 decamers)	Coffee Berry Disease resistant (high, medium and normal) and susceptible (high and normal) arabica genotypes *Total = 45*	Very low polymorphism (5.2% bands are polymorphic) and 3 markers linked to CBD resistance.	Cultivars specific groupings.
Lashermes et al. (1999)	RFLP (23 probes)	Arabicas (3 acc.), robustas (3 acc.), *C. congensis* (2 acc.), *C. eugenioides* (1 acc.), *C. sp.* Moloundou (1 acc.) *Total = 10*	Extermely low diversity in arabicas with some fixed heterozygosity. Many alleles in arabica were shared by canephoroid and'/or eugenioides species.	I = canephoras II = arabicas III = *C. congensis* IV = *C. eugenioides*
	GISH (fluorescence probes: *C. canphora* + *C. eugenioides)*	*C. arabica*, *C. canephora*, *C. eugenioides*	Detected 22 chromosomes each belonging to each constituent genomes*C. canphora* and *C. eugenioides.*	–
Lashermes et al. (2000a)	AFLP (42 primer pairs) canephora gene introgression into HdeT and its derivatives	Cultivated arabica (17 acc.), wild arabicas (6 acc.), introgressed arabicas (19 acc.), HdeT (2 acc.) cultivated robusta (3 acc.), wild robusta (5 acc.) *Total = 52*	No. of polymorphic bands in robustas (945) 9 times higher than arabicas (109), 50% introgression from robusta into HdeT.	Introgressed arabicas weakly grouped together, strong clustering of cultivated and wild robustas.
Lashermes et al. (2000b)	RFLP (16 single copy probes)	Segregating population of 14 F2 from F1 arabica hybrid plants and 70 BC plants generated from arabusta and arabica hybrid	3/16 probes polymorphic for arabicas. 11/16 probes polymorphic for arabusta. Confirm diploidised behaviour of arabica.	–

Combes et al. (2000)	SSR (11 primer pairs)	Arabica (32 acc.), robusta (10 acc.), 13 *Coffea* taxa and 2 *Psilanthus* taxa	Very low diversity and heterozygosity in artabicas, low to moderate diversity and heterozygosity in robustas, moderate to high transferability	–
Rajkumar et al. (2000)	RAPD (41 primers), ISSR (6 primers)	32 putative different accessions of*C. euginoides, C. salvatrix, C. stenophylla, C. dewevreii*	All tested primers polymorphic between species, but resulted in almost monomorphic patterns between accessions of a species.	High inter-species but negligible intra-species (between accessions) diversity.
Anthony et al. (2001)	RAPD (150 decamers)	Wild and semi wild cultivars of arabica *Total = 118*	16 polymorphic primers with 29 polymorphic amplicons out of a total of 106.	6 clusters based upon origin and pedigree.
Steiger et al. (2002)	AFLP (6 primer pairs)	Arabicas: Bourbon (4 acc.), Catimor (8 acc.), Red Catui (6 acc.), Yellow Catui (3 acc.), Catui (1 acc.), Caturra (2 acc.), Mokka hybrid (MH, 8 acc.), Typica (26 acc.), Diploid: robusta (1 acc.), *C. liberica* (2 acc.) *Total = 58 arabicas, 3 diploids*	Very high genetic similarity within the arabica groups. Typica least diverse (97% sharing among accs.), mean genetic similarity 0.935 (range 0.767 to 1.0).	I = 22 Typica acc., 2 Bourbon, 1 Catui II = 5 MH, 1 Typica III = 5 Catui, 2 Caturra IV = 2 Typica, 1 MH V = 4 Catimor VI = 2 libericas 12 accessions could not be grouped.
Anthony et al. (2002)	AFLP (37 primer pairs), SSR (6 primer pairs)	Confirmatory samples (1 each of Typica and Bourbon), Commercial cultivars (6 acc.), arabica mutants (2 acc.), arabica di-haploid (1 acc.), cultivars from collections (4 acc.), wild arabicas (11 acc.) *Total = 26*	High diversity among wild arabicas (90% of all) as compared to cultivated. High similarity between Typica and Bourbon.	Except for Typica and Bourbon as clusters, other accessions were grouped very weakly.

(Continued)

Table 4. (Continued)

Study source/ Reference	Techniques used	Germplasm studied	Important results	Phylogeny grouping and relationship
Prakash et al. (2002)	AFLP (36 primer pairs) gene introgression into S-26 derived tetraploid arabicas	17 introgressed genotypes (S- 288 x Kent arabica), *C. arabica* (5 acc.), S-288, Kent arabica *C. liberica* (5 acc.) *Total = 29*	Very low polymirphism in arabicas (only 35 bands), less alien genome introgression into the introgressed arabicas .	No evident relationship between introgressed cultivars.
Herrera et al. (2002a)	AFLP (30 primer pairs), SSR (19 Primer pairs)	From 6 triploid hybrids (*C. arabica* x *C. canephora*), 6 BC progenies were produced @ 10 plants per progeny *Total = 60*	Low introgression from canephora genome into arabica genotypes (13.8% of all the genotypes had canephora specific alleles.	–
Herrera et al. (2002b)	RFLP (11 probes), SSR (13 primer pairs)	1 tetraploid hybrid plant (arabica x tetraploid canephora i.e. IF181T) crossed with *C. arabica* (Et 30) to produce F1 hybrid. 28 P1 BC plants and 45 P2 BC plants	Very high level of geneflow from canephora to arabicas.	–
Bardini et al. (2003)	Tubulin based polymorphism (TBP: β-tubulin intron 1 specific primer pair)	Arabica (5 acc.), robusta (9 acc.)	No polymorphism in arabica and some polymorphism in robustas.	–
Silviera et al. (2003)	RAPD (18 primers/ restriction enzyme combinations)	90 individuals belonging to 6 Sarchimor progenies, 2 cultivars and a progeny (all arabicas)	38.5% and 68.5% variations within and between populations. Low to moderate diversity detected.	Clustering according to progenitors.

Baruah et al. (2003)	SSR (9 primer pairs)	Arabicas (45 acc.), robustas (6 acc.), 13 *Coffea* spp., 4 *Psilanthus* spp.	Diversity and PIC values High in robustas, and low in arabicas, robust cross species transferability.	Fixed heterozygosity (ca. 100%) in arabicas proposed due to presence of duplicate loci.
Bhat et al. (2004)	EST-SSR (9 primer pairs)	Arabicas (15 acc.), robustas (8 acc.), 14 *Coffea* spp., 4 *Psilanthus* spp.	Low diversity/ very low PIC for arabicas, Low-high diversity/ PIC values for robustas, robust cross species transferability with moderate diversity.	Fixed heterozygosity (ca. 100%) in arabicas proposed due to presence of duplicate loci.
Chaparro et al. (2004)	RAPD (42 decamers)	47 explored arabica accessions	24 (57%) primers produced moderate polymorphism. 34.2% bands are polymorphic. High diversity in wild arabicas.	One loose cluster containing 22 accessions and other dispersed. No defined basis for clustering.
Poncet et al. (2004)	SSR (110 primer pairs)	*C. canephora* (11 acc.), *C. eugenioides* (4 acc.), *C. liberica* (4 acc.), *C. heterocalyx* (1 acc.), *C. sp.* Moloundou (2 acc.), *C. pseudozanguebariae* (10 acc.) *Total = 32*	High transferability among species (81.4%), high genetic diversity and PIC in canephora and pseudozanguebariae,	–
Moncada and McCouch (2004)	SSR (34 primer pairs)	Arabica cultivars (10 acc.), HdeT (2 acc.), wild arabicas (11 acc.), robusta (1 acc.), *C. eugenioides* (2 acc.), *C. liberica* (2 acc.), *C. congensis* (1 acc.), *C. kapakata* (1 acc.) *Total = 30*	More alleles/high PIC value in diploids, followed by wild arabicas, cultivated arabicas. But SSRs unearthed more diversity in cultivated arabicas than by RFLP or RAPDs.	I = cultivated arabicas, few wild arabicas, HdeT II = wild arabicas Others: *C. eugenioides* closer to arabicas, *C. kapakata* different from all; *C. liberica*, *C. canephora* and*C. congensis* closer to each other.

(Continued)

Table 4. (Continued)

Study source/ Reference	Techniques used	Germplasm studied	Important results	Phylogeny grouping and relationship
Aggarwal et al. (2004b)	RAPD (37 primers), AFLP (7 6-bp selective primer pair sets), SSR (150 primer pairs)	16 superior selections of coffee (14 arabica and 2 robusta) developed in India	Considerably more alleles/ high PIC value in diploid robustas compared to arabicas. SSRs unearthed more diversity in cultivated arabicas than by AFLP or RAPDs.	Clear separation between robusta and arabica selections. Reference panels of discriminating, polymorphic markers could be generated for individualization of each selection.
Prakash et al. (2005)	AFLP (15 primer pairs), SSR (12 primer pairs)	Indian collection of robusta genotypes (60 acce.), core collection robusta (14 acc.), *C. congensis* (3 acc.)	72 alleles with 12 SSRs, 205 polymorphic bands with AFLP, i.e. high level of polymorphism.	Clusters according to five diversity groups of robusta coffee.
Aga et al. (2005)	ISSR (15 primers)	160 genotypes belonging to 16 wild populations of arabica coffee	Moderate level of polymorphism (25% polymorphic bands).	Clustering of genotypes based on geographical location.
Masumbuko and Bryngelsson (2006)	ISSR (6 primers)	Arabica (100 acc. from 5 regions), *C. eugenioides* (2 acc.), *C. zanguiberae* (5 acc.), *C. mufindiensis* (3 acc.)	Low variability for arabicas.	Grouping based on species and geographical location.
Aggarwal et al. (2007)	EST-SSR (18 primer pairs)	Arabica (15 acc.), robusta (8 acc.), 13 *Coffea* spp., 4 *Psilanthus* spp.	16 polymorphic pairs, comparable PIC (0.42) and diversity in both arabica and robusta. High cross species transferability/sequence conservation across the species.	Clustering based on pedigree and geographical origin.

Table 5. DNA markers based studies for deciphering species relationship among *Coffea* taxa

Study source/ Reference	Technique	Species assessed	Important results	Groupings and relationship
Berthou et al. (1983)	Chloroplast (cp) and mitochondrial (mt) DNA RFLP polymorphism using *Hpa* II enzyme and *Sal* I enzyme respectively	Arabica (2 acc.), robusta (2 acc.), *C. eugenioides* (2 acc.), *C. congensis* (2 acc.), 'nana' taxon, *C. excelsa*, *C. liberica*, *Paracoffea ebracteolata*, *C. arabusta*	No difference in cp-DNA of (arabica and eugenioides) and (canephora and congensis). Similarity between mt-DNA of (arabica, eugenioides, congensis) and (canephora and 'nana' taxon), unique patterns for other species	*cp – DNA* Group A: *C. arabica*, *C. eugenioides*, *C. congensis* Group A': *C. canephora*, 'nana' taxon *mt – DNA* Group 1: *C. arabica*, *C. eugenioides* Group 2:'nana' *C. arabica* has an ancester similar to *C. eugenioides*.
Lashermes et al. (1996c)	Chloroplast DNA RFLP in *atpB-rbcL* intergenic region	52 trees from 25 *Coffea* taxa *Total = 52*	Low divergence, 12 different plastomes were detected. Maternal inheritance of chloroplast confirmed.	– *Coffea* is monophyletic and recent in origin
Orozco-Castillo et al. (1996)	Chloroplast PCR-RFLP (*trn*T-L, *trn*L, *trn*L-F)	Arabica (4 acc.), robusta (3 acc.), *C. eugenioides* (2 acc.), *C. liberica* (2 acc.), C. stenophylla, *C. racemosa*, *C. humilis*, *C. pseudozanguebariae*, *C. congensis*, *C. sessiflora*, *C. breviceps* (1 acc. each); *Total = 18*	Amplicons from *trn*L, *trn*L-F shows enzyme site polymorphism.	I = *C. arabica*, *C. eugenioides*, *C. humilis*, *C. stenophylla* II = *C. canephora*, *C. liberica*, *C. breviceps*, *C. congensis* III = *C. pseudozanguebariae*, *C. sessiflora*, *C. racemosa*
	Mitochondria PCR-RFLP (V7 rDNA)		Monomorphic in all the species.	–
Lashermes et al. (1997)	ITS 1 and 2 region of nuclear ribosomal DNA	26 *Coffea* taxa, 3 *Psilanthus* taxa *Total = 37*	ITS 2 region informative. Intra-species sequence variants were observed.	Strong geographical correspondence among the clusters.

(Continued)

Table 5. (Continued)

Study source/ Reference	Technique	Species assessed	Important results	Groupings and relationship
Cros et al. (1998)	*Trn*L-*trn*F intergenic spacer of chloroplast DNA	23 *Coffea* taxa and 2 *Psilanthus* taxa	Low sequence divergence (only 20 types obtained). No difference between *C. canephora*, *C. congensis* and *C. breviceps* and among *C. eugenioides*, *C. arabica* and *C. sp.* Moloundou.	Strong geographical correspondence among the clusters. Supports radial mode and recent origin of *Coffea* taxa in Africa.
Saini et al. (2000)	RAPD (20 primers), ISSR (10 primers)	15 *Coffea* spp., 4 *Psilanthus* spp. Individual as well as pooled samples	High diversity across different species	Well resolved species clusters showing correspondence to their botanical types, as well as geographical origin; *C. kapakata* was indicated to be a *Coffea* sp. rather than belonging to *Psilanthus* or *Psilanthopsis* genra. Four endemic paracoffea species appeared as a distinct cluster under related genus *Psilanthus.*

to genetic uniformity (Lashermes et al., 1996d). In addition, many of the present day cultivars are suggested to have originated from single mutations having greater effect (Krug and Carvalho, 1951). All these factors suggest a narrow genetic base of arabica, which is largely vindicated by DNA markers based assessment of coffee germplasm. The first ever study to this end (Lashermes et al., 1993) using PCR based RAPD markers revealed that while there were sufficient inter-specific variations between arabica and canephora, there was almost no detectable variation within the few arabica accessions analysed in the study. Subsequently, a large number of studies (Table 4) carried out using molecular markers viz., isozymes, RFLP, RAPD, AFLP to detect variability in arabica and robusta coffee germplasm revealed:

a) Invariably a very low level of variability in cultivated arabica, which could be detected only when a very large number of markers were used for analysis;
b) Suitability of molecular approaches like RAPD and AFLP (that provided relatively high multiplex ratios) to resolve the subtle variation in arabica germplasm compared to approaches like isozymes and RFLP that generally failed to detect the same (Berthou and Trouslot, 1977; Lashermes et al., 1996a; Lashermes et al., 1999);
c) Considerably high genetic variability in the *C. canephora* and *C. congensis* germplasm compared to arabicas (Lashermes et al., 1993, Prakash et al., 2005), which could be detected rather easily even by RFLP analysis (Lashermes et al., 1999);
d) Relatively higher genetic diversity in the wild collections of arabica than cultivated ones, but much less than that seen for the diploid robusta germplasm (Lashermes et al., 1996d; Orozco-Castillo et al., 1996; Lashermes et al., 2000a; Anthony et al., 2001, 2002; Chaparro et al., 2004; Moncada and McCouch, 2004; Aga et al., 2005);
e) Arabica cultivar like HdeT or its derivatives exhibit more diversity than cultivated as well as wild arabica type (Lashermes et al., 2000a); it is possibly due to the genetic introgression that probably happened during natural hybridisation event underlying HdeT evolution;
f) Multiplex markers like AFLPs could distinguish and group different accessions belonging to six arabica cultivars (Bourbon, Catimor, Catuai, Caturra, Mokka hybrid, Typica) based on their origin and pedigree but without any cultivar specific signatures (Steiger et al., 2002);

In recent years, with the development and availability of coffee specific SSR markers (Table 3), the same are now also being tried for coffee germplasm characterization. The few studies carried out using a limited number of SSR markers while in line with the earlier observations of low variability in cultivated arabica compared to wild arabica accessions and diploid coffee germplasm, but expectedly and desirably demonstrate the SSRs utility over other marker approaches in their ability to more efficiently detect the inherent low variability of arabicas (Aggarwal et al., 2004a, 2004b, 2007; Baruah et al., 2003; Bhat et al., 2004; Moncada and McCouch, 2004).

The efforts to date, thus, demonstrate that PCR based DNA markers can be used reliably for identification and ascertaining the genetic diversity in the available cultivated and wild exotic germplasm of coffee, and that the microsatellites are the most efficient and desirable markers for the purpose. Nevertheless, the experience from these studies also highlights the overall low efficiency of the molecular marker approaches for coffee genepool characterization. The data clearly establish that despite their high-genetic resolution, DNA typing need to be carried out at large scale to resolve the low variation inherent in the coffee genepool especially of arabica, thus making the whole exercise resource intensive and practically non-viable. The later constrain warrants development of ways and means to increase the efficiency of DNA markers based expensive but unavoidable approaches for coffee genetic analysis. An easy way to achieve this can be creation of molecular data banks enumerating DNA marker (approach) efficiencies and more importantly, DNA polymorphism status (using defined markers/guidelines) for the analyzed germplasm, for use as reference resource (Aggarwal et al., 2004b). Based on our experience, we propose construction of 'Reference DNA polymorphism databases' of elite coffee germplasm available world-wide using standard repeatable markers (such as microsatellites) that can then be used by the coffee geneticists/breeders community for various advantages namely, for: a) better management, utilization, registration and IPR protection of elite coffee germplasm; b) selection of suitable material/genotypes and informative DNA markers for breeding and linkage analysis; c) exchange of germplasm, etc.

4.2. Genetic Diversity in Secondary Genepool of Coffee

Secondary genepool of coffee comprises >100 related diploid species of *Coffea* and *Psilanthus*, which represents an immensely valuable storehouse for the desirable genetic variability/gene(s) of interest needed for genetic improvement of the cultivated coffee. This potential assumes even more significance, especially considering that: different coffee species interbreed freely and relatively easily with high levels of gene flow (Berthaud and Charrier, 1988; Louarn, 1993), and there exists considerable genetic diversity in this pool (Berthaud and Charrier, 1988; Lashermes et al., 1993; Orozco-Castillo et al., 1996; Saini et al., 2004, our unpublished data). But to exploit this enormous secondary genepool for genetic improvement of coffee, there is a need for comprehensive studies on genetic characterization that can reveal its genetic base as well as broad generic affinities among/across populations and species. This task can best be accomplished using DNA marker approaches that provide high-genetic resolution and thus have become central to the germplasm diversity studies. Accordingly, even in case of coffee the DNA markers have now been used increasingly to characterize the related wild genepool.

4.3. Phylogenetic Studies for Species Relationship using DNA Markers

More than 100 species have been described under the genus *Coffea* mainly based on differences in morphology, size and ecological adaptations. Interestingly, most of these species can interbreed freely suggesting them to be of relatively recent origin, and also the possibility of many of these purported species may actually be natural hybrids rather than true type species. The situation warrants systematic, comprehensive studies using high-resolution genetic approaches as possible by DNA marker analysis for deciphering interspecific relationships and redefine some of the coffee species/taxonomy. Some such studies carried out in recent years are summarized in Table 5 (few also in Table 4). As becomes evident from these studies, a number of DNA marker approaches (viz., evolutionary informative DNA signatures of organelle genomes (cp-DNA, mt-DNA) and/or ITS (*I*nternally *T*ranscribed *S*pacer) region of the nuclear 5S rDNA, and nuclear length polymorphism based markers like RAPDs, AFLPs and ISSRs, have been tried (generally on a limited number of coffee species/accessions) with varying success. Briefly, the results from these studies show:

1. Low polymorphism in organelle DNA providing poor resolution of species relationships, but providing support for monophyly of *Coffea* genus, and also suggesting *C. eugioniodes* as possible maternal progenitor of *C. arabica* (Berthou et al., 1983; Lashermes et al., 1996c, Oroczo-Castillo et al., 1996, Cros et al., 1998).
2. ITS region is phylogenetically more informative than organelle DNA signatures for deciphering species level divergence among *Coffea* species (Lashermes et al., 1997, Aggarwal, 2005), but needs to be evaluated more thoroughly for its overall potential. Lashermes et al. (1997) using only its sub-domain ITS2 specific polymorphism could resolve relationships of 26 coffee species, but also encountered confounding intra-species specific variation (almost similar to seen in the inter-species comparisons). Contrary to this, our detailed studies suggest that to obtain confident assessment of coffee species relationships one need to consider the complete ITS region polymorphism rather than a part of variation across its sub-domains ITS1 or ITS2 (Aggarwal, 2005).
3. Compared to the domain specific sequence signatures, length polymorphism based DNA markers like RAPD, AFLP, SSR and ISSR (which generally mirror the nuclear genomic variation), provide more resolution and relatively consistent results about the species relationship between different *Coffea* species. Most of these studies broadly support the conventional coffee taxonomic relationships and/or strong indicate geographical correspondance (Lashermes et al., 1993, 2000b; Oroczco-Castillo et al., 1996; Saini et al., 2000; Steiger et al., 2002; Moncada and McCouch, 2004; Aggarwal, 2005; Aggarwal et al., 2007; see Tables 4 and 5).
4. In addition, our empirical findings involving comparisons of taxonomic interpretations of 19 species of *Coffea* and *Psilanthus*, based on DNA polymorphisms (across evolutionarily informative sequences viz., intergenic *trn*L regions of cp-DNA, 16S rDNA of mt-DNA, and nuclear ITS region, as well as, mobility

based nuclear markers- RAPD, ISSR and genomic-/EST-SSRs), representing each of the 3 genomic compartments of the cell (chloroplast, mitochondrion, nucleus) strongly suggest that organelle genomes may not be ideal indicators of evolutionary/taxonomic relationships of coffee species (Aggarwal, 2005). We expect this apparent failure of the organelle DNA to be due to the relatively easy inter-species crossing which is known to happen in natural conditions rather frequently across diploid coffee species.

The above studies amply demonstrate the potential of DNA polymorphism based analysis for understanding coffee systematics and phylogeography, but also brings in focus the need for devising standard guidelines and choice of suitable DNA markers/signatures for robust and reliable inferences. Some of the major consideration that need to be addressed are: a) whether organelle DNA/genomes provide suitable targets for coffee taxonomic studies; b) what and how many samples should be used to represent a given coffee species; c) among different nuclear markers (RAPD, AFLP, ISSR, SSR), which are the most suited; and d) finally how many markers should one use to obtain high level of confidence in resulting interpretations.

4.4. DNA Markers to Assess Gene Flow and Introgression into Cultivated Coffee

Inter-specific hybridisation and production of natural fertile hybrids are well documented in coffee (Louarn, 1993; Charrier, 1978) between and across ploidy barriers. Many of such natural hybrids, especially of arabica types, contain desirable traits that are expected to have resulted as a result of gene flow from the diploid coffee species surpassing ploidy barriers. Some such hybrids like HdeT and S-26 have been widely used in rust resistance breeding with successful development of few improved introgressed arabica cultivars showing some rust resistance, like Catimor, Sarchimor, S-26 (Bettencourt, 1973; Sreenivasan, 1986; Sreenivasan et al., 1993; Charrier and Eskes, 1998). However, identification of the original coffee sources of desirable rust resistance gene(s) and their introgression in *arabica* coffee has remained a matter of speculation and interest for long. Only recently using DNA markers based analysis, it has become possible to answer some of these important questions. AFLP analysis of HdeT and introgressed arabica cultivars (Catimor, Sarchimor) has suggested *C. canephora* as the putative source (Lashermes et al., 2000b). Similarly, AFLP typing of S-26 (putative natural spontaneous hybrid identified in India), its derived resistant Indian arabica variety S 795 and related breeding materials have indicated *C. liberica* as the source for putative resistance factor SH-3 (Prakash et al., 2002). Both these studies provide support to the theory that gene flow/introgressions from diploid coffee species to tetraploid *C arabic*a might have happened quite often leading to successful gene flow to the cultivated coffee genepool, which in turn also suggest the practical feasibility and scope for attempting systematic interspecific hybridisation program for gene introgression and coffee improvement.

4.5. Origin and Evolution of *Coffea arabica* as Deciphered from DNA Markers

The evolution of arabica coffee has always been an intriguing puzzle for the researchers because it is the only autogamous and tetraploid species in the *Coffea* genus. The classical cytogenetic studies suggest that it might have originated from coupling of unreduced gametes from a cross involving *C. eugeniodes* and any one of the *C. canephora*, *C. congensis*, and *C. liberica* (Cramer, 1957; Carvalho, 1952; Narasimhaswamy, 1962), followed by genetic regulation of synapsis helping in the progressive diploidisation and evolution of present day amphidiploid *C. arabica* from the archetype tetraploid intermediate (Charrier and Berthaud, 1985). However, meiotic pairing in interspecific hybrids generated from *C. eugenoides* and *C. canephora* was found better (Louarn, 1976) than in dihaploid plants of *C. arabica* (Mendes and Bachhi, 1940; Berthaud, 1976), suggesting possibility of chromosomal re-patterning after the polyploidization event. The recent DNA based studies using conserved regions from nuclear and/or chloroplast genome have not only substantiated some of the above conclusions but also provided better understanding of the evolution and origin of arabica coffee. The trnL-trnF intergenic spacer of cp-DNA of *C. arabica* was found similar to that of *C. eugenioides* and *C*. sp. Moloundou suggesting that the maternal species of *C. arabica* could be any one of these or a similar taxon (Cros et al., 1998). On the other hand, analysis of the ITS2 spacer sequences of nuclear 5S rDNA domain (Lashermes et al., 1996a; Lashermes et al., 1997) suggested one of the canephoroid species (*C. canephora*, *C. congensis* and *C. brevipeps*) as the other progenitor species, and possibility of genomic repatterning (rDNA homogenisation via elimination of sequence specific to the other putative maternal progenitor). Subsequently, using RFLP markers and GISH (genomic *in-situ* hybridisation) analysis Lashermes et al. (1999) meticulously demonstrated the presence of 22 chromosomes belonging each to canephoroid (C) and eugenioides (E) type genomes, and amphidiploid nature of *C. arabica*. Moreover, the observed RFLP allelic diversity across different genera (considering a divergence timeline of 5-25 MYA for the most common ancestor of coffee) suggested a relatively recent speciation of *C. arabica* over 1 MYA (Lashermes et al., 1999).

Despite the above progress in our understanding of the arabica evolution, there are many questions/speculations that need to be answered and/or validated. Some of these are: a) the extent and pattern of genomic reorganization, and whether the arabica genome is like a segmental allotetraploid; b) the underlying mechanism responsible for diploid bahaviour- pairing specific gene(s) and/or structural differentiation of chromosomes; and c) the possibility of duplicate loci (one specific to each constituent genome) as indicated in some recent SSR based diversity analysis of arabica germplasm (Baruah et al., 2003; Bhat et al., 2005; Hendre, 2006). It is hoped that availability/use of the advance high genetic-resolution capable DNA based comparative genomic resources/tools (linkage maps, sequence databases, microarray chips etc.) would help provide answers to these intricate questions in near future.

4.6. Molecular Linkage Maps of Coffee

The utility of DNA marker based genetic maps in linkage studies and genetic improvement of crop species in now well established. The potential of such maps is expected to be even more in case of difficult plant species like, trees including coffee, where conventional breeding efforts are severely constrained due to lack of genetic markers, screening tools and long generation cycles (outlined in earlier section). It is expected that availability of dense DNA marker based linkage maps would greatly help in identifying marker tags/QTLs for gene(s) controlling different qualitative and quantitative traits of agronomical importance and of interest to coffee breeders, which in turn would promise MAS based accelerated breeding of improved coffee genotypes.

In recent years, a modest beginning has been made to generate molecular linkage maps of coffee with limited but promising success. Till date, five DNA marker based linkage maps have been developed using RAPD, AFLP, RFLP and SSR markers in different genetic backgrounds (Table 6). Four of these maps are of diploid coffee genome, represented by two *C. canephora* and other two interspecific maps. The first DNA marker based linkage map was constructed using RAPD and RFLP markers and double-haploid mapping population of a *C. canephora* clone IF-200 (Paillard et al., 1996). Subsequently, the same mapping population was used to develop another map wherein AFLP and SSR markers were also used (Lashermes, 2001). In general the studies revealed relatively low levels of mappable polymorphism, which seemingly led to the attempts involving the use of interspecific backcross (BC1) populations that were expected to be more informative, and/or use of more SSR markers. The first such interspecific map was developed by Ky et al. (2000) based on AFLP and RFLP marker analysis of the BC1 progeny {(*C. pseudozanguebariae* X *C. liberica* var. *dewevrei*) X *C. liberica* var. *dewevrei*}, and more recently a similar map has been developed using another interspecific BC1 population {(*Coffea heterocalyx* X *C. canephora*) X *C. canephora*} and AFLP and SSR markers (Coulibaly et al., 2003b). These maps had more mapped markers with relatively better genome coverage than of *C. canephora*, and thus suggested the utility of the approach (inter-specific population). The studies also reinforce the potential and significance of the secondary genepool for basic, as well as, applied research on coffee genetics and improvement.

We at CCMB, have now developed a relatively well populated framework linkage map for cultivated diploid robusta, under the 1st 'Indian Initiative on Coffee Genomics', using a more practical alternative of using a pesudo-testcross population (Hendre, 2006). This first generation map (Figure 2) developed using a trait-specific (drought tolerance) mapping population has a total of 374 mapped markers (comprising 185 RAPDs, 118 AFLPs, 71 SSRs, and a moderate marker density of 3.3 cM centimorgans) spread over 11 major and 5 minor linkage groups. This is the most elaborate map of diploid coffee genome till to date, and also has a very large number of mapped SSRs. Thus, it is expected to be of great value in future studies on identification of QTLs for drought tolerance (one most important

Table 6. DNA marker based linkage maps available till date for coffee genome

Study Reference	Paillard et al. (1996)	Ky et al. (2000)	Lashermes et al. (2001)	Coulibaly et al. (2003b)	Pearl et al. (2004)
Species used for mapping	*C. canephora* P. ex Froehner	Interspecific	*C. canephora* P. ex Froehner	Interspecific	*C. arabica*
Parents	IF-200	{(*C. pseudozanguebariae* X *C. liberica* var. *dewevrei*) × *C. liberica* var. *dewevrei*}	IF-200	{(*Coffea heterocalyx* × *C. canephora*) × *C. canephora*}	Mokka Hybrid x Catimor Hybrid
Population	DH (n = 85)	BC (n = 62)	DH (n = 92), BC (n = 44)	BC (n = 74)	F1 (n = 60)
DNA Markers used	RFLP, RAPD	RFLP, AFLP	RFLP, RAPD, AFLP, SSR	AFLP, SSR	AFLP
No. of loci/ primers/ probes used to screen population	RAPD = 250 primers RFLP = 250 genomic probes and 75 cDNA probes	AFLP = 30 primer pairs RFLP = same as Paillard et al. (1996)	RFLP = same as Paillard et al. (1996) RAPD = over 90 primers AFLP = 6 primer pairs SSR = 18 primers	AFLP = 12 primer pairs SSR = 113 loci (Rovellie et al., 2000)+ 47 loci (Dufour et al., 2001)	AFLP = 288 primer pairs
No. of mapped loci	RAPD = 100 bands RFLP = 47 loci*Total = 147*	AFLP = 167 bands RFLP = 13 loci Phenotypic marker = 1 *Total = 181*	DH: RFLP = 36 loci RAPD = 11 bands AFLP = 97 bands SSR = 18 loci Total = 162 TC: RFLP = 32 loci RAPD = 6 bands AFLP = 53 bands SSR = 18 loci Total = 109 bands *Total (DH) = 160*	AFLP = 160 out of 188 polymorphic bands SSR = 29 out of 36 polymorphic loci Phenotypic marker = 1 *Total = 190*	AFLP = 464 bands *Total = 464*

(Continued)

Table 6. (Continued)

Study Reference	Paillard et al. (1996)	Ky et al. (2000)	Lashermes et al. (2001)	Coulibaly et al. (2003b)	Pearl et al. (2004)
Length of map (average marker density)	1402 cM (9.5 cM)	1,114 cM (6.15 cM)	1041 cM (6.5 cM)	1,360 cM (7.2)	1802.8 cM (3.9 cM)
%SD (bands and loci of polymprohic bands or loci)	RAPD = 20% RFLP = 12%	30%	DH = 44% TC = 14%	AFLP = 15.9% SSRs = 16.6%	25%
No. of linkage groups	15	14	11	15	31

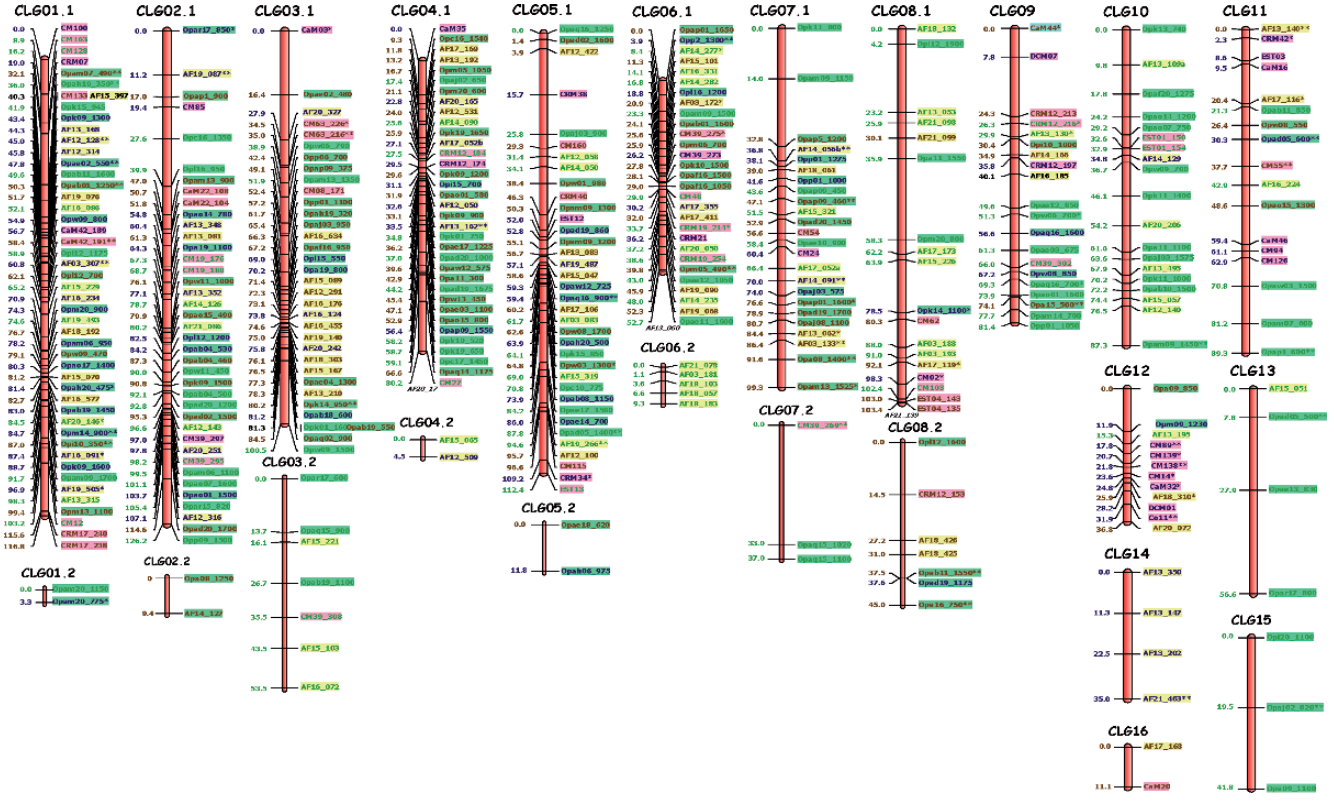


Figure 2. An integrated framework linkage map of robusta coffee (*C. canephora*) developed using a *pseudo-testcross* strategy and DNA markers (AFLP, RAPD, SSRs). The map comprises 11 major and 5 minor linkage groups, and has a total of 374 mapped markers. Prefixes Op-, AF- in the marker names indicates RAPD, AFLP respectively, while CM-/CR-/CaM- represents the mapped SSRs. The small groups shown below some of the major CLGs, represent the floating markers of the respective CLG placed in an 'Accessory group/map" (unpublished; Hendre, 2006)

trait of interest to breeders in India and elsewhere), and in realization of MAS based molecular breeding for accelerated robusta coffee improvement programs.

In contrast to diploid coffee, there had been very little progress in case of mapping of the tetraploid genome of arabica coffee. Till date only one AFLP based linkage map has been developed for arabica genome (Pearl et al., 2004) using a pseudo-F2 population and pseudo-test cross approach (Grattapaglia and Sederoff, 1994). The length (ca. 1800 cM) and marker density (3.9 cM) of this map are significantly higher than observed for any of the diploid maps (that ranges from ca. 1000 cM to 1400 cM, with low to moderate marker density of 6.15–9.5 cM). But the map has a disadvantage that it is based on only AFLP markers, which although are good for map saturation, fine-gene mapping, but remain difficult to be uniquely characterized/identified and thus do not make ideal, stable, easily revisitable genomic landmarks like SSRs for use in subsequent studies/practical breeding schemes.

To our understanding, the most important hurdle in mapping arabica genome is its highly homogenous state with very little detectable variation at DNA level in the cultivated genepool (expected to be due to historical reasons mentioned in the beginning), which is further aggravated with it being a tetraploid. Thus, creating a useful reasonably saturated molecular/genetic reference linkage map of arabica coffee (which is of major interest to all coffee producing countries, including India) is going to be a mammoth task, which is not realistically feasible for any individual group/institute or probably even for any one country. Any worthwhile attempts to this end thus, warrant active interaction between global partners through international initiative like many other international programmes that have been started in recent years for generating mapping related resources for targeted crop plants, viz., International Tritiaceae Mapping Programme (ITMI), Genomic Network for Solanaceous crops (Sol; http://www.sgn.cornell.edu/solanaceae-project/index.pl), Loblolly Pines Genome Project (http://dendrome.ucdavis.edu/lpgp/), International Cotton Genome Initiative (http://icgi.tamu.edu/), International Grape Genome Project (http://www.vitaceae.org/), PROMUSA for banana improvement etc. The Department of Biotechnology, India along with the Coffee Board, India, initiated efforts to discuss the needs, feasibility, and possible mechanisms of extending the national initiative to a wider International platform for accelerating the coffee improvement efforts. These motivated efforts led to the ‘Bangalore Declaration’ in January 2004, followed by conceptualization and origin of a formal international group i.e. ICGI (International Coffee Genomic Initiative) in October 2004 in Bangalore during the ASIC2004 meet. ICGI had its first meeting in Paris, France in April 2005, which led to formalization of the International Coffee Genomic Network (ICGN; http://www.coffeegenome.org/about/launching_meeting.htm), wherein creation of saturated coffee maps with international participation have been identified as one of the major research priorities in immediate future.

4.7. Mapping of Quantitative and Qualitative Traits in Coffee

It is now well established that most of the plant traits of agronomical interest are quantitative, controlled by multiple genes that need to be judiciously manipulated to develop genetically improved germplasm, but hitherto were not accessible through conventional breeding approaches. With the development and availability of DNA marker based linkage maps for many animal and plant species, identification, mapping and selection of Quantitative Trait Loci (QTLs), have become practically feasible and achievable. The QTLs can be identified by monitoring mapped or unmapped DNA markers in conjunction with target trait using a segregating mapping population (Hackett, 2002), and/or diverse genotypes employing dis-equilibrium mapping (Gupta et al., 2005). Though having a molecular linkage map is not a prerequisite for linking a QTL, it is essential to map them for efficient selection and subsequent transfer of the trait. In case of coffee, QTL mapping is in its infancy with few successful efforts in recent years. In the first such study in coffee, few RAPD markers were linked with resistance to coffee berry disease (CBD) caused by *Colletotrichum kahawae* (Agwanda et al., 1997), using the indirect approach of diverse genotypes (5 susceptible and 8 resistant arabica cultivars/selections) and unmapped RAPD markers. The study revealed 24 candidate linked markers of which only three showed strong association with putative CBD resistance *T* gene (from Catimor) and proposed for use in marker assisted selection (MAS). On the other hand, Coulibaly et al. (2003a) could identify 3 fertility restorer QTLs using the partial AFLP based interspecific map. Similarly, QTL responsible for trigonelline content (an important compound responsible for the taste and flavour of coffee) was identified on the other interspecific linkage map developed by Ky et al. (2000). In addition, it is suggested that putative QTLs for caffeine, chlorogenic acid and sucrose content in diploid coffee (Ky et al., 2000), and for source-sink traits in arabica (Pearl et al., 2004), have been identified. In contrast to the above few reports on mapping of quantitative traits, there is one study wherein RFLP marker could be linked to self-incompatibility (validated to be a qualitative trait) on the robusta DH map (Lashermes et al., 1996b) and further validated on the intraspecific AFLP linkage map (Coulibaly et al., 2002).

Although few, the above reports are encouraging and give the hope that with the all increasing emphasis, excitement and international efforts on coffee genomics, elaborate coffee maps and markers would become available soon paving the way for identification of molecular tags/QTLs (for most of the important qualitative/quantitative traits of interest) and MAS based coffee breeding.

5. CONCLUSIONS

Coffee is a an important crop for the world socio-economy, which provides livelihood to the millions of people worldwide with sustenance of few third world countries solely dependent on coffee cultivation. Similar to other crop

plants, there is a continuous need for genetically improved coffee germplasm to meet the ever-changing demands of environment and markets, but it is also a difficult crop to conventional genetic improvement efforts. The problem is more acute in case of arabica coffee that for historical reasons suffers from a very narrow genetic base making genetic improvement more difficult to be realized. The constraints and barriers of conventional coffee improvement warrant newer powerful molecular approaches that can provide the requisite speed, efficiency, and directionality to breeding efforts. In this context, the DNA markers/technologies that provide high-genetic resolution becomes important and thus need to be integrated in research on genetics and improvement of coffee, which till recently was treated more like an orphaned plant species in context of molecular genomic work. It is fortunate that in recent years the realization has dawned on the coffee community worldwide, and modest but sincere efforts have been initiated that amply demonstrate the promise of the DNA markers/approaches in genetic studies on coffee. DNA analysis of the existing coffee germplasm resources indicate very low genetic diversity in cultivated genepool and highlight the need for exploitation of wild arabicas and the secondary gene pool (consisting of diploid species) for desired variability/gene(s) for coffee improvement. The potential of wild germplasm is exemplified recently by the discovery of 3 arabica plants devoid of caffeine in conventional breeding program that included explored arabica germplasm from Ethiopia (Silvarolla et al., 2004). Moreover, interspecies geneflow is widespread in coffee. Considered together, these suggest that exploration of extant germplasm/secondary genepool followed by conventional breeding approaches combined with advanced genotyping methods based on DNA markers, is a practically realistic proposition to boost the genetic improvement of coffee.

6. FUTURE PERSPECTIVES

Key for modern genetic improvement programmes includes application of DNA marker technology for characterization of desirable variability, construction of saturated molecular linkage maps with wider genome coverage and using robust, easily transferable markers like SSRs or SNPs. Availability of such resources would be the beginning for QTL mapping and application of Marker Assisted Selection in coffee. But to make this potential of DNA tools a reality, and also considering the major constraints and possibilities at global level the following activities need to be taken up on priority:

1. Development of good informative markers for genetic studies: Very few microsatellite markers, which are the most useful for genetic linkage studies, are available for coffee, and only a fraction of these are informative for elite arabica coffee germplasm (as evidenced in our work in India and similar studies worldwide). The situation thus necessitates the development of a very large number of microsatellite markers (many fold more than what have been developed till date) to have enough informative, usable markers for

rational deployment in genetic-linkage analysis/improvement programs of *Coffea arabica*.

2. Genetic diversity in coffee genepool and need for reference polymorphism panels: Enlarging the desirable genepool by concerted international efforts involving various research groups working on arabica coffee from different countries by: identification of sets of elite germplasm (being used or having potential for breeding programs) available with different groups worldwide to create a global resource of arabica expectedly defining the broadest genetic base; exchange of the identified germplasm resources (possibly in the form of DNA samples initially and later the shortlisted germplasm as per the need); DNA fingerprinting of germplasm using pre-defined standard protocols and robust DNA/SSR markers, by multiple labs to generate 'reference polymorphism panels', which can be shared to serve as a ready reference tool to identify most diverged, potentially useful germplasm for deployment in arabica breeding programs.
3. Molecular linkage mapping in coffee: This most important but also the most onerous activity, especially in case of arabica coffee, can best be achieved probably by: selecting/developing one or two common mapping population(s) using diverse elite parental genotypes, at mutually agreed center(s), followed by extensive DNA typing of the selected population(s) by multiple groups using different sets of markers (including coffee specific SSRs available at their end). This approach would ensure generation of enormous data in relatively short time, and expectedly would provide enough allelic polymorphism to develop a well spread molecular linkage map of arabica genome, which otherwise may remain a mirage.

It is hoped that the above would soon be feasible under the newly formalized International Coffee Genomics Network (ICGN), and would help create the wanting knowledge base and essential DNA resources paving a new and successful route to genetic improvement of coffee. Some immediate applications of these resources (information/tools) would be:

- Availability of rich repertoire of coffee specific PCR based efficient DNA markers would greatly facilitate genetic linkage analysis and germplasm characterization.
- Development of saturated reference molecular linkage maps would help fill the existing gaps in the coffee genetics, and tremendously help complement conventional genetic improvement efforts by the possibilities of MAS based molecular breeding. Availability of well-populated reference maps would also be useful for comparative genomic studies, involving comparisons with interspecific maps of coffee and many other species mainly of Solanaceous crops using anchor markers. Such studies can be useful to understand evolution of coffee gene(s)/genome, in finding synteny or even microsynteny relations with tomato or other Solanaceous (Cornell University) genomes, and utilizing genetic markers/tools developed under the international SOL Genomics network for coffee improvement.
- The success on the above fronts, in turn would make it much easier to identify DNA marker tags/QTLs linked to gene(s) of agronomic importance. Once estab-

lished, these would serve as efficient screening tools for difficult quantitative traits and selection of desirable segregating progenies in a breeding program, thus leading to considerable gains in time and less physical and financial resources.

As a result, coffee improvement may get impetus that will result in development of new coffee cultivars and varieties, which in turn may rejuvenate and lease a new life to the economics of world coffee. This will be helpful in providing proper livelihood and economic sustenance of the farmers as well as satisfaction to the coffee drinkers by having a good brew in their cup.

ACKNOWLEDGEMENTS

Authors thank the Department of Biotechnology, Government of India, New Delhi, India for the financial support under the "1st National Initiative on Coffee Genomics", the Director, CCMB, Hyderabad for the facilities and the encouragement. Thanks are due to Dr R Naidu, Director Research, Coffee Board, Bangalore and colleagues from CCRI, Chikmagalur for providing coffee plant materials used in various studies on marker development, diversity and coffee mapping. PSH thanks CSIR, India for research followship during his Doctoral Research program.

REFERENCES

Aga E, Bekele E, Bryngelsson T (2005) Inter-simple sequence repeat variation in forest coffee trees (*Coffea arabica* L.) populations from Ethiopia. Genetica 124:213–221

Aggarwal RK (2005) DNA fingerprinting of coffee germplasm for conservation and better management and development of molecular map of coffee: molecular characterization and linkage analysis. Completion report, submitted to Department of Biotechnology, New Delhi, India, Project No. BT/PR1302/AGR/08/095/98, p 48

Aggarwal RK, Baruah A, Naik V, Hendre PS, Ashruf A, Rajendrakumar P, Rajkumar R, Annapurna V, Phanindranath R, Prakash NS, Srinivasan CS, Singh L (2004a) Development and characterization of coffee specific microsatellite markers for use as potential genetic markers. In: Proceedings of ASIC 2004: 20th International conference on coffee sciences, Banglore, India, 11–15 October 2004

Aggarwal RK, Rajkumar R, Rajendrakumar P, Hendre PS, Baruah A, Phanindranath R, Annapurna V, Prakash NS, Santaram A, Srinivasan CS, Singh L (2004b) Fingerprinting of Indian coffee selections and development of reference DNA polymorphism panels for creating molecular IDs for variety identification. In: Proceedings of ASIC 2004: 20th international conference on coffee sciences, Banglore, India, 11–15 October 2004

Aggarwal RK, Hendre PS, Varshney RK, Bhat PR, Krishnakumar V, Singh L (2007) Identification, characterization and utilization of EST-derived genic microsatellite markers for genome analyses of coffee and related species, Theor Appl Genet 114:359–372

Agwanda C, Lashermes P, Trouslot P, Combes M-C, Charrier A (1997) Identification of RAPD markers for resistance to coffee berry disease, *Colletotrichum kahawae*, in arabica coffee. Euphytica 97: 241–248

Anim-Kwapong E, Adu-Ampomah Y (2004) Robusta coffee improvement in Ghana: achievements and prospects, 20th conference of ASIC, Bangalore, India

Anthony F (1992) Les ressources génétiques des caféiers: collecte, gestion d'un conservatoire et évaluation de la diversité génétique. Collection TDM (81), ORSTOM (ed), Paris

Anthony F, Bertrand B, Quiros O, Lashermes P, Berthaud J, Charrier A (2001) Genetic diversity of wild coffee (*Coffea arabica* L.) using molecular markers. Euphytica 118:53–65

Anthony F, Combes M-C, Astoga C, Bertrand B, Graziosi G, Lashermes P (2002) The origin of cultivated *Coffea arabica* revealed by AFLP and SSR markers. Theor Appl Genet 104:894–900

Bardini M, Lee D, Donini P, Mariani A, Gianì S, Toschi M, Lowe C, Breviario D (2004) Tubulin-based polymorphism (TBP): a new tool, based on functionally relevant sequences, to assess genetic diversity in plant species. Genome 47:281–291

Baruah A, Naik V, Hendre PS, Rajkumar R, Rajendrakumar P, Aggarwal RK (2003) Isolation and characterization of nine microsatellite markers from *Coffea arabica* L., showing wide cross-species amplifications. Mol Ecol Notes 3:647–650

Berthaud J (1976) Etude cytogcénétiqe d'un haploid de *Coffea arabica* L. Café Cacao Thé 20:91–96

Berthaud J, Charrier A (1988) Genetic resources of *Coffea*. In: Clarke RJ, Macrae R (eds) *Coffee*, Elsevier Applied Science, New York, pp 1–41

Berthou F, Trouslot P (1977) L'analyse du polymorphisme enzymatique dans le genre *Coffea*: adaptation d'une method d'électrophoréses en série. 8th conference of ASIC, Abidjan, Côte d'Ivore, 373–383

Berthou F, Mathieu C, Vedel F (1983) Chloroplast and mitochondrial DNA variation as indicator of phylogenetic relationships in the genus *Coffea* L. Theor Appl Genet 65:77–84

Bettencourt A (1973) Cosideraçoes gerais sobre o "Hibrido de Timor", Instituto Agronomico.

Bhat PR, Krishnakumar V, Hendre PS, Rajendrakumar P, Varshney RK, Aggarwal RK (2005) Identification and characterization of expressed sequence tags-derived simple sequence repeat markers from robusta coffee variety 'C×R' (an interspecific hybrid of *Coffea canephora* and *Coffea congensis*). Mol Ecol Notes 5:80–85

Cardenas J (2001) The world coffee crisis. Opening statement at the first world coffee conference, London organised on the 17 May 2001

Carvalho A (1952) Taxonomia de *Coffea arabica* L. caracteres morfologicas dos haploids. Bragantia 12:201–212

Carvalho A (1988) Principles and practice of coffee plant breeding for productivity and quality factors: *Coffea arabica*. In: Clarke RJ, Macrae R (eds) *Coffee* Vol. 4. *Agronomy*, Elsevier Applied Science, London, pp 129–165

Chaparro AP, Cristancho, MA, Cortina, HA, Gaitán, AL (2004) Genetic variability of *Coffea arabica* L. accessions from Ethiopia evaluated with RAPDs. Genet Res Crop Evol 51:291–297

Chaparro AP, Cristancho MA, Cortina HA, Gaitán AL (2004) Genetic variability of *Coffea arabica* L. accessions from Ethiopia evaluated with RAPDs. Genet Res Crop Evol 51:291–297

Charrier A (1978) La structure génétique das des caféiers spontanés de la région malgache (Mascarocoffea), Mémoires ORSTOM (87), (ed) ORSTOM, Paris

Charrier A, Berthaud J (1985) Botanical classification of coffee. In: Clifford MN, Wilson KC, (eds) Coffee: botany, biochemistry and production of beans and beverage. Croom Helm, London, pp 13–47

Charrier A, Eskes AB (1998) les caféiers. In: Charrier A, Jacquot M, Hamon S, Nicolas, D (eds) L'amélioration des plantes tropicales, Collection Repéres. CIRAD et ORSTOM, Paris, pp 171–196

Combes M-C, Andrzejewski S, Anthony F, Bertrand B, Rovelli P, Graziosi G, Lashermes P (2000) Characterization of microsatellite loci in *Coffea arabica* and related coffee species. Mol Ecol 9: 1178–1180

Coulibaly I, Noirot M, Lorieux M, Charrier A, Hamon S, Louarn J (2002) Introgression of self-compatibility from *Coffea heterocalyx* to the cultivated species *Coffea canephora*. Theor Appl Genet 105:994–999

Coulibaly I, Louarn J, Lorieux M, Charrier A, Hamon S, Noirot M (2003a) Pollen viability restoration in a *Coffea canephora* P and *C heterocalyx* Stoffelen backcross. QTL identification for marker-assisted selection. Theor Appl Genet 106:311–316

Coulibaly I, Revol B, Noirot M, Poncet V, Lorieux M, Carasco-Lacombe C, Minier J, Dufour M, Hamon P (2003b) AFLP and SSR polymorphism in a *Coffea* interspecific backcross progeny [(*C. heterocalyx* × *C. canephora*) × *C. canephora*]. Theor Appl Genet 107:1148–1155

Coutoron E, Lashermes P, Charrier A (1998) First intergeneric hybrids (*Psilanthus ebracteolatus* Hiern × *Coffea arabica* L.) in coffee trees. Can J Bot 76:542–546

Cramer PJS (1957) In: Wellman FL, (ed) Review of literature of coffee research in indonesia. SIC International American Institute of Agricultural Sciences, Turrialba, Costa Rica

Cros J, Combes M-C, Trouslot P, Anthony F, Hamon S, Charrier A, Lashermes P (1998) Phylogenetic analysis of chloroplast DNA variation in *Coffea* L. Mol Phyl Evol 9:109–117

Dufour M, Hamon P, Noirot M, Ristrerucci AM, Brottier P, Vico V, Leroy T (2001) Potential use of SSR markers for *Coffea* spp. genetic mapping. 19th Int Sci Colloquium on Coffee, Trieste, Italy

Etienne H, Anthony F, Dussert S, Fernandez D, Lashermes P, Berthrand B (2002) Biotechnological applications for the improvement of coffee (*Coffea arabica* L.). Vitro Cell Dev Biol-Plant 38:129–138

FAO (1968) FAO mission to Ethiopia 1964–1965. In:FAO, (ed), Report by FAO. Rome, Italy, p 200

Fazuoli LC, Gallo PB, Martins ALM, Guerreiro-Filho O, Medina-Filho HP, Bordignon R, Gonçalves W (2001) Efficiency of early selection in the Icatu coffee. 19th conference of ASIC, Trieste, Italy

Ferwerda FP (1976) Coffee. In: Simmonds NW (ed) Evolution of crop plants. Longman, London, pp 257–260

Grattapaglia D, Sederoff R (1994) Genetic linkage maps of *Eucalyptus grandis* and *Eukalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. Genetics 137:1121–1137

Guillaumet JL, Hallé F (1978) Echantillonnage du materiel *Coffea arabica* récolté en Ethiopia. In: Charrier A (ed) Etude de la structure et de la variabilité génétique des caféiers. IFCC, Paris, Bulletin no. 14, pp 13–18

Gupta PK, Rustigi S, Kulwal PL (2005) Linkage disequilibrium and association studies in higher plants: present status and future prospects. Plant Mol Biol 57:461–485

Hackett CA (2002) Statistical methods for QTL mapping in cereals. Plant Mol Biol 48:585–599

Hendre PS (2006) Development of microsatellite markers and construction of framework linkage map of robusta coffee (*Coffea canephora*). PhD Thesis, Centre for Cellular & Molecular Biology, Hyderabad, submitted to Jawaharlal Nehru University, New Delhi, India, p 230

Herrera JC, Combes M-C, Anthony F, Lashermes P (2001) Efficient use of coffee genetic resources: Molecular analyses of genome interactions in the arabusta hybrid (*Coffea arabica* × *C. canephora*). 19th conference of ASIC, Trieste, Italy

Herrera JC, Combes M-C, Anthony F, Lashermes P (2002a) Gene introgression into *Coffea carabica* by way of triploid hybrids (*C. arabica* × *C. canephora*). Heredity 89:488–494

Herrera JC, Combes M-C, Anthony F. Charrier A, Lashermes P (2002b) Introgression into the allotetraploid coffee (*Coffea arabica* L.): Segregation and recombination of the *C. canephora* genome in the tetraploid interspecific hybrid (*C. arabica* × *C. canephora*). Theor Appl Genet 104:661–668

ICO (2006) International Coffee Organisation, London, http://dev.ico.org/

Kaplinsky R (2004) Competitions policy and the global coffee and cocoa value chains. Paper prepared for UNCTAD

Krug CA, Carvalho A (1951) The genetics of coffee. Adv Genet 4:127–158

Ky C-L, Barre P, Lorieux M, Trouslot P, Akaffou S, Louarn J, Charrier A, Hamon S, Noirot M (2000) Interspecific genetic linkage map, segregation distortion and genetic conversion in coffee (*Coffea* sp.). Theor Appl Genet 101:669–676

Lashermes P, Cros J, Marmey P, Charrier A (1993) Use of random amplified DNA markers to analyse genetic variability and relationships of *Coffea* species. Genet Res Crop Evol 40:91–99

Lashermes P, Combes M-C, Trouslot P, Anthony F, Charrier A (1996a) Molecular analysis of the origin and genetic diversity of *Coffea arabica* L.: Implications for coffee improvement Proc EUCARPIA meeting on tropical plants. Montpellier, pp 23–29

Lashermes P, Coutoron E, Charrier A (1996b) Inheritance and genetic mapping of self incompatibility in *Coffea canephora* Pierre. Theor Appl Genet 93:458–462

Lashermes P, Cros J, Combes M-C, Trouslot P, Anthony F, Hamon S, Charrier A (1996c) Inheritance and restriction fragment length polymorphism of chloroplast DNA in the genus *Coffea* L. Theor Appl Genet 93:626–632

Lashermes P, Trouslot P, Anthony F, Combes M-C, Charrier A (1996d) Genetic diversity for RAPD markers between cultivated and wild accessions of *Coffea arabica*. Euphytica 87:59–64

Lashermes P, Combes M-C, Trouslot P, Charrier A (1997) Phylogenetic relationship of coffee-tree species (*Coffea* L.) as inferred from ITS sequences of nuclear ribosomal DNA. Theor Appl Genet 94:947–955

Lashermes P, Combes M-C, Robert J, Trouslot P, D'Hont A, Anthony F, Charrier A (1999) Molecular characterisation and origin of the *Coffea arabica* L. genome. Mol Gen Genet 261:159–266

Lashermes P, Andrzejewski S, Bertrand B, Combes M-C, Dusseri S, Graziosi G, Trouslot P, Anthony F (2000a) Molecular analysis of introgression breeding in coffee (*Coffea arabica* L.). Theor Appl Genet 100:139–146

Lashermes P, Paczek V, Trouslot P, Combes M-C, Couturon E, Charrier A (2000b) Single-locus inheritance in the allotetraploid *Coffa arabica* L. and interspecifc hybrid *C. arabica* × *C. canephora*. J Hered 91:81–85

Lashermes P, Combes M-C, Prakash NS, Trouslot P, Lorieux M, Charrier A (2001) Genetic linkage map of *Coffea canephora*: effect of segregating distortion and analysis of recombination rate in male and female meioses. Genome 44:589–596

Lin C, Muller LA, Mc Carthy J, Crouzillat D, Pétiard V, Tansksley SD (2005) Coffee and tomato share common gene repertoires as revealed by deep sequencing of seed and cherry transcripts. Theor Appl Genet 112:114–130

Litt M, Luty JA (1989) A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. Am J Hum Genet 44:397–401

Louarn J (1976) Hybrides interspecifiques entre *Coffea canephora* Pierre et *C eugenioides* Moore. Café Cacao Thé 20:433–452

Louarn J (1993) La fertilité des hybrides intrspécifiques et les relations génomiques entre les caféiers diplo d' origines africaines (Genre *Coffea* L. sous-genre *Coffea*). Thèse de l. Université Paris XI, Orsay, France

Masumbuko LI, Bryngelsson T (2006) Inter simple sequence repeat (ISSR) analysis of diploid coffee species and cultivated *Coffea arabica* L. from Tanzania. Genet Res Crop Evol 53:357–166

Mendes AJT, Bachhi O (1940) Observacoes citologicas em *Coffea* V Uma cariedade haploide (di-haploide) de *C. arabica* L. Instituto Agrinomico de Campinas, Boletin Tecnoci, 77

Moncada P, McCouch S (2004) Simple sequence repeat diversity in diploid and tetraploid *Coffea* species. Genome 47:501–509

Montagnon C, Leroy T, Charmetant P, Yap A, Legnate H, Berthaud J, Charrier A (2004) Outcome of two decades of reciprocal recurrent selection applied to *C. canephora* in Côte d'Ivoire: New outstanding hybrids available for growers. 20th conference of ASIC, Bangalore, India

Moreno G (1990) Etude de polymorphisme de l'hybride de Timor en vue de l'amélioration due caféier *arabica*. Thése Docteur-ingenieur, ENSA, Montpellier, France, p 127

Narasimhaswamy RL (1962) Some thoughts on the origin of *C. arabica* L. Coffee, Turrialba: 4:1–5

Orozco-Castillo C, Chalmers KJ, Waugh R, Powell W (1994) Detection of genetic diversity and selective gene introgression in coffee using RAPD markers. Theor Appl Genet 87:934–940

Orozco-Castillo C, Chalmers KJ, Powell W, Waugh R (1996) RAPD and organellar specific PCR re-affirms taxonomic relationship within the genus *Coffea*. Plant Cell Rep 15:337–341

Osorio N (2004) Lessons from the world coffee crisis: a serious problem for sustainable development. Submission to UNCTAD XI, São Paulo, Brazil, June 2004

Paillard M, Lander ES, Pétiard V (1996) Construction of a linkage map in coffee. Theor Appl Genet 93:41–47

Pearl HM, Nagai C, Moore PH, Steiger DL, Osgood RV, Ming R (2004) Construction of genetic map for arabica coffee. Theor Appl Genet 108:829–835

Poncet V, Hamon P, Minier J, Carasco C, Hamon S, Noirot M (2004) SSR cross-amplification and variation within coffee trees (*Coffea* spp.). Genome 47:1071–1081

Ponte S (2001) The "latte revolution"? Winners and losers in the restructuring of the global coffee marketing chain. Centre for Development Research Working Paper No. 1.3. Copenhagen

Prakash NS, Combes M-C, Somana N, Lashermes P (2002) AFLP analysis of introgression in coffee cultivars (*Coffea arabica* L.) derived from a natural interspecific hybrid. Euphytica 124:265–271

Prakash NS, Combes M-C, Dussert S, Naveen S, Lashermes P (2005) Analysis of genetic diversity in Indian robusta coffee genepool (*Coffea canephora*) in comparison with a representative core collection using SSRs and AFLPs. Genet Res Crop Evol 52:333–343

Rajkumar R, Santaram A, Srinivasan CS, Aggarwal RK (2000) Assessment of intra-population genetic variation in few species of *Coffea* using random amplified polymorphic DNA markers. In: Abstracts of PLACROSYM XIV: International Conference on Plantation Crops, Hyderabad, India, 12–15 December 2000, P-143, p 107

Rovelli P, Mettulio R, Anthony F, Anzueto F, Lashermes P, Graziosi G (2000) Microsatellitess in *Coffea arabica* L. In: Sera T, Soccol CR, Pandey A, Roussos S, (eds) Coffee Biotechnology and Quality Kluwer Academic Publishers, Dordrecht, Netherlands, pp 123–133

Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, Mullikin JC, Mortimore BJ, Willey DL et al (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 409:928–933

Saini N, Rajkumar R, Santaram A, Srinivasan CS, Aggarwal, RK (2000) Understanding generic affinities among some of the *Coffea* species and related taxa using ISSR and RAPD markers. In: Abstracts of PLACROSYM XIV: International conference on plantation crops, Hyderabad, India, 12–15 December 2000, 0–42, p 106

Silvarolla MB, Mazzafera P, Fazuoli LC (2004) A naturally decaffeinated arabica coffee. Nature 429:826

Silviera SR, Ruas PM, Ruas CF, Sera T, Carvalho VP, Coelho ASG (2003) Assessment of genetic variability within and among coffee progenies and cultivars using RAPD markers. Genet Mol Biol 26:329–336

Sreenivasan MS (1986) Coffee germplasm in India. J Plantation Crops 16:313–319

Sreenivasan MS, Ram AS, Prakash NS (1993) Tetraploid interspecific hybrids in coffee breeding in India. 15th conference of ASIC, Montpellier, France, pp 226–233

Steiger DL, Nagai C, Moore PH, Morden CW, Osgood RV, Ming R (2002) AFLP analysis of genetic diversity within and among *Coffea arabica* cultivars. Theor Appl Genet 105:209–215

Tautz D (1989) Hypervariablity of simple sequences as a general source of polymorphic DNA markers. Nucleic Acids Res 17:6463–6471

Van de Vossen HAM (1985) Coffee selection and breeding. In: Clifford MN, Wilson KC, (eds) Coffee, botany, biochemistry and production of beans and beverage. Croom Helm, London & Sydney, pp 48–97

Van de Vossen HAM (2004) Coffee breeding and selection: Review of achievements and challenges. 20th conference of ASIC, Bangalore, India

Vos P, Hogers R, Bleeker M, Reijans M, Van der Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23:4407–4414

Wasserman M (2002) Trouble in coffee lands. Reg Rev Q2:4–12

Weber JL, May, PE (1989) Abundant class of human DNA polymorphism, which can be typed using the polymerase chain reaction. Am J Hum Genet 44:388–396

Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res 18:6531–6535

Wyman AR, White RA (1980) A highly polymorphic locus in the human zeta-globin gene complex. Proc Natl Acadl Sci USA 80:5022–5026

CHAPTER 16

GENOMICS OF ROOT NODULATION IN SOYBEAN

KYUJUNG VAN[1], MOON YOUNG KIM[1,2] AND SUK-HA LEE[1,2,*]

[1]*Department of Plant Science, Seoul National University, Seoul 151-921, The Republic of Korea*
[2]*Research Institute for Agriculture and Life Sciences, Seoul National University, Seoul 151-921, The Republic of Korea*

Abstract: Soybean is a suitable crop material for studying root nodulation and full genome sequencing because of its economic value. This review introduces the "nodulation" phenomenon that occurs in legume root systems such as the soybean. In addition, the paper identifies and discusses nodulation mutants (e.g., non-nodulation, ineffective nodulation, and super-/hypernodulation) and the genetic loci that control nodulation. The advent of genomics, proteomics, metabolomics, etc., has greatly contributed in improving our understanding of the symbiotic interactions between legume plants and Rhizobia, particularly for the identification of nodulation-related genes. Furthermore, molecular gene identification should be combined with biochemical pathways for nodulation in order to better understand the symbiotic interactions between legume and Rhizobia.

1. INTRODUCTION

The Leguminosae plant family is the third largest of the angiosperms, has spread to every continent, adapted to the tropics and the Arctic and comes in different varieties ranging from annual herbs to trees (Long 1989). *Rhizobium* and *Bradyrhizobium* are unique among microorganisms in their ability to develop nitrogen-fixing nodules on the roots of leguminous plants, which may be the most highly-evolved system for fixing nitrogen from the atmosphere.

Legume plants have two main sources of nitrogen including combined nitrate or ammonia nitrogen and biologically-fixed nitrogen from the atmosphere in the root nodule. As biological nitrogen fixation is of agronomic importance reducing the need for chemical nitrogen fertilizer for agriculturally-important crops such as

*Author for correspondence: sukhalee@snu.ac.kr

R.K. Varshney and R. Tuberosa (eds.), Genomics Assisted Crop Improvement: Vol. 2: Genomics Applications in Crops, 435–452.

soybean and alfalfa (Stacey et al. 2006), the utilization of nitrogen fixation in the root nodule is given considerable emphasis. Legume plants show a wide range of nodulation response: non-nodulation, ineffective nodulation, normal nodulation and super-/hypernodulation. In addition to the biological importance of these mutants for characterizing the molecular basis of symbiotic association, non-nodulating mutants can be used as a control to estimate the biological N_2 fixation. Furthermore, super-/hypernodulating mutants can reduce the nitrogen fertilizer requirements for crops, thereby reducing the impact of agricultural practices on the environment.

This paper deals with the genetic and molecular analysis of soybean symbiotic mutants. Emphasis is placed on the genetic and genomic components of the nodulating soybean mutant loci and the methods used in identifying the genes involved in nodulation.

2. NODULATION IN LEGUME AND SOYBEAN: PLANT-*RHIZOBIUM* INTERACTION

Many recent studies have focused on legumes rather than the model plant *Arabidopsis thaliana* because legumes show symbiotic relationships between plants and soil-borne bacteria and they have *Rhizobia* for nitrogen fixation (Sato and Tabata 2006; Shoemaker et al. 2006; Town 2006). In terms of genome full sequencing, the soybean draws more attention than other leguminous plants because of its economic importance in the symbiotic interactions between plants and *Bradyrhizobium* (Sutton 1983). During this symbiotic relationship, *Rhizobia* provide the host plant with nitrogen in the form of ammonia (as do the amides in the alfalfa and pea nodules and the ureides allantoin and allantoic acid in the soybean and cowpea nodules) inside the N_2 -fixing nodules on roots of leguminous plants; in turn, they receive carbohydrates and other compounds from the plant (Long 2001; Riely et al. 2004). This symbiotic nitrogen fixation is beneficial to the farmers because it leads to a significant decrease in the use of N-containing fertilizers (Riely et al. 2004; Stacey et al. 2006). External factors, such as light intensity, temperature, pH, soil moisture, etc., closely regulate the symbiotic formation of nodules (Lie 1974; Lee et al. 1998); in legume plants, the amount of nitrogen fertilizer also controls nodulation and nitrogen fixation. Externally-supplied nitrogen in the form of nitrate prevents symbiosis by inhibiting the infection process of the nitrogen-fixing bacteria (Carroll and Mathew 1990; Lee et al. 1998). Internally, this symbiosis is regulated by 'autoregulation' (Delves et al. 1986; Van et al. 2005; Lestari et al. 2005, 2006b), the primary functions of which are to control the number and growth of the nodules on the root; it functions by initiating the nodule primordial signal to the leaf, which produces a shoot-derived inhibitor for to restrict the progress of younger nodule primordia (Lee et al. 1991; Gresshoff 2003; Van et al. 2005).

The symbiosis between the soil bacteria and leguminous plants is characterized by a specific multi-step signal cascade (Israel et al. 1986); the first step is the release of flavonoids from the plant root to attract compatible *Rhizobia* (Schultze and Kondorosi 1998; Riely et al. 2004; Geurts et al. 2005; Stacey et al. 2006). Flavonoids

induce nodulation (nod) genes to encode enzymes for the synthesis of specific lipochitooligosaccharides nodulation signals (Nod factors) (Dénarié et al. 1996; Long 1996; Oldroyd et al. 2001). In turn, the nod factors trigger many initial steps of the signal transduction pathway for the root hair infection process after these factors are perceived by plant receptors (Riely et al. 2004; Stacey et al. 2006). Although flavonoid nod gene-induction is required for nodulation, precise nodulation can only occur if the nod genes are appropriately expressed quantitatively, spatially and temporally (Loh and Stacey 2003). The strong constitutive expression of the nodulation genes leads to defective, reduced nodulation phenotypes (Knight et al. 1986). These results indicate that the nod genes are involved in both positive activation and negative regulation of the *nod* genes (Cren et al. 1995; Fellay et al. 1998; Kondorosi et al. 1989). Nol A and nod D2 are two key components in the negative regulation of the nodulation genes in *B. japonicum*. In addition, Loh and Stacey (2003) suggest that a key role it has multiple forms the *nol A* gene plays in allowing *B. japonicum* to nodulate its plant hosts. The Nod signal or components of the nod signal and the feedback regulated the nodulation genes of *B. japonicum* by the induction of nol A by chitin led to an increase of the repressor protein, nod D2 (Loh and Stacey 2001). The nodulation genes are feedback-regulated by the nod signal, which contains either a tetrameric or a pentameric chitin backbone (Loh and Stacey 2003).

These nod factors initiate root hair curling and induce the formation of nodule primordia by activating the root cortical cell cycle; this is then followed by the bacterial infection into the root hairs by the infection threads (Schultze and Kondorosi, 1998; Stacey et al. 2006). When the infection thread reaches the cells of the developing primordium, the bacteria are released into symbiosomes and then are differentiated into bacterioids inside a plant cell (Schultze and Kondorosi 1998; Stacey et al. 2006). This microaerobic nodule environment is able to conduct nitrogen fixation because of the oxygen permeability carrier; inside the infected cell, an oxygen-binding protein, leghemoglobin, regulates and delivers oxygen (Lestari et al. 2005). Since differentiated structures of various nod factors and polysaccharides are present, host specificity is possible because of differences in the substrate specificity of the enzymes (Schultze and Kondorosi, 1998). Thus, high host specificity is a very important control for the interaction between plants and soil borne-bacteria, which include *Rhizobium*, *Mesorhizobium*, *Azorhizobium*, *Sinorhizobium*, and *Bradyrhizobium* (Schultze and Kondorosi 1998). Consequently, the formation of effective soybean nodules is a complex and highly-regulated process that requires the production and exchange of specific molecular signals between the host plant and the bacterial symbiont (Kosslak et al. 1987).

Among the signal components for symbiotic relationships, in nodule development ethylene is a negative regulator in the premature arrest of Rhizobial infection in pea (Lee and LaRue 1992), *Lotus*, and *Siratro* (Nukui et al. 2000). It has been suggested that aminoethoxyvinylglcyine (AVG) or Ag^{+}, which are inhibitors of ethylene formation and perception, might stimulate nodule production in *Medicago truncatula* roots (Schmidt et al. 1999); however, ethylene does not always act as a

nodule inhibitor in every legumes. Treatment with AVG and exogenous ethylene did not inhibit nodulation in soybeans although ethylene production was increased by infection by *B. japonicum* (Lee and LaRue 1992). In addition, Schmidt et al. (1999) compared the nodule development in soybean mutants that were either less responsive to ethylene or deficient in the usual number of nodules having Ag^+. Neither the number of nodules nor ethylene sensitivity differed from the wild type, which indicates that ethylene is not significant in regulating soybean nodule formation. Thus, the feedback regulation is the other component in the signal cascade in the soybean nodule development by suppressing the development of younger nodules by the older nodules (Bhuvaneswari et al. 1981; Pierce and Bauer 1983).

3. NODULATION MUTANTS IN SOYBEAN

3.1. Isolation of Nodulation Mutant

The nodulation mutants of soybeans have been obtained either from the normal soybean population or from ethylmethane sulfonate (EMS), sodium azide or gamma irradiation mutagenesis (Carroll et al. 1985a, 1985b; Gresshoff and Delves 1986; Vance et al. 1988). Although both gamma rays and sodium azide mutagenesis increased the mutation frequency for chlorophyll deficiency, EMS treatment caused a 10-fold increase in the frequency of mutation. The M_2 generation of 'Bragg' seed that was mutagenized with EMS showed a chlorophyll deficiency of about 2.8% (Gresshoff and Delves 1986). The isolated mutants showed three major phenotypes: non-nodulation, ineffective nodulation and super-/hypernodulation.

Non-nodulating mutants have been isolated in several legume species including peas (Kneen and LaRue 1984, 1988; Borisov et al. 2003), chickpeas (Davis et al. 1985), common beans (Davis et al. 1988) and sweet clover (Kneen and LaRue 1988). In model legume plants, the *nodule inception* (*nin*) mutant of *L. japonicus* and *M. truncatula* hair curling (*hcl*) mutants were identified, and these mutants showed impaired nodule development or the upstream of signal transduction pathway for symbiotic interaction (Schauser et al. 1999; Catoira et al. 2001). Other non-nodulating model legume mutants were isolated by Stacey et al. (2006). A line with naturally-occurring, non-nodulation was also reported in soybeans (Williams and Lynch, 1954), and the non-nodulation trait was conditioned by a single recessive gene (*rj*). Root hair curling occurred, but infection threads did not develop in this mutant (Tanners and Anderson 1963). A soybean mutant, screened by Clark (1957) exhibited no nodulation in soil cultures but did exhibit occasional and irregular nodulation in sand-nutrient solution culture. Mathews et al. (1987) isolated three non-nodulation mutants (nod 49, nod 139 and nod 772) from Bragg after the mutagenesis of seeds with EMS and selected for non-nodulation in the M_2 in the presence of *B. japonicum* USDA 136. Mutants nod 49 and nod 139 lacked curled root hairs when inoculated with *B. japonicum* USDA 110; the mutant nod 772 showed occasional hair curling. Complementary tests indicated that nod 49 and nod 772 were allelic to rj_1, but nod 139 was not (Delves et al. 1988).

Three possible reasons for the occurrence of non-nodulation in these mutants were suggested by Gresshoff and Delves (1986). First, the non-nodulation mutants had altered production of substances in root exudates. These root exudates interact with the *nod D* gene product to behave as an activator of the nodulation genes in the nitrogen-fixing bacteria. Second, these mutants may be less sensitive to the bacterial signal during the infection processes. Third, there was a strong active defense response of these mutants against bacterial infection (Gresshoff and Delves 1986).

Ineffective nodulation of soybean was conditioned by dominant alleles. The Rj_2, Rj_3, and Rj_4 genes showed ineffective nodulation with specific *Bradyrhizobium* strains. All of these dominant alleles condition the formation of small white nodules (Gresshoff et al. 1987). Non-nodulation and ineffective mutants that were defective in nodule formation are useful for understanding interactions between the host and bacteria. In addition to study the underlying mechanisms of symbiosis, these mutants can be also used as controls for measuring the contribution of symbiotic N_2 fixation to the total N economy of plant (Gresshoff et al. 1987).

Super- or hyper-nodulating mutants with enhanced nodulation and partial tolerance of nodulation in the presence of high levels of exogenous NO_3 have been developed in several legume crop plants such as peas (Jacobsen and Feenstra 1984), white beans (Park and Buttery 1988) and soybeans (Carroll et al. 1985a, 1985b). *L. japonicus* also has a hypernodulating mutant, *sym78* mutant *hypernodulation aberrant root1-1* (*har1-1*) encoded an LRR-RLK, which was characterized by high homology with CLAVATA1 (Krusell et al. 2002; Nishimura et al. 2002). *M. truncatula* super numeric nodules mutant also encoded SUNN proteins similar to HAR1 proteins (Schnabel et al. 2005). While limited information is available for peas and white beans, supernodulation has been well characterized in soybean (Carroll et al. 1985a, 1985b; Gresshoff and Delves 1986; Gresshoff et al. 1987; Searle et al. 2003).

Several groups of scientists have independently isolated super- or hypernodulating soybean mutants by using chemical mutagenesis from Bragg (Carroll et al. 1985a, 1985b), Williams (Gremaud and Harper 1989), Enrei (Akao and Kouchi 1992) and Sinpaldalkong 2 (Lee et al. 1997). The mutagen EMS has been commonly applied to produce several super-/hypernodulating mutants. More recently, an induced mutagenesis using fast neutron resulted in a supernodulating soybean mutant (Men et al. 2002). Supernodulating soybean mutants produce ten times the number of nodules as wild types, whereas hypernodulating mutants have two to five times as many nodules as wild types and exhibit a nodulation between supernodulation and wild-type nodulation (Kim et al. 2005). The contrast in nodulation responses between the super-/hypernodulating mutants and the wild type from which they are derived provides unique biological material for the study of host plant factors, which may be involved in the nodulation process.

Suppression of nodulation in the younger regions of the root caused by prior inoculation of more mature regions of the root has been referred to as an autoregulatory response (Pierce and Bauer 1983). However, the autoregulatory response governing nodulation was not present in supernodulating soybean mutants (Olsson

et al. 1989). Thus, the mutated gene of the *nts* mutant may be the same gene or closely linked to the gene that altered the autoregulatory signal. A general model for the autoregulation of nodulation was proposed based on the presence of a symbiotic nodulation inhibitor in the wild type (Gresshoff and Delves 1986). In wild type plants, the shoot produces a signal after it is stimulated by the root-derived compound, which is a product of nodule meristematic center. The shoot-derived signal acts as an inhibitor of symbiotic development at the continued cell division stage. The supernodulating mutant was thought to lack the ability to produce shoot-derived inhibitors.

Genetic studies and linkage mapping for these supernodulating mutants revealed that the mutated locus is located on linkage group (LG) H and contains separate alleles (Landau-Ellis et al. 1991; Kolchinsky et al. 1997). More recently, map-based cloning with markers pUTG-132 and UQC-IS1, which flank *NTS-1* on LG H, was performed to identify a gene conferring nodulation autoregulation (Searle et al. 2003). The expression of this gene in shoots is responsible for controlling the nodule meristem proliferation and communicating with distant nodules; thus is called the *Glycine max* nodule autoregulation receptor-like protein kinase (*GmNARK*) gene. Nonsense and missense mutations that comprise single-base changes on the gene were found to lead to supernodulation and hypernodulation phenotypes in the mutants (Searle et al. 2003).

3.2. Genetic Loci Controlling Nodulation in Soybean

Even though much is known about the bacterial contribution to the symbiotic nodulation of legumes, knowledge of the host contribution has thus far received less attention in the identification of the plant genes involved in nodulation. A few dozen of genes are assumed to be involved in nodulation and N_2 fixation, and range from host plant recognition to nodule formation. Since the nodulation trait has been recognized as governed by qualitative genes, classical genetic approaches have generally been applied to identify those genes. As mentioned above, several genetic loci for nodulation in soybean have been found by using the chemical mutants altered in the level of nodulation. William and Lynch (1954) first reported the recessive gene rj_1 in a spontaneous non-nodulating mutant that controls a restriction of nodulation with almost all *Bradyrhizobia* (Table 1). Since then, a serial of the *rj* loci from rj_2 to rj_7 have been identified as responsible for ineffective or redundant nodulation (Table 1). Three dominant genes, Rj_2 (Caldwell 1966), Rj_3 (Vest 1970), and Rj_4 (Vest and Caldwell 1972) controlled strain-specific restricted nodulation, which possessed genetic independence (Devine and O'Neill 1986). Mapping the genes that contribute to this phenotype change in nodule formation was conducted to discover which genes were altered resulting in non-nodulation (Matthews et al. 2001). The non-nodulation genes rj_1, Rj_2 and Rj_4 were mapped on separate linkage groups (LG D1b+CLG11, LG J and GmEST-BFM) depending on search results of the breeder's toolbox at SoyBase (http://soybeanbreederstoolbox.org), however, the Rj_3 gene has not yet been mapped. In the early stages of nodulation, the non-nodulation genes in soybeans may be involved in recognition of host plant to bacterial infection because

Table 1. List of the nodulation genes defined in soybean

Gene	Function	LG	References
rj_1	Non-nodulation	D1b+CLG11	http://soybeanbreederstoolbox.org/
Rj_2	Non-nodulation	LG J	Lohnes et al. 1993
Rj_3	Non-nodulation	ND[1)]	–
Rj_4	Non-nodulation	GmEST-BFM	Matthews et al. 2001
Rj_5	Normal nodulation	ND[1)]	–
rj_6	Non-nodulation	ND[1)]	–
rj (*nts*)	Super-/hypernodulation	LG H	Kolchinsky et al. 1997

[1)] Not determined

they condition specificity or susceptibility of certain strains of *B. japonicum*. The additional rj_6 non-nodulation gene has been proposed in the two mutants, nod 139 and NN 5, which were chemically mutagenized from William (Pracht et al. 1993), while the Rj_5 locus was symbolized as a normal nodulation of Harosoy 63. In contrast, the recently described locus rj_7 has been designated as super-/hypernodulation of nod-type mutant and the En 6500 mutant (Vuong and Harper 2000). This locus should be same as *nts* mutant locus derived from mutagenized Bragg. Nevertheless, few molecular studies on soybean mutants that are defective in nodule formation have been performed, which contrast with studies on the legumes, *L. japonicus*, *M. truncatula* and *P. sativum* (Stacey et al. 2006).

Unlike *rj*-type loci, extensive studies have been performed on the other locus *nts* involved in excessive nodule proliferation (Table 1). A remarkable achievement in positional cloning of the nts gene conferring super- or hypernodulation was accomplished by Gresshoff and his colleagues (2003); *nts* was the first gene that was isolated in the soybean. This gene, *GmNARK*, is responsible for autoregulation of nodule formation.

3.3. Application of Molecular Markers for Enhancing Nodulation

The evaluation of nonsense and missense mutations of the *GmNARK* gene led to supernodulation and hypernodulation mutants (Searle et al. 2003). The single nucleotide polymorphism (SNP) of *GmNARK*, which is between supernodulating mutant SS2-2 and wild-type Sinpaldalkong 2, was converted into specific PCR marker (Kim et al. 2005). The sequence mutation of A→T in *GmNARK* between SS2-2 and Sinpaldalkong 2 directly affect phenotypic variations of nodulation. SNP markers that are directly linked to a specific trait would be a very useful DNA marker and provide the advantages of complete linkage with no error in selection.

It is becoming clear that SNPs have great potential as genetic markers to directly map important genes and facilitate selecting desirable traits, or excluding deleterious traits in crop breeding (Kim et al. 2004). Nodulation traits in the soybean are difficult to assay phenotypically because of their subterranean nature and the requirement to inoculate with *Bradyrhizobia*. Thus, the single nucleotide amplified

polymorphic (SNAP) marker for supernodulation is able to distinguish between normal nodulation and supernodulation in both the cross populations of SS2-2 and soybean genotypes (Kim et al. 2005). Marker-assisted selection of this SNAP marker to improve the efficiency of nodule trait selection is widely applicable in soybean breeding activities including parent assessment, recovery of recurrent parent in backcrossing and trait selection in single seed descendents at early generation.

4. GENE EXPRESSION PROFILING FOR NODULATION

Classically, nodulating plants were biochemically and physiologically analyzed by examining the mechanisms of carbon supply, nitrogen fixation, ammonia assimilation and nitrogen transport, such as the expression of leghemoglobin (Gresshoff 2003). As molecular techniques were developed, differential screening was widely adopted to survey gene expression profiling with cDNA libraries (Table 2). Mylona et al. (1995) identified early nodulins and late functions that are associated with nitrogen fixation, such as leghemoglobin, glutamine synthetase and uricase. Nodule-specific cDNAs that are shown to be induced or enhanced during the first days of the symbiotic interaction were also identified by means of differential display (Goormachtig et al. 1995). Other methods of differential screening, such as differential hybridization, subtractive hybridization, differential display and EST and cold-plaque screening were used to study nodulation researches with various leguminous host plants. With the advantage for surveying tissue specificity, various plants could be screened for early nodulation-related genes (see Lievens et al. 2001). Gene expression patterns that were identified by differential screening were confirmed by using northern blot analysis, RT-PCR, RNase protection assay or quantitative RT-PCR for lower expression levels (Lievens et al. 2001; Gresshoff 2003).

The cDNA-amplified fragment length polymorphism (cDNA-AFLP) approach was used to study differential gene expression of nodulation (Simoes-Araujo et al. 2002; Kistner et al. 2005; Lestari et al. 2006a). Compared to differential display, this gel-based method could overcome problems such as low reproducibility, difficulty in detecting rare transcripts and generation of false positives (Bachem et al. 1998). The cDNA-AFLP approach was also used to differentially discriminate genes expressed on a large scale because direct side-by-side comparison among samples is possible, and prior sequence information is not required (Kistner et al. 2005; Lestari et al. 2006a). Simoes-Araujo et al. (2002) surveyed the expression profiles from cowpea nodules during heat stress. From about 600 bands, 55 transcripts-derived fragments (TDFs) were up-regulated and nine were down-regulated; they show similarities to wound-induced proteins, disease resistance protein, etc. Using cDNA-AFLP, 344 differentially-expressed genes were detected from about 7000 TDFs generated from *L. japonicus* roots (Kistner et al. 2005). Approximately 2% of the TDFs were up-regulated during the arbuscular mycorrhiza symbiosis, and a ser-protease and a cys-protease were activated during root nodule development. More specifically, 65 TDFs from Sinpaldalkong 2 (wild type), and 40 TDFs from SS2-2

Table 2. Lists of gene expression profiling in nodulation research

Methods	Species	References
Differential screening	*Sesbania rostrata* (Differential display)	Goormachtig et al. 1995
	S. rostrata(Differential display)	Lievens et al. 2001
	Various species	Lievens et al. 2001
cDNA-AFLP	Cowpea	Simoes-Araujo et al. 2002
	L. japonicus	Kistner et al. 2005
	Soybean	Lestari et al. 2006a
SAGE	*L. japonicus*	Asamizu et al. 2005
Microarray	*L. japonicus*	Colebatch et al. 2002, 2004 Kouchi et al. 2004
	M. truncatula	El Yahyaoui et al. 2004 Hohnjec et al. 2005 Lohar et al. 2006 Küster et al. 2004, 2006 Mitra and Long 2004 Mitra et al. 2004 Starker et al. 2006
	Soybean	Maguire et al. 2002 Lee et al. 2004
In Silico transcriptome	*M. truncatula*	Fedorova et al. 2002 Tesfaye et al. 2006
Proteomics	Clover	Morris and Djodjevic 2001
	M. truncatula	Bestel-Corre et al. 2002 Mathesius et al. 2001, 2002, 2003 Rolfe et al. 2003
	Melilotus alba	Mathesius et al. 2002
	Pea	Saalbach et al. 2002
	Soybean	Hwang, C.H. (personal communication) Panter et al. 2000 Wan et al. 2005 Winzer et al. 1999
	Trifolium subterraneum	Mathesius et al. 2002
	White sweetclover	Natera et al. 2000

(supernodulating mutant) were differentially expressed in response to *B. japonicum* symbiotic relationships (Lestari et al. 2006a). Interestingly, *GmNARK*, which was previously known as involved in autoregulation of nodulation, was detected by this study, thus confirming cDNA-AFLP as a powerful method for identifying interesting genes that had no prior assumptions.

Serial analysis of gene expression (SAGE) was also used to compare the gene expression patterns between roots and nodulating roots in *L. japonicus*; it should be mentioned that unequivocal gene identification through SAGE may be cumbersome because the technology is based on a 14-base pair tag sequence (Asamizu et al. 2005). Unlike cDNA-AFLP, SAGE was only deployed in the model

plants Arabidopsis and rice because neither EST nor genome sequences from these organisms were available. To model leguminous plants, *L. japonicus* was investigated for the transcript profiles because of the EST sequencing availability in the public databases as well as portions of the genome sequences. From SAGE libraries constructed from uninfected and nodulating roots, approximately 400 tag species from more than 80,000 SAGE tags were significantly expressed either in the uninfected or nodulating roots. This study also showed that transcriptional induction was different among leghemoglobin gene paralogs (Asamizu et al. 2005).

Molecular genetic technologies such as gene tagging and gene sequencing helped the understanding of the nodulation mechanism, the *Nod* and *Nol* genes for nodulation and the fix genes for nitrogen fixation (Gresshoff 2003). After that, microarray transcript analysis was developed to survey genes with an active role in nodulation (Table 2) as well as map-based the gene cloning method for specific genes (Endre et al. 2002; Stracke et al. 2002; Searle et al. 2003). Many studies on gene expression profiling by microarray were conducted with the model legume, *M. truncatula*. To identify genes expressed in the developing and mature nodules, and to detect differential expression of those genes, wild type and mutants were used to identify differentially-expressed genes in the symbiotic relationship between *M. truncatula* and *S. meliloti* (El Yahyaoui et al. 2004; Küster et al. 2004; Manthey et al. 2004). During nodule initiation, formation and function, El Yahyaoui et al. (2004) identified more than 750 differentially-expressed genes grouped into four categories: strongly up-regulated both in young and mature nodules, down-regulated during nodulation, activated mainly in mature nodules and genes that were transiently induced three to four days after inoculation. Additional genes were identified using a number of mutants of *M. truncatula* and *S. meliloti* (Mitra and Long 2004; Mitra et al. 2004; Starker et al. 2006). Among these studies, Mitra et al. (2004) and Starker et al. (2006) used an Affymetrix "symbiosis" gene chip to obtain about 10,000 *M. truncatula* tentative consensus sequences (TCs) that were available from TIGR and probed with the *S. meliloti* genome. The genes activated in the first 72 hours of nodule initiation were identified by a 6,000 unigene *M. truncatula* cDNA microarray (Lohar et al. 2005). In the other model legume, *L. japonicus*, cDNA microarray was used to survey the genes responsible for global changes in the development of root nodules and the metabolic differentiations toward nitrogen fixation (Colebatch et al. 2002, 2004). In addition, Kouchi et al. (2004) investigated gene expression profiling in early stages of root nodule development.

In silico-based gene discovery and tissue profiling with EST, collections have been performed for many plant species. Since EST databases of legume model plants were only available in limited numbers, an *in silico*-approach has been rarely adopted (Table 2). Fedorova et al. (2002) identified nodule-specific genes from *M. truncatula* EST databases (http://www.tigr.org/tdb/mtgi) that contained more than 140,000 sequences generated from 30 cDNA libraries. There were 340 putative TCs that were expressed only in root nodules; nodule-specific TCs were subdivided into nine groups by the predicted function, such as nodulins, plantacyanin, agglutinin, embryo-specific protein, calmodulin-like proteins, purine permease, etc. Both

studies used macroarrays and Northern blot analysis to experimentally confirm gene expression identified by *in silico* transcriptome profiling. Combined with targeted macroarrays, Tesfaye et al. (2006) used the same approach and identified 81 nodule-specific TCs that were involved in carbon and nitrogen metabolism under different conditions of nitrogen supply.

Proteomics that were carried out by protein separation by two-dimensional gel electrophoresis, multi-dimensional liquid chromatography, peptide sequencing by mass spectrometry, and bioinformatics were more developed (Trevaskis et al. 2002; Jorrín et al. 2006). Most research focused on *M. truncatula* and soybean because the EST and genome sequencings were intensively performed although various legume species were used for studying differential gene profiling in nodulation (Table 2). The proteome profiles were compared between three legumes (*M. truncatula*, *Melilotus alba* and *Trifolium subterraneum*) to evaluate cross-species similarities (Mathesius et al. 2002), but publicly available *M. truncatula* EST databases only identified the proteins and peptide mass fingerprinting. Therefore, Mathesius et al. (2001) were able to efficiently identify and analyze 485 proteins by peptide mass fingerprinting; 179 matched with the *M. truncatula* EST databases. *M. truncatula* root protein profiles were analyzed by time-course after inoculation with either *S. meliloti* or the arbuscular mycorrhizal *Glomus mosseae* (Bestel-Corre et al. 2002). Leghemoglobin that was identified by matrix-assisted laser desorption/ionization-time of flight-mass spectrometry after trypsin digestion was induced in nodulated roots of *M. truncatula*. Mathesius et al. (2003) surveyed proteins in response to nanomolar to micromolar concentrations of bacterial N-acyl homoserine lactone (AHL), which play important roles in eukaryote-prokaryote interactions. More than 150 proteins were significantly changed, and 99 were identified by peptide mass fingerprinting.

Using proteomic analysis, soybean peribacteroid membrane (PBM) proteins possibly involved in protein processing were identified (Panter et al. 2002). These results suggested that these PBM homologs might play a role in protein translocation, folding, maturation or degradation in symbiosomes. A nodule-specific 53-kDa late nodulin from a soybean symbiosome membrane was also identified by proteomics, and its function and expression were confirmed by Northern blot analysis (Winzer et al. 1999). Wan et al. (2005) surveyed protein profiles of root hairs inoculated with *B. japonicum*. Of 37 protein spots, 27 were identified by database comparisons, and included both the proteins response to rhizobial inoculation and novel proteins. Comparative proteomics using a supernodulating mutant showed that malate dehydrogenase, peroxidase and the alcohol dehydrogenase in particular were up-regulated during nodulation in both the wild and supernodulating mutants (Hwang C.H. personal communication).

Other than *M. truncatula* and soybean, clover, pea and white sweet clover were used to investigate the *Rhizobium*-legume symbiotic relationship (Table 2). Natera et al. (2001) generated proteome maps after comparison between free-living bacterium grown in the culture and the bacteroid form isolated from root nodules. More than 250 proteins, such as leghemoglobin and NifH (nitrogenase H), were

induced or up-regulated in the nodule compared with the root, wherein about 350 proteins, such as glutamine synthetase, urease, a urea-amide binding protein and a PII (nitrogen regulatory protein) isoform were down-regulated in the bacteroid form. After isolating the PBM and peribacteroid space fractions from pea root nodules, 46 of 89 protein spots were identified through a database search (Saalbach et al. 2002). Endomembrane proteins, such as V-ATPase, BIP and an integral membrane protein, were detected in the PBM fraction, which suggests a role for the endomembrane system in PBM biogenesis. In clover, Morris and Djordjevic et al. (2001) identified 16 protein spots differentially expressed. Eight of 22 constitutively-expressed spots were similar to the putative protein homologies that were involved in several pathogenesis and stress-related proteins. With these results, the proteome maps were able to construct in a cultivar-specific legume-*Rhizobium* interaction.

All molecular technologies and knowledge, such as the transcriptome, proteome, metabolome and the study of mutations coupled with reverse genetics, can be used to identify genes that more intensely confer nodulation and symbiotic relationships. The characterized genes and proteins identified by these recently developed methods should be studied with the biochemical pathway for nodule formation to clarify the legume-*Rhizobium* interaction.

5. CONCLUSIONS AND PERSPECTIVES

Symbiotic N_2-fixation is a complex physiological process that is influenced by the interaction of genetic elements between legume plants and nitrogen-fixing bacteria. While substantial progress has been made in identifying bacterial genes and gene products that contribute to nodulation and N_2 fixation, relatively little progress has been made in understanding the contribution that host plant genes make to symbiosis. Host plant genetic control of symbiotic association has been well documented through classical genetic studies; however, physiological and biochemical manifestations of these genes in the host were poorly understood.

The genes identified in the nodulation soybean mutants conditioned four major phenotypes: 1) non-nodulating, 2) ineffective nodulation (little or no N_2 fixation), 3) hypernodulating and 4) supernodulating. These mutants are unique biological materials for characterizing the host plant regulation of nodulation and nitrogen fixation.

With the rapid development of functional plant genomics, the large numbers of expressed sequence tags (EST) are publicly available for soybean plants. In addition, recent advances in cDNA, oligonucleotide and Affymetrix microarray for large transcriptomic studies on nodulation, mass spectrometry (MS) techniques combined with two dimensional polyacrylamide gel electrophoresis (2-D PAGE) and the establishment of protein database are feasible for identifying novel candidate genes that might be implicated in nodulation.

The availability of nodulation mutants in soybeans in combination with the recent advances in omics approaches will accelerate our understanding of the molecular basis of symbiotic association by the functional characterization of nodule biology.

The integration of gene identifications and the basic biochemical pathway for nodulation would be helpful for a more complete understanding of the signaling cascade for the symbiotic relationships between legume plants and Rhizobia.

ACKNOWLEDGEMENTS

This research was supported by a grant (code no. CG3121) from the Crop Functional Genomics Center of the 21st Century Frontier Research Program funded by the Ministry of Science and Technology (MOST), Republic of Korea. Dr. K. Van is recipient of a fellowship from the BK21 program granted by the Ministry of Education & Human Resources Development (ME & HRD), the Republic of Korea. We also thank the National Instrumentation Center for Environmental Management at Seoul National University in Korea.

REFERENCES

Akao S, Kouchi H (1992) A supernodulating mutant isolated from soybean cultivar Enrei. Soil Sci Plant Nutr 38:183–187

Asamizu E, Nakamura Y, Sato S, Tabata S (2005) Comparison of the transcript profiles from the root and the nodulating root of the model legume *Lotus japonicus* by serial analysis of gene expression. Mol Plant Microbe Interact 18:487–498

Bachem CWB, Oomen RJFJ, Visser RGF (1998) Transcript imaging with cDNA-AFLP: a step-by-step protocol. Plant Mol Biol Reporter 16:157–173

Bestel-Corre G, Dumas-Gaudot E, Poinsot V, Dieu M, Dierick JF, Tuinen DV, Remacle J, Gianinazzi-Pearson V, Gianinazzi S (2002) Proteome analysis and identification of symbiosis-related protein form *Medicago truncatula* Gaertn by two-dimensional electrophoresis and mass spectrometry. Electophoresis 23:122–137

Bhuvaneswari TV, Bhagwat AA, Bauer WD (1981) Transient susceptibility of root cells in four common legumes to nodulation by Rhizobia. Plant Physiol 68:1144–1149

Borisov AY, Madsen LH, Tsyganov VE, Umehara Y, Voroshilova VA, Batagov AO, Sandal N, Mortensen A, Schauser L, Ellis N, Tikhonovich IA, Stougaard J (2003) The *Sym35* gene required for root nodule development in pea is an ortholog of Nin from *Lotus japonicus*. Plant Physiol 131:1009–1017

Caldwell BE (1966) Inheritance of a stain-specific ineffective nodulation in soybean. Crop Sci 6:427–428

Carroll BJ, Mathew A (1990) Nitrate inhibition of nodulation in legumes. In: Gresshoff, PM (ed) Molecular biology of symbiotic nitrogen fixation. CRC Press, Boca Raton, FL, pp 159–180

Carroll BJ, McNeil DL, Gresshoff PM (1985a) A supernodulation and nitrate-tolerant symbiotic (*nts*) soybean mutant. Plant Physiol 78:34–40

Carroll BJ, McNeil DL, Gresshoff PM (1985b) Isolation and properties of soybean [Glycine max (L.) Merr] mutants that nodulate in the presence of high nitrate concentrations. Proc Natl Acad Sci USA, 82:4162–4166

Catoira R, Timmers AC, Maillet F, Galera C, Penmetsa RV, Cook D, Denarie J, Gough C (2001) The *HCL* gene of *Medicago truncatula* controls *Rhizobium*-induced root hair curling. Development 128:1507–1518

Clark FE (1957) Nodulation responses of two near-isogenic lines of the soybean. Can J Microbiol 3:113–123

Colebatch G, Kloska S, Trevaski B, Freund S, Altman T, Udvardi MK (2002) Novel aspects of symbiotic nitrogen fixation uncovered by transcript profiling with cDNA arrays. Mol Plant Microbe Interact 15:411–420

Colebatch G, Desbrosses G, Ott T, Krusell L, Montanari O, Kloska S, Kopka J, Udvardi MK (2004) Global changes in transcription orchestrate metabolic differentiation during symbiotic nitrogen fixation in *Lotus japonicus*. Plant J 39:487–512

Cren M, Kondorosi A, Kondorosi E (1995) *NolR* controls expression of the *Rhizobium meliloti* nodulation genes involved in the core Nod factor synthesis. Mol Microbiol 15:733–747

Davis TM, Foster KW, Phillips DA (1985) Nodulation mutants in chickpea. Crop Sci 25:345–348

Davis JHC, Giller KE, Kipe-Nolt J, ad Awah M (1988) Non-nodulating mutants in common bean. Crop Sci 28:859–860

Delves AC, Mathews A, Day A, Carter AS, Carroll BJ, Gresshoff, PM (1986) Regulation of the soybean-*Rhizobium* nodule symbiosis by shoot and root factors. Plant Physiol 82:588–590

Delves AC, Carroll BJ, Gresshoff, PM (1988) Genetic analysis and complementation studies in a number of mutant supernodulating soybean lines. J Genet 67:1–8

Dénarié J, Debelle F, Prome J-C (1996) Rhizobium lipo-chitooligosaccharide nodulation factors: signaling molecules mediating recognition and morphogenesis. Annu Rev Biochem 65:503–535

Devine TE, O'Neill JJ (1986) Registration of BARC-2 (*Rj4*) and Barc-3 (*rj4*) soybean germplasm. Crop Sci 26:1263–1264

Devine TE, O'Neill JJ (1993) Genetic independence of the nodulation-response gene loci-*Rj1*, *Rj2*, and *Rj4*-in soybean. J Hered 84:140–142

El Yahyaoui F, Kuster H, Ben Amor B, Hohnjec N, Puhler A, Becker A, Gouzy J, Vernie T, Gough C, Niebel A, Godiard L, Gamas P (2004) Expression profiling in *Medicago truncatula* identifies more than 750 genes differentially expressed during nodulation, including many potential regulators of the symbiotic program. Plant Physiol 136:3159–3176

Endre G, Kereszt A, Kevei Z, Mihacea S, Kalo P, Kiss GB (2002) A receptor kinase regulating symbiotic nodule development. Nature 417:962–966

Fedorova M, van de Mortel J, Matsumoto PA, Cho, J, Town, CD, VandenBosch, KA, Gantt, JS, Vance, CP (2002) Genome-wide identification of nodule-specific transcripts in the model legume *Medicago truncatula*. Plant Physiol 130:519–537

Fellay R, Hanin M, Montorzi G, Frey J, Freiberg C, Golinowski W, Staehelin C, Broughton WJ, Jabbouri, S (1998) *nodD2* of *Rhizobium* sp. NGR234 involved in the repression of the *nodABC* operon. Mol Microbiol 27:11039–11050

Geurts R, Fedorova E, Bisseling, T (2005) Nod factor signaling genes and their function in the early stages of *Rhizobium* infection. Curr Opin Plant Biol 8:346–352

Goormachtig S, Valerio-Lepiniex, M, Szczyglowski, K, Van Montagu M, Holsters M, de Bruijin FJ (1995) Use of differential display to identify novel *Sesbania rostrata* genes enhanced by *Azorhizobium caulinodans* infection. Mol Plant Microbe Interact 8:816–824

Gremaud MF, Harper JE (1989) Selection and initial characterization of partially nitrate tolerant nodulation mutants of soybean. Plant Physiol 89:169–173

Gresshoff PM (2003) Post-genomic insights into plant nodulation symbioses. Genome Biol 4:201–205

Gresshoff PM, Delves AC (1986) Plant genetic approaches to symbiotic nodulation and nitrogen fixation in legumes. Plant Gene Res 3:159–206

Gresshoff PM, Olsson JE, Day DA, Schuller KA, Mathews A, Delves AC, KrotzKy A, Price GD, Carroll BJ (1987) Plant host genetics of nodulation in soybean. In: Verma DPS, Brisson N (eds) Molecular genetics of plant-microbe interactions. M. Nijhoff Publisher, Dordrecht, The Netherlands, pp 885–890

Hohnjec N, Vieweg MF, Pühler A, Becker A, Küster H (2005) Overlaps in the transcriptional profiles of *Medicago truncatula* roots inoculated with two different *Glomus* fungi provide insights into the genetic program activated during arbuscular mycorrhiza. Plant Physiol 137:1283–1301

Israel DW, Mathis JN, Barbour WM, Elkan GH (1986) Symbiotic effectiveness and host-strain interactions of *Rhizobium fredii* USDA 191 on different soybean cultivars. Appl Environ Microbiol 51:898–903

Jacobsen E, Feenstra WJ (1984) A new pea mutant with efficient nodulation in the presence of nitrate. Plant Sci Lett 33:337–344

Jorrín JV, Rubiales D, Dumas-Gaudot E, Recorbet G, Maldonado A, Castillejo MA, Curto M (2006) Proteomics: a promising approach to study biotic interaction in legumes. A review. Euphytica 147: 37–47

Kim MY, Ha B-G, Jun T-H, Hwang E-Y, Van K, Kuk YI, Lee S-H (2004) Single nucleotide polymorphism discovery and linkage mapping of lipoxgenase-2 gene (*Lx2*) in soybean. Euphytica 135: 169–177

Kim MY, Van K, Lestari P, Moon J-K, Lee S-H (2005) SNP identification and SNAP marker development for a *GmNARK* gene controlling supernodulation in soybean. Theor Appl Genet 110: 1003–1010

Kistner C, Winzer T, Pitzschke A, Mulder L, Sato S, Kaneko T, Tabata S, Sandal N, Stougaard J, Webb KJ, Szczyglowski K, Parniske M (2005) Seven *Lotus japonicus* genes required for transcriptional reprogramming of the root during fungal and bacterial symbiosis. Plant Cell 17: 2217–2229

Kneen BE, LaRue TA (1984) Nodulation resistant mutant of *Pisum sativum* (L.). J Hered 75:238–240

Kneen BE, LaRue TA (1988) Induced symbiosis mutant pea (*Pisum sativum*) and sweetclover (*Mililotus alba annua*). Plant Sci 58:177–182

Knight CD, Rossen L, Robertson JG, Wells B, Downie JA (1986) Nodulation inhibition of *Rhizobium leguminosarum* multicopy *nodABC* genes and analysis of early stages of plant infection. J Bacteriol 166:552–558

Kolchinsky A, Landau-Ellis D, Gresshoff PM (1997) Map order and linkage distance of molecular markers close to the supernodulation (*nts-1*) locus of soybean. Mol Genet Genomics 254:29–36

Kondorosi E, Gyuris J, Schmidt J, John M, Duda E, Hoffeman B, Schell J, Kondorosi A (1989) Positive and negative regulation of *nod* gene expression in *Rhizobium meliloti* is required for optimal nodulation. EMBO J 8:1331–1340

Kosslak RM, Booklanf R, Barkei J, Paaren H, Appelbaum ER (1987) Induction of *B. japonicum* common nod genes by flavones isolated from *Glycine max*. Proc Natl Acad Sci USA 84:7428–7432

Kouchi H, Shimomura K, Hata S, Hirota A, Wu GJ, Kumagai H, Tajima S, Suganuma N, Suzuki A, Aoki T, Hayashi M, Yokoyama T, Ohyama T, Asamizu E, Kuwata C, Shibata D, Tabata S (2004) Large-scale analysis of gene expression profiles during early stages of root nodule formation in a model legume, *Lotus japonicus*. DNA Res 11:264–274

Krusell L, Madsen LH, Sato S, Aubet G, Genua A, Szczyglowski K, Duc G, Kaneko T, Tabata S, de Bruijin F, Pajuelo E, Sandal N, Stougaard J (2002) Shoot control of root development and nodulation in mediated by a receptor-like kinase. Nature 420:422–426

Küster H, Hohnjec N, Krajinski F, El Yahyaoui F, Manthey K, Gouzy J, Dondrup M, Meyer F, Kalinowski J, Brechenmacher L, van Tuinen D, Gianinazzi-Pearson V, Pühler A, Gamas P, Becker A (2004) Construction and validation of cDNA-based Mt6k-RIT macro- and micro-arrays to explore root endosymbioses in the model legume *Medicago trauncatula*. J Biotechnol 108:95–113

Küster H, Vieweg MF, Manthey K, Baier MC, Hohnjec N, Perlick AM (2007) Identification and expression regulation of symbiotically activated legume genes. Phytochemistry 68:8–18

Landau-Ellis D, Angemuller S, Shoemaker RC, Gresshoff PM (1991) The genetic locus controlling supernodulation in soybean (*Glycine max* L.) co-segregates tightly with a cloned molecular marker. Mol Genet Genomics 228:221–226

Lee KH, LaRue TA (1992) Exogenous ethylene inhibits nodulation of *Pisum sativum* L. cv. Sparkle. Plant Physiol 100:1759–1763

Lee HS, Lee S-H (1998) Introduction, development, and characterization of supernodulating soybean mutant – nitrate inhibition of nodulation and nitrogen fixation in supernodulating soybean mutant. Korean J Crop Sci 43:23–27

Lee SH, Ashley DA, Boerma HR (1991) Regulation of nodule development in supernodulating mutants and wild-type soybean. Crop Sci 31:688–693

Lee HS, Chae YA, Park EH, Kim YW, Yun KI, Lee SH (1997) Introduction, development, and characterization of supernodulating soybean mutant – mutagenesis of soybean and selection of supernodulating soybean mutant. Korean J Crop Sci 42:247–253

Lee H, Hur C-G, Oh CJ, Kim HB, Park S-Y, An CS (2004) Analysis of the root nodule-enhanced transcriptome in soybean. Mol Cells 18:53–62

Lestari P, Van K, Kim MY, Lee S-H (2005) Symbiotic effectiveness of *Bradyrhizobium japonicum* USDA 110 in a supernodulating soybean mutant SS2-2. Korean J Crop Sci 50:125–130

Lestari P, Van K, Kim MY, Hwang CH, Lee S-H (2006a) Differentially expressed genes related to symbiotic association in a supernodulating soybean mutant and its wild type by cDNA-AFLP. Mol Plant Pathol 7:235–247

Lestari P, Van K, Kim MY, Lee B-W, Lee S-H (2006b) Newly featured infection events in a supernodulating soybean mutant SS2-2 by *Bradyrhizobium japonicum*. Can J Microbiol 52:328–335

Lie TA (1974) Environmental effects on nodulation and symbiotic nitrogen fixation. In: Quispel A (ed) The biology of nitrogen fixation. North-Holland Publishing Company, Amsterdam, pp 555–582

Lievens S, Goormachtig S, Holsters M (2001) A critical evaluation of differential display as a tool to identify genes involved in legume nodulation: looking back and looking forward. Nucleic Acids Res 29:3459–3468

Loh J, Stacey G (2001) Feedback regulation of the *Bradyrhizobium japonicum* nodulation genes. Mol Microbiol 41:1357–1364

Loh J, Stacey G (2003) Nodulation gene regulation in *Bradyrhizobium japonicum*: a unique integration of global regulatory circuits. Appl Environ Microbiol 69:10–17

Lohar DP, Sharopova N, Endre G, Penuela S, Samac D, Town C, Silverstein KA, VandenBosch KA (2006) Transcript analysis of early nodulation events in *Medicago truncatula*. Plant Physiol 140: 221–234

Lohnes DG, Wagner RE, Bernard RL (1993) Soybean genes *Rj2*, *Rmd*, and *Rps2* in linkage group 19. J Hered 84:109–111

Long SR (1989) Rhizobium-legume nodulation: life together in the underground. Cell 56:203–214

Long SR (1996) *Rhizobium* symbiosis: nod factors in perspective. Plant Cell 8:1885–1898

Long SR (2001) Genes and signals in the *rhizobium*-legume symbiosis. Plant Physiol 125:69–72

Maguire TL, Grimmond S, Forrest A, Iturbe-Ormaetxe I, Meksem K, Gresshoff PM (2002) Tissue-specific gene expression monitored by cDNA microarray analysis of soybean (*Glycine max*). J Plant Physiol 159:1361–1374

Manthey K, Krajinski F, Hohnjec N, Firnhaber C, Pühler A, Perlick AM, Küster H (2004) Transcriptome profiling in root nodules and arbuscular mycorrhiza identifies a collection of novel gene induced during *Medicago truncatula* root endosymbioses. Mol Plant Microbe Interact 17:1063–1077

Mathesius U, Keijzers G, Natera SH, Weinman JJ, Djordjevic MA, Rolfe BG (2001) Establishment of a root proteome reference map for the model legume *Medicago truncatula* using the expressed sequence tag database for peptide mass fingerprinting. Proteomics 1:1424–1440

Mathesius U, Imin N, Chen H, Djordjevic MA, Weinman JJ, Natera SH, Morris AC, Kerim T, Paul S, Menzel C, Weiller GF, Rolfe BG (2002) Evaluation of proteome reference maps for cross-species identification of proteins by peptide mass fingerprinting. Proteomics 2:1288–1303

Mathesius U, Mulders S, Gao MS, Teplitski M, Caetano-Anolles G, Rolfe BG, Bauer WD (2003) Extensive and specific responses of a eukaryotic to bacterial quorum-sensing signals. Proc Natl Acad Sci USA 100:1444–1449

Mathews A, Carroll BJ, Gresshoff PM (1987) Characterization of non-nodulation mutants of soybean (*Glycine max* [L.] Merril): *Bradyrhizobium* effects and absence of root hair curling. J Plant Physiol 131:349–361

Matthews BF, Devine TE, Weisemann JM, Beard HS, Lewers KS, MacDonald MH, Park Y-B, Maiti R, Lin J-J, Kuo J, Pedroni MJ, Cregan PB, Saunders JA (2001) Incorporation of sequenced cDNA and genomic markers into the soybean genetic map. Crop Sci 41:516–521

Men AE, Laniya TS, Iturbe-Dramaetxe I, Gresshoff I, Jiang Q, Carroll BJ, Gresshoff PM (2002) Fast neutron mutagenesis of soybean (*Glycine soja* L.) produces a supernodulating mutant containing a large deletion in Linkage group H. Genome Lett 3:147–155

Mitra RM, Long SR (2004) Plant and bacterial symbiotic mutants define three transcriptionally distinct stages in the development of the *Medicago truncatula*/*Sinorhizobium meliloti* symbiosis. Plant Physiol 134:595–604

Mitra RM, Shaw SL, Long SR (2004) Six nonnodulating plant mutants defective for Nod-factor-induced transcriptional changes associated with the legume-Rhizobia symbiosis. Proc Natl Acad Sci USA 101:10217–10222

Morris AC, Djordjevic MA (2001) Proteome analysis of cultivar-specific interaction between *Rhizobium leguminosarum* biovar *trifolii* and subterranean clover cultivar Woogenellup. Electrophoresis 22: 586–598

Mylona P, Pawlowski K, Bisseling T (1995) Symbiotic nitrogen fixation. Plant Cell 7:869–885

Natera SHA, Guerreiro N, Djordjevic MA (2000) Proteome analysis of differentially displayed protein as a tool for the investigation of symbiosis. Mol Plant Microbe Interact 13:995–1009

Nishimura R, Hayashi M, Wu, GJ, Kouchi H, Imaizumi-Anraku, H, Murakami Y, Kawasaki S, Akao S, Ohmori M, Nagasawa M, Harada K, Kawaguchi M (2002) HAR1 mediates systemic regulation of symbiotic organ development. Nature 420:426–429

Nukui N, Ezura H, Yuhashi K, Yasuta T, Minamisawa G (2000) Effects of ethylene precursor and inhibitors for ethylene biosynthesis and perception on nodulation in *Lotus japonicus* and *Macroptilium atropurpureum*. Plant Cell Physiol 41:893–897

Oldroyd GED (2001) Dissecting symbiosis: development in Nod factor signal transduction. Ann Bot 87:709–718

Olsson JE, Nakao P, Bohlool B, Gresshoff PM (1989) Lack of systemic suppression of nodulation in split root systems of supernodulating soybean mutants. Plant Physiol 73:286–290

Park SJ, Buttery BR (1988) Nodulation mutants of white bean (*Phaseolus vulgaris* L.) induced by ethyl-methane sulphonate. Can J Plant Sci 68:199–202

Patner S, Thompson R, de Bruxelles G, Laver D, Trevaskis B, Udvardi M (2000) Identification with proteomics of novel proteins associated with the peribacteriod membrane of soybean root nodules. Mol Plant Microbe Interact 13:325–333

Pierce M, Bauer WD (1983) A rapid regulatory response governing nodulation in soybean. Plant Physiol 73:286–290

Pracht JE, Nickell CD, Harper JE (1993) Genetic analysis of a hypernodulating mutant of soybean. Soybean Genet Newsl 20:107–111

Riely BK, Ané J-M, Penmetsa RV, Cook DR (2004) Genetic and genomic analysis in model legumes bring nod-factor signaling to center stage. Curr Opin Plant Biol 7:408–413

Rolfe BG, Mathesius U, Djordjevic M, Weinman J, Hocart C, Weiller G, Bauer WD (2003) Proteomic analysis of legume–microbe interactions. Comp Funct Genomics 4:225–228

Saalbach G, Erik P, Wienkoop S (2002) Characterization by proteomics of peribacteriod space and peribacteriod membrane preparations from pea (*Pisum sativum*) symbiosomes. Proteomics 2:325–337

Sato S, Tabata S (2006) *Lotus japonicus* as a platform for legume research. Curr Opin Plant Biol 9:128–132

Schauser L, Roussis A, Stiller J, Stougarrd J (1999) A plant regulator controlling development of symbiotic root nodules. Nature 402:191–195

Schmidt JS, Harper JE, Hoffman TK, Bent AF (1999) Regulation of soybean nodulation independent of ethylene signaling. Plant Physiol 119:951–959

Schnabel E, Journet EP, de Carvalho-Niebel F, Duc G, Frugoli J (2005) The *Medicago truncatula SUNN* gene encodes a CLV1-like leucine-rich repeat receptor kinase that regulates nodule number and root length. Plant Mol Biol 58:809–822

Schultze M, Kondorosi A (1998) Regulation of symbiotic root nodule development. Ann Rev Genet 32:33–57

Searle IR, Men AM, Laniya TS, Buzas DM, Iturbe-Ormaetxe I, Caroll BJ, Gresshoff PM (2003) Long-distance signaling for nodulation control in legumes requires a CLAVATA1-like receptor kinase. Science 299:109–112

Shoemaker RC, Schlueter J, Doyle JJ (2006) Paleopolyploidy and gene duplication in soybean and other legumes. Curr Opin Plant Biol 9:104–109

Simoes-Araujo JL, Rodrigues RL, de A Gerhardt LB, Mondego JM, Alves-Ferreira M, Rumjanek NG, Margis-Pinheiro M (2002) Identification of differentially expressed gene by cDNA-AFLP technique during heat stress in cowpea nodules. FEBS Lett 515:44–50

Stacey G, Libault M, Brenchenmacher L, Wan J, May GD (2006) Genetics and functional genomics of legume nodulation. Curr Opin Plant Biol 9:110–121

Starker CG, Parra-Colmenares AL, Smith L, Mitra RM, Long SR (2006) Nitrogen fixation mutants of *Medicago truncatula* fail to support plant and bacterial symbiotic gene expression. Plant Physiol 140:671–680

Stracke S, Kistner C, Yoshida S, Mulder I, Sato S, Kaneko T, Tabata S, Sandal N, Stougaard J, Szczyglowski K, Parniske M (2002) A plant receptor-like kinase requires for both bacterial and fungal symbiosis. Nature 417:959–962

Sutton SD (1983) Nodule development and senescence. In: Broughton WJ (ed) Nitrogen fixation, Vol. 3. Legumes, Clarendon Press, Oxford, pp 144–209

Tanner JW, Anderson IC (1963) Investigations on non-nodulating and nodulating soybean strains. Can J Plant Sci 43:542–545

Tesfaye M, Samac DA, Vance CP (2006) Insights into symbiotic nitrogen fixation in *Medicago truncatula*. Mol Plant Microbe Interact 19:330–341

Town CD (2006) Annotating the genome of *Medicago truncatula*. Curr Opin Plant Biol 9:122–127

Trevaskis B, Colebatch G, Desbrosses G, Wandrey M, Wienkoop S, Saalbach G, Udvardi M (2002) Differentiation of plant cells during symbiotic nitrogen fixation. Compar Funct Genomics 3:151–157

Van K, Kim K-S, Ha B-K, Jun T-H, Jang H-J, Kim MY, Lee S-H (2005) Molecular marker characterization of a supernodulating soybean mutant, SS2-2. Korean J Breeding 37:35–42

Vance CP, Egli MN, Griffith SM, Miller SS (1988) Plant regulated aspects of nodulation and N fixation. Plant, Cell Environ 11:413–427

Vest G (1970) Rj – a gene conditioning ineffective nodulation in soybean. Crop Sci 10:34–35

Vest G, Caldwell BE (1972) *Rj4*-a gene conditioning ineffective nodulation in soybean. Crop Sci 12:692–693

Vuong TD, Harper JE (2000) Inheritance and allelism analysis of hypernodulating genes in the NOD3-7 and NOD2-4 soybean mutants. Crop Sci 40:700–703

Wan J, Torres M, Ganapathy A, Thelen J, DaGue BB, Mooney B, Xu D, Stacey G (2005) Proteomic analysis of soybean root hairs after infection by *Brayrhizobium japonicum*. Mol Plant Microbe Interact 18:458–467

William LF, Lynch DL (1954) Inheritance of a non-nodulating character in the soybean. Agron J 46:28–29

Winzer T, Bairl A, Linder M, Linder D, Werner D, Müller P (1999) A novel 53–kDa nodulin of the symbiosome membrane of soybean nodules, controlled by *Brayrhizobium japonicum*. Mol Plant Microbe Interact 12:218–228

CHAPTER 17

GENOMICS OF WHEAT DOMESTICATION

CARLO POZZI[1] AND FRANCESCO SALAMINI[1,2,*]

[1]*Fondazione Parco Tecnologico Padano, via Einstein Loc. C.na Codazza, 26900 Lodi, Italy*
[2]*Università degli Studi di Milano, Dip. Produzione Vegetale, via Celoria 2, 20133 Milano, Italy*

Abstract: The review covers several issues concerning the state of molecular knowledge of the effects induced by domestication and breeding on the wheat crop. Genes at the root of the domestication syndrome are currently the focus of an active research which frequently uses comparative genomics approaches. Conclusions drawn on available data indicate that the domestication syndrome is originated by "sudden" genetic events, controlled by few major pleiotropic genes. These events were followed by the accumulation of a larger set of minor mutations, having a multifactorial mode of inheritance. Moreover the organization of nucleotide variability enables the detection of domestication-related molecular footprints, suggesting that the genomic regions more responsible for genetic variation and more related to domestication are reduced when compared to the whole genome size. The polyploidy history of the domesticated wheats is presented, making a specific mention to the origin of the wheat A, B, D and G genomes and to the molecular control of chromosome pairing in polyploids. A general presentation is also provided on the genomic changes which have accompanied the emergence of domesticated wheats. What follows is a molecular information on: i) the wheat adaptation to the environment (genomics of photoperiod, vernalization, heading date, plant height, and erect plant type); ii) the effect of domestication on seed-related yield components (genomics of seed size, grain hardness, and tillering); iii) modification of traits affecting harvestability (emergence of free-threshing seeds, rachis toughness, and presence of ear awns). Genetic bottlenecks which have been associated to wheat domestication and breeding are considered in a final section. The relatively young history of the wheat crop, the presumably small founder population of this gene pool, and the intensive long-term selection for agronomic traits did set the basis for a reduced genetic variability of the genus.

1. INTRODUCTION

The most important grass subfamilies, including *Pooideae*, radiated 50–80 million years ago (MYA). The *Triticeae* and *Poeae* (*Lolium rigidum*) tribes separated ca. 35 MYA, *Hordeum* diverged from the *Triticum/Aegilops* lineage ca. 11 MYA,

*Author for correspondence: francesco.salamini@tecnoparco.org

R.K. Varshney and R. Tuberosa (eds.), Genomics Assisted Crop Improvement: Vol. 2: Genomics Applications in Crops, 453–481.

and *Secale* from the same lineage 7 MYA. Several members of the grass family (*Poaceae*) – wheat (genus *Triticum*), rice (genus *Oryza*) and maize (genus *Zea*) – were domesticated in a time frame comprised between 8,000 and 12,000 years from present (BP). Agricultural systems based on these crops made possible, at the time, to support the civilization of South-West Asia, China, and South America (Diamond, 1997; summarized in Pozzi et al., 2002). Events leading to an increase of wheat ploidy to 4n should be moved back to some 300,000 years (Akhunov et al., 2003), and to approximately 8,000 years ago the synthesis of the 6n genome of bread wheat (Salamini et al., 2002).

Specific alleles of groups of tightly linked loci (genes known as domestication syndrome factors, DSFs), have been favoured during domestication (Le Thierry d'Ennequin et al., 1999; Lin et al., 1995). The noted strong linkage among DSFs is thought to facilitate the maintenance of the phenotypic identity of domesticated populations coexisting in the same area with wild lineages, a case observed for rice (Cai and Morishima, 2002). Alternatively, the clustering in specific linkage groups of domestication-related genes with relevance to plant adaptation has been explained by the effect of balancing selection in large natural populations: in this perspective, clustering of genes predisposed wild crops to be more rapidly domesticated, while the subsequent domestication process fixed useful mutations within the cluster (Lin et al., 1995). Fifty-six QTLs for 11 domestication-related traits underlying the difference between wild *T. dicoccoides* and *T. durum* are clustered in seven linkage intervals, mostly located on only 4 chromosomes of the tetraploid wheat (Peng et al., 2003). DSFs have been probably selected as a set of genetic factors common to different crops (Buckler et al., 2001), thus justifying the believe that domestication of the grasses was genetically convergent, as reported for maize, rice and sorghum (Paterson et al., 1995). This implies that homologous, if not orthologous, genes/alleles have been involved in the domestication of these particular plants.

When comparative genomics has been used to investigate the relative location of DSFs in sorghum, rice, and maize, three QTLs affecting seed size were found to occupy syntenic positions in the three species; for grain shattering, QTLs map to a single locus in sorghum, three loci in rice and 10 loci in maize, and the sorghum single locus corresponds to a single rice QTL and to two maize QTLs located on duplicated chromosomal regions (Doebley and Stec, 1993). Experiments were also carried out to map domestication QTLs in wheat: Peng et al. (2003) found 7 major wheat DSFs, and mapped them in gene-rich regions. They concluded that "*the high pleiotropy and/or tight linkage of most wheat domestication QTLs suggest an important role of recombination in either consolidation of positive mutations within the clusters and/or in reducing the antagonism between artificial and background selection*." Several studies are currently addressing the molecular basis of domestication traits, and, interestingly, the picture emerging is similar across species. For example, plant height and flowering time are affected by mutations in the orthologous genes *RhtD1a* of wheat, *SLR1* of rice, and *D8* of maize (Ikeda et al., 2001).

Recent experiments approaching the possibility to map and clone QTLs (Doebley et al., 1997; Yano et al., 2000; reviewed in Salvi and Tuberosa, 2005) allow to draw some conclusions concerning the process of plant domestication: (i) the domestication syndrome originated by "sudden" genetic events, controlled by few major pleiotropic genes (Li et al. 1995; Lin et al., 1995). These events were followed by the accumulation of a larger set of minor mutations (Barton and Keightley, 2002); (ii) a multifactorial mode of inheritance characterizes domestication traits. It is likely that genetic changes in these traits are required to overcome a certain threshold if the developmental program of the plant has to shift towards the domesticated phenotype. This hypothesis relies on the observation that in natural populations genetic variation may exist without phenotypic variation (cryptic genetic variation). A consideration of the level and pattern of linkage disequilibrium and of the organization of nucleotide variability (reviewed in Morgante and Salamini, 2003) enables the detection of domestication-related molecular footprints, suggesting that the genomic regions more responsible for genetic variation and more related to domestication are reduced when compared to the whole genome size.

That selection acting on specific regions of the genome may be sufficient to explain the domestication syndrome is, to some extent, evident. In fact about 2% of maize genes investigated reveal a history of selection (Wright et al., 2005). This can be determined by Tajima's D statistics (Wright and Gaut, 2005), which detects deviations from neutrality by examining, in a set of genotypes, the average nucleotide differences and the number of segregating sites (Tajiima, 1989). An alternative explanation for the genes found to be under selection may reside in the population history of the species under consideration. For example, the occurrence of bottlenecks, and how these have affected nucleotide variation, may lead to the same results induced by positive selection (selective sweeps; Tajiima, 1989). Bottlenecks, as a matter of facts, contribute to decrease molecular polymorphism.

In conclusion, it is currently not easy to infer whether loci, existing today in a molecularly modified form, were fixed due to population history or were selected during the domestication process or during the subsequent breeding phase (Wright et al., 2005).

2. IMPORTANCE OF PLOIDY DURING DOMESTICATION

Polyploidy originates mainly from unreduced gametes which have twice the number of chromosomes expected from a species (Leitch and Bennet, 1997). When two unreduced gametes of the same species join in a zygote, the resulting progeny is autopolypolyfloid Unreduced gametes from different but interfertile species can also result in a zygote (allopolyploidy) with a double number of chromosomes compared to the average number present in the progenitor species. Allopolyploidization may involve intergenomic recombination (Soltis, 2005) and rapid loss of DNA (Ozkan et al., 2001), whereby subsequent diploidization restores disomic genetics (Levy and Feldman, 2002).

Often polyploid species are sterile owing to the formation of abnormal multivalents during meiosis. Allotetraploid and allohexaploid should have two or three different copies of each nuclear gene and thus polyploids can be viewed as a combination of genomes with independent past histories. Recent molecular genetic studies have made clear that even the simple genome of Arabidopsis is the result of nonrandom patterns of arrangement of duplicated genes, probably due to ancient duplication or polyplodization events (Kowalski et al., 1994; Vision et al., 2000; Ozkan et al., 2001; Li et al., 2004). In Arabidopsis, however, the diploidization process is advanced, with only approximately 50% of genes having a discernible paralog and 19% locating in a region consistent with a chromosomal or segmental duplication event (Paterson et al., 2000). The question is whether polyploid formation has a selective advantage. QTL mapping efforts suggest that there may exist a relationship between polyploidy and crop productivity (Freeling and Thomas, 2006), that is, that non linear changes in phenotype are made possible as a direct result of polyploidy.

Specific polyploids arose recurrently during flowering plant evolution (Soltis, 2005), accompanied by rapid genome restructuring (Leicht and Bennet, 1997).

2.1. Ploidy and Wheats Evolution

Ploidy changes are at the core of cereal evolution. These species can have 14 (diploid), 28 (tetraploid) or 42 (hexaploid) chromosomes. The diploid genomes that contributed to the formation of the modern hexaploids are several: genomes A (A^m and A^urespectively present today in *T. monococcum* and *T. urartu*), D, and S (S, S^s, S^b, S^l, S^{sh}). Two genomes found in polyploids wheats were given new names, B and G, because their diploid progenitors were at the time not known (Kilian et al., 2006). The genus *Triticum* includes cultivated species cytogenetically associated in 4 groups: einkorn (genome A^mA^m), emmer (genome A^uA^uBB), *T. timopheevii* (A^uA^uGG) and bread wheat (A^uA^uBBDD).

2.2. Origin of the Wheat Genomes

Monococcum (einkorn, genome A^mA^m) was probably the first wheat to be cultivated. The wheat was successively substituted, starting in the bronze age, predominantly by tetraploids and hexaploids with naked seeds. Wild emmer derives from a hybridization event dating back of 300,000 years (Akhunov et al., 2003). A diploid donor with the A genome (*T. urartu*, genome A^uA^u) hybridized with a second species, a representative of the *Aegilops* section *Sitopsis* (close to *Ae. speltoides*; Kilian et al., 2006), the donor of the B genome (Huang et al., 2002). The newly synthesized species (*T. turgidum* ssp. *dicoccoides* , genome A^uA^uBB, wild emmer) has given rise to the current tetraploid varieties of the domesticated 4n wheat *T. turgidum* ssp. *durum*, (free-threshing), and *T. turgidum* ssp. *dicoccum* (hulled). In comparison, the hexaploid wheat (*T. aestivum*, genome A^uA^uBBDD) is a young polyploid: the taxon exists only since 8,000 years (Nesbitt and Samuel, 1996).

Its origin is from an additional polyploidization event between an early domestication tetraploid *T. turgidum* and the diploid donor of the D genome, *Ae. tauschii* (genome DD).

2.2.1. The A genome

Several studies point to a high level of similarity between the chromosomes of *T. monococcum* (A^m) and chromosomes A^u of *T. aestivum* (Luo et al., 2000; Provan et al., 2004). This suggests that A^u and A^m genomes have not diverged significantly since their separation from a common progenitor. A quantification of their difference is provided by Brandolini et al. (2006), based on AFLP markers describing putative chromosome A-gene fragments (i.e. AFLP bands not deriving from the B genome, and not specific for the hexaploids, as it would be if they were derived from the D genome). The authors used 41 accessions of A-genome, 3 of AG-genome wheats (*T. timopheevii*), 19 of AB-genome wheats, 15 of ABD-genome wheats, and 1 of the D-genome donor *Ae. tauschii*. The data indicate that the A^u genome of *T. urartu* is more similar by around 20% to the A genome of tetraploid and hexaplod wheats, compared with the A^m genome of *T. monococcum*. The difference is smaller when the comparison of the A^u genome with the A genome of *T. timopheevi* is considered. The study supports the current belief that only a part of the variation existing in the wheat diploid gene pool is present in the polyploid wheats (Dvorak et al,. 2006).

The domestication of the A genome has been followed by studying einkorn wheat which was domesticated from its wild progenitor, *T. boeoticum* (also genome A^mA^m). The two taxa differ mainly in plant and seed size and in seed dormancy. The site of einkorn domestication has been identified from the analysis of 288 amplified fragment length polymorphisms (AFLP) marker loci. *T. boeoticum* populations from the western foothills of the Karakadag mountains of southweast Turkey are more similar to einkorn than are other wild populations, which points to this area as the centre of its domestication (Heun et al., 1997).

2.2.2. The B genome

Wild emmer (genome AABB) has brittle ears, a character lost during the domestication that led to *Trititum dicoccum*. A study based on AFLP at 204 loci (Ozkan et al., 2002) demonstrated that domesticated tetraploid wheats are closely related to wild populations from south-eastern Turkey. In the Fertile Crescent at least two major races of wild emmer exist, and it was the one present in the south-eastern Turkish areas which has, apparently, been taken into cultivation. The same study proved that hulled emmer may have an origin independent from free-threshing 4n wheats. Mori et al. (2003) investigated chloroplast microsatellite variation in populations, from the Kartal Dagi Mountains and concluded that emmer was domesticated twice, once in the Kartal Dagi region and a second time in an unidentified site. Ozkan et al. (2005) excluded the Kartal Dagi as site and suggested that wild emmer from the Karaca Dag region and the Sulaimanyia region (Iran-Iraq border) were at the root of domesticated emmer. Luo et al. (2006; unpublished. results)

confirmed that wild emmer consists of two populations, with domesticated emmer deriving from the southern population. Moreover, they provide evidence for an extensive gene flow between wild and domesticated emmer, and conclude that either emmer was domesticated in south-eastern Turkey or that it was domesticated independently twice, and later on the Levantine gene pool was absorbed into the south-eastern domesticated gene pool. An early appearance of domesticated hulled and free-threshing emmer is consistent with early archaeological evidences concerning domesticated tetraploid free-threshing wheats (Salamini et al., 2002).

The origin of the B genome is still not definitively clear (Huang et al., 2002). The genome of *Ae. speltoides*, a probable progenitor, is designated S from the name of the *Aegilops* section *Sitopsis* (Huang et al., 2002; Slageren, 1994). A second point concerning the B genome origin is that this is highly related to the origin of the G genome of *T. timopheevi* (AAGG). The AABB and AAGG genome origins have been attributed to the same single hybridization event (Provan et al., 2004) or to separate alloploidization events (Rodriguez et al., 2000). Huang et al. (2002) suggest that some loci of the S and of the G genome are closely related, leading to the conclusion that parts of their genomes may have the same origin in an *Ae. speltoides* progenitor. This evidence agrees with results of other studies implicating *Ae. speltoides* as the donor of the G genome. On the other hand, none of the loci of the *Aegilops* genome analyzed in the Huang's study (2002) closely resemble the hortologous locus of the B genome. Kilian et al. (2006) extensively evaluated if *A. speltoides* was the donor of the B and G genomes in tetraploids wheats. Five *Aegilops* species of the *Sitopsis* section were compared by AFLP fingerprinting to diploid, tetraploid, and nullitetrasomic *T. aestivum* wheat lines: the B genome could be attributed to S chromosomes of *Ae. speltoides*. The same study haplotyped eight genes in *Aegilops* and *Triticum* lines: B and G genomes share again the highest sequence similarity with *Ae. speltoides,* but they sequestered different samples of *Ae. speltoides* haplotype diversity. This evidence would prove that the hybridization events that led to AABB and AAGG genomes occurred independently.

In a recent study, Petersen et al. (2006) intended to overcome the limitation of most of the published wheat phylogenetic analyses, which are based on a small number of taxa, or are biased by the choice of polyploid wheat accessions having an *a priori* assumed role as progenitors of domesticated forms (Buchner et al., 2004), or consider few *Aegilops* and/or *Triticum* species (Zhang et al., 2002; Sallares and Brown, 2004). The authors studied two single-copy nuclear genes, *DMC1*, and *EF-G*, together with a plastid gene (*NADH* subunit F) from each of the A, B, and D genomes of hexaploid wheat and from the A and B genomes of the tetraploid progenitor of *T. turgidum*. The genes were amplified from accessions of diploid *Aegilops/Triticum* species, of tetraploid *T. turgidum* and from one accession of *T. aestivum*. According to these results, the D genome derives from *A. tauschii*, the A genome from *T. urartu* and the B genome from *Ae. speltoides*.

2.2.3. Hexaploid wheats

A wild hexaploid (AABBDD) genome lineage does not exist. In this sense, hexaploid wheat is the result of a hybridization between a domesticated from of emmer (*T. turgidum* most probably spp. *dicoccum*) and the diploid *Ae. tauschii* (genome DD; Kihara, 1944; McFadden and Sears, 1946; Cox, 1998; Feldman, 2001; Salamini et al., 2002). *Ae. taushii* consists of the gene pools S and T, loosely related with the two subspecies *strangulata* and *tauschii* (Dvorak et al., 1998). RFLP studies indicated that the D genome is more closely related to the S than to the T gene pool. However, the finding of parallel polymorphisms in the wheat D genome and in the *Ae. tauschii* gene pools and the existence of wheat haplotypes closely related to the T gene pool negate the idea that 6n wheat originated via a single hyrbidization event (Caldwell et al, 2004; Talbert et al., 1998). Thus, the Chinese Spring D genome appears to be a conglomerate of contributions from both the T and S *Ae. tauschii* gene pools (Peng et al., 2004; Akhunov et al., 2003; Akhunov et al., 2006).

The seven (1A–7A) A chromosomes of hexaploid wheats are related, through common ancestors, to homeologous chromosomes B and D; that is, chromosomes 1A, 1B and 1D are syntenic as are 2A, 2B and 2D and so on. This is because the wheat A, B, and D genomes diverged between 2.5 and 4.5 MYA from their common ancestor (Huang et al., 2002). Hulled, brittle-rachis forms of *T. aestivum* have been frequently reported to be ancestral to the free-threshing forms. The emergence of free-threshing hexaploids is however still not completely clarified. Hulled forms of *T. aestivum* may derive from hybridization between free-threshing hexaploids and hulled emmer wheat (Dvorak et al., 2006). The molecular relationships of the putative progenitors - at 4n level several forms of *T. turgidum*, and the 6n hulled form of *T. spelta* - and the cultivated tetraploids remain thus elusive (Dvorak and Akhunov, 2005).

2.3. Molecular Control of Chromosome Pairing

The control of chromosome pairing provides insights on wheat evolution. *Ae. speltoides* forms are known that surppress pairing among homologous chromosomes (*Ph1* activity). If *Ph1* genotypes participate in allopolyploidization, the hybrids acquire a bivalent chromosome pairing, the case of *T. dicoccoides* (Sears, 1976). This was most probably a very important step in wheat polyploidization because *T. dicoccoides* is at the root of all polyploid wheats (with the exception of the wild AAGG genome; Kilian et al., 2006). Meiotic chromosome pairing in hexaploid wheats is in fact confined to strict homologous chromosomes, despite the presence in the genome of homoeologous chromosomes. This strict control contributes to the integrity of the A, B, and D genomes, and the cytological outcome is a diploid-like behaviour of the species leading to 21 bivalents at meiotic metaphase I. The control of chromosome pairing depends on a group of suppressing and promoting pairing homoeologous (*Ph*) genes (Sears, 1976), with *Ph1* located on the long arm of chromosome 5B. *Ph2* maps to the short arm of chromosome

3D (Mello-Sampayo, 1971; Sutton et al., 2003), and a further locus on the short arm of chromosome 3A (Driscoll, 1972). In the absence of *Ph1*, homoeologous recombination takes place among A, B, and D chromosomes, included also those of related species or genera (Koebner and Shepherd, 1985).

The *Ph1* locus has recently been cloned in *T. aestivum* (Griffiths et al., 2006) and found to have a complex structure consisting of a segment of subtelomeric heterochromatin inserted in a cluster of *cell division control 2* (*cdc2*) genes. Rearrangements of the region which seems responsible for pairing control may have arose on poplyploidization or pre-existed in the diploid genome donors.

The molecular bases for chromosomal recognition and pairing during early meiosis in hexaploid wheat are still unknown. An hypothesis states that *Ph1* affects basic components of chromosome structure. Mutants at *Ph1* show an altered chromosome/chromatin organization and compaction, in meiotic and in somatic cells, but also premature separation of sister chromatids (Aragon-Alcaide et al., 1997; Vega and Feldman, 1998). In addition, *Ph1* mutant have alterations in the relative arrangement of homologous chromosomes in both meiotic and somatic cell (Mikahilova et al., 1998). Some lines of *Ae. speltoides* do not show *Ph1*-like activity and allow chromosome pairing, leading to tetravalent formation and, possibly, intergenomic translocations. The AAGG genome *T. araraticum* has extensive DNA loss and chromosomal rearrangements relative to *T. dicoccoides* (Ozkan et al., 2001) which are, in part, the result of intergenomic G-A translocations. This observation, combined with the fact that today *T. araricum* – the supposed wild progenitor of the AAGG wheats - has an active *Ph1* allele, may suggest that *Ph1* was suppressed during the formation of the AAGG genome and later on restored via genetic segregation.

2.4. Genomic Changes During Domestication

Allopolyploidization was responsible for a rapid wheats genome evolution, including the elimination of single copy DNA, activation and silencing of specific genes and activation of retrotransposons (Ozkan et al., 2001; Kashkush et al., 2002; He et al., 2003; Gu et al., 2004). Feldman et al. (1997) conclude that the physical basis for a diploid-like meiotic behaviour of polyploidy wheat is under the control of two molecular processes: I) the non-random elimination of DNA sequences; II) the homoeologous pairing suppression by the *Ph1* locus. The first phenomenon increased the differentiation between homoeologous chromosomes, thus preventing intergenomic recombinations.

A gene sequence comparison approach has been frequently followed to investigate wheat evolution. A comparison of two large orthologous HMW-glutenin regions located on chromosome 1AS of *T. monococcum* and *T. turgidum* sbsp. *durum* indicates that sequence conservation is restricted to relatively small regions containing orthologous genes (Wicker et al., 2003). In this context, it is surprising that more than 90% of DNA sequences are not or are loosely conserved, even in the recently diverged orthologous glutenin regions of the tetraploid wheats. The

genome-specific differential amplification of the glutenin loci has most probably contributed to the genome divergence noted in the regions compared. A second study, dealing with the high molecular weight (HMW) glutenin loci (Gu et al., 2004), and based on sequencing of 307 kb region, has considered the *T. turgidum* A genome compared to the orthologous regions of the B genome of the same wheat and the D genome of *Ae. tauschii*. Several inactive genes were detected and, curiously, this was found more frequently in the A genome. In addition, a large number or retroelements was recognised (Li et al., 2004). The explanation proposed for this finding may indicate that polyploidization acts as a buffer against the accumulation of deleterious mutations (Wendel, 2000). The presence and activity of retroelements contributes to rapid genome expansion, as pointed out by Kong et al. (2004). The authors provide a comparison between orthologous regions including the high-molecular weight (HMW) glutenin of the wheat B and D genomes. The differences they observe between the B genomes of *T. turgidum* and *T. aestivum* could be due to events occurred since the appearance of the hexaploid wheats.

Belyavev et al. (2000), using genomic *in situ* hybridization, revealed a chromosome pattern interpreted as a process of enrichment of the A genome with repetitive sequences deriving from the B genome. Similar genomic changes may provide a physical basis for phenomena noted in polyploid wheats, such as gene silencing, or non random distribution of markers (typical of B, but not of A chromosomes; Peng et al., 1999). When synthetic allohexaploids were analysed, a fast non-random sequence elimination was observed in the first hybrid generations (Ozkan et al., 2001; Shaked et al., 2001). When synthetic genomes were compared in *Aegilops-Triticum* lines and in their diploid progenitors, it was noted that allopolyploidy induced a very early DNA sequence elimination (Feldman et al., 1997). Both expansion and reduction of wheat genomes have been described by Chantret et al. (2005).

In this context it is evident that transposable elements have been active in these genomes since the divergence of the A, B, and D chromosomes, as well as after the allopolyploidization event. For example, the high rate of genomic rearrangements at the *Ha* locus supports a plastic and dynamic nature of genomes. Issues raised by grass genomes studies are important when the "stabilization" of other allopolyploids genomes is under discussion, particularly when considering the possibility of bringing new species of synthetic origin into agricultural systems.

3. GENOMICS OF PHOTOPERIOD, VERNALIZATION, HEADING DATE, PLANT HEIGHT, AND ERECT PLANT TYPE

Modern plant breeding is supposed to have utilized, and to continue to use, novel variants of genes which have been fixed as "domestication-related alleles" (Paterson et al., 1995; Salamini et al., 2003). In barley, for example, domestication resulted in the elimination of *Btr* (*brittle rachis*) alleles causing brittleness of the rachis; later on, additional loci conditioning rachis weakness were identified and thus also became a target for breeding (reviewed in Wanlong and Gill, 2006).

Flowering time, plant height, and other yield components, like tillering capacity, are additional domestication traits currently under phenotypical, and molecular selection. In cereals, comparative genomics has progressively clarified that domestication and breeding operate, at least in part, on the same genes and metabolic pathways controlling, across different species, some of the mentioned traits (Buckler et al., 2001).

In wheat, spike development and date of heading are controlled by three major groups of genes: photoperiod response genes (located on 5A and 5D); vernalization response genes (5A, 5B, and 5D); and earliness *per se* genes (Bullrich et al., 2002; Shindo et al., 2003).

3.1. Photoperiod

Modern breeding has largely exploited the genetic variability at genes controlling flowering time and photoperiod sensitivity. The modification of time of flowering is important for regional adaptation and, eventually, for dry matter accumulation in all cereals. In this sense, short-day flowering wild grasses have been transformed into domesticated varieties in which flowering time is now unaffected by day length (Buckler et al., 2001). The photoperiodic and vernalization genes respond to environmental changes by delaying or accelerating flowering, thus avoiding the exposure of flower primordia to deleterious extremes of temperature. In winter wheat, which requires vernalization when grown in Northern latitudes, specific genes delay ear initiation after the establishment of the seedling in autumn. As a result, the temperature-sensitive stages of ear development do not occur during the winter season. Where summers are too hot for wheat growing, on the other hand, the dominant photoperiod-insensitive alleles *Ppd1*, *Ppd2*, and *Ppd3* accelerate the onset of ear initiation and development as in the early spring day length begins to increase. This allows the plant to complete the reproductive cycle before the start of the dry season (Worland, 1996; Law and Worland, 1997).

The spreading of photoperiodism-supportive wheat alleles reflects the adaptation achieved by breeders operating under varied environmental conditions all around the globe. A particular allele may have been used in preference to other equally potent, only as a consequence of its availability. For example, most European and North American spring wheat depend upon the dominant *Vrn-1A* allele on 5A for their spring habit, and this only reflects their common parentage. In fact, *Vrn-1D* and *5D*, frequent in Chinese and Japanese spring wheat (Stelmack, 1990), would have been also available and are equally suitable, but due to stochastic choices they were not used in Europe. Alleles controlling wheat flowering time are, thus, good markers to follow the world colonization by wheats.

One of the first breeding development was the introduction, early in the XX century, of the gene *Ppd1* from the Japanese variety Akagomugi into Italian breeding programs. This was important for the development of Italian early maturing wheats. This allele moved, around 1930, to Brazil, where it was used in the breeding of the early Brazilian wheats. The *Ppd1* allele was successively picked up by N. Borlaug

in his Mexican wheat programme (Borlaug, 1968), based on the transfer of the short plant height from the variety Norin 10 to modern wheats. The wheat breeding achievement of N. Borlaug and CIMMYT is now known as "green revolution". A third movement of genes was to Australia, where the allele *Ppd2* arrived from England and *Ppd3* from Indian wheats. Both were necessary for the Australian wheat industry to develop new varieties more adapted to dry weather conditions.

The *Ppd1* and *Ppd2* genes are located on chromosomes 2D and 2B respectively, and *Ppd3* on 2A. *Ppd1*, while shortening life cycle, reduces plant height together with tillering, but also spikelet number. The two effects are opposite, in agronomic terms, even if the increase in spikelet fertility compensates for reduction in spikelet number.

3.2. Vernalization

The vernalization process requires a plant to be exposed to a long period of low temperatures to become competent to flower, while preventing the winter damage to the cold-sensitive flowering meristems. In winter wheats, the promotion of flowering is conditioned by long exposure to low temperature. These varieties are sown in the fall, contrary to spring wheats which do not require vernalization and can be planted both in spring and fall. Two are the genes involved in this response: *VRN1* and *VRN2* (Tranquilli and Dubcovsky, 1999). The former provides most of the variation in vernalization requirement of polyploids wheats. In these wheats, the gene is critical for adaptation to autumn sowing: the presence of *VRN1* alternative alleles divides wheat varieties into the winter and spring market classes. This allelic variation has to be related to a strong selection pressure carried out in contrasting environments during the last 8,000 years of wheat domestication and adaptation. Most of the *Triticeae* have a winter habit, a trait recessive to spring habit, where the recessive allele *vrn1* is the ancestral form. The *VRN2* of grasses is a dominant repressor of flowering down-regulated by both vernalization and short day.

The genes *VRN1* and *VRN2* have been cloned (Yan et al., 2003; Yan et al., 2004). The *VRN1* gene is similar to the Arabidopsis meristem identity gene *APETALA 1 (AP1)*, which initiates the transition of the apical meristem from the vegetative to the reproductive state. In diploid wheats and in the A genome of hexaploid wheats, the molecular nature of the dominant *Vrn1-A1* allele, which supports the spring growth habit, has been associated to mutations located in the promoter region, especially in the 50 bp immediately upstream of the ATG, where the level of variation is higher. Exceptions are several dominant alleles like *Vrn-A1*, *Vrn-B1*, *Vrn-D1* of hexaploid wheat, and *Vrn-H1* of barley which are not polymorphic in the promoter region relative to alleles with a wild type (non mutated) promoter. In a recent experiment, Fu et al. (2005) have sequenced the latter *VRN1* alleles and found a large deletion within the first intron of the gene. It has been hypothesized that a 2.8 kb region of intron I includes putative regulatory elements important for vernalization. Two further vernalization genes affect the vernalization response (Yan et al., 2006), barley *SVRN-H3* and wheat *VRN-B4*, which are located respectively to chromosomes

7HS and 7b . Both genes are now referred to as *VRN3*. The two genes are the orthologous of the Arabidopsis *FT* gene, which promotes flowering. These genes are expressed at high level in *Vrn3* plants, and support early flowering which is dominant to delayed. The wheat *Vrn3* dominant allele is characterized by an insertion of a retroelement in the promoter, while the barley gene has a mutation in the first intron, a mutation which differentiates dominant from recessive alleles (Yan et al., 2006).

VRN2 encodes a dominant repressor of flowering that controls negatively *VRN1* (Yan et al., 2004). The gene belongs to the zinc finger transcription factors class. The molecular nature of *VRN2* and the significant epistatic interactions observed for the gene products of *VRN2* towards the *VRN1* promoter, both support the conclusion that the two genes are acting in the same pathway. As the vernalization process reduces the abundance of the VRN2 protein, the *VRN1* gene transcript increases, leading to the competence to flower. Vernalization requirement can be replaced by interrupting long day treatment with a change to short days in the photoperiod, which is associated with a down-regulation of the *VRN2* flowering repressor (Dubcovsky et al., 2006). In hexaploid winter wheat, the downregulation of the *VRN2* gene anticipates flowering for more than one month, demonstrating that the gene plays a role in the regulation of vernalization.

Spring wheats are double recessive at the two loci: the *vrn2* allele encodes an inactive repressor that cannot interact with the *VRN1* promoter. As a consequence, no effect on flowering time can be associated in this background to the molecular variation present in the promoter. The available model proposes that independent mutations in the promoter regions of genes of winter wheat, and barley have resulted in the loss of the recognition site of the *VRN2* repressor, and therefore, in a dominant spring growth habit (*Vrn1* allele).

3.3. Heading Time

Flowering time is regulated by day length (intercepted by photoperiod responsive genes) or by exposure to cold temperatures (mediated by vernalization responsive genes). Other genes exist controlling “earliness per se”, i.e. genes which regulate flowering time independently of the previously mentioned environmental signals. These are responsible for the fine-tuning of wheat flowering time. Bullrich et al. (2002) have mapped to chromosome 1AmL the “earliness per se” gene *Eps-Am1* in *T. monococcum*.

QTL mapping for loci controlling flowering time has been carried out by Peng *et al*. (2003) in a cross between the wild *T. dicoccoides* and the domesticated *T. durum*. The wild parent had typical traits: short stature, late flowering, small seeds, poor yield, and yield components. The authors individuate, by single- and linked-QTL analyses, up to 76 QTL effects. Allelic interactions supporting QTL factors followed dominance, recessive, additive, and heterotic models of inheritance (20% followed the latter). The map position of the QTLs controlling different traits and mapping to the same chromosomes were frequently overlapping. The

most significant QTLs were clustered in limited intervals of chromosomes 1B, 2A, 3A and 5A. The traits scored in the experiment were: spike number/plant; spike weight/plant; single spike weight; kernel number/plant; kernel number/spike; kernel number/spikelet; 100-grain weight; grain yield/plant; and spikelet number/spike. Flowering time QTLs (four) mapped on chromosomes 2A, 4B, 5A and 6B. The wild parent was sensitive to day length and flowered later compared to the cultivar. The authors conclude that the wild allele for the QTL on 5A increases the value of heading date and is responsible for late flowering of *T. dicoccoides*, whereas the wild alleles on 2A, 4B, and 6B accelerate flowering date.

Traditional rice cultivars can be sensitive to photoperiod; because of this, their cultivation can be practiced only once in one season. In most tropical and subtropical countries modern varieties having a photoperiod-insensitive allele of the *Se1*/*Hd-1* gene can be repeatedly grown in the same year (Khush, 2001). *Se1*/*Hd-1* is an orthologous of the closely related *CONSTANS* gene which controls photoperiod response in Arabidopsis (Yano et al., 2000). The *Hd3a* rice gene is controlling the transition to flowering. The gene has been isolated and reveals orthologous to the flowering promoting gene *FLOWERING LOCUS T* (*FT*) of Arabidopsis (Kojima et al., 2002). Breeding for early maturity in rice has also involved the *ef* (*early flowering*) genes (Khush, 2001).

The sowing-harvesting period of wheat can be similarly reduced exploiting the already mentioned *Ppd1* and *Ppd2* genes, that cause photoperiod insensitivity (Börner et al., 1998; reviewed in Griffiths et al., 2006).

3.4. Plant Height

Lodging is a detrimental phenomenon which accentuated its negative role in the 1960s, when the amount of fertilizers supplied to wheat and rice increased significantly. The problem was overcome by the deployment of new semi-dwarf varieties that were lodging-resistant. The success of the new genotypes was also depending on their capacity of partitioning a higher proportion of photosynthates into the grain, leading to a dramatic yield increases: the “green revolution” (reviewed in Hedden, 2003). The “green revolution” in rice is based on the use of mutations in genes controlling a key-step in gibberellin biosynthesis (for example, the *Semidwarf1 Sd1* gene encodes for a GA20-oxidase enzyme; reviewed in Pozzi et al., 2004). In wheat, semi-dominant mutations of the homoeologous/duplicated *Rht-B1a* and *Rht-D1a* genes conferring dwarfism characterize recent high yielding cultivars (Börner et al., 1996; Peng et al., 1999; reviewed in Hedden, 2003). *Rht-B1b* or *Rht-D1b* genes, present today in most wheat varieties, have been transferred from the Japanese variety Norin 10 into CIMMYT germplasm under the direction of N.Borlaug. Starting with CIMMYT germplasm, the same genes spread world-wide to wheat breeding programs. *Rht8*, also capable to reduce wheat plant height by 10 cm, derives from the Japanese variety Akagomugi. Probably the Korean peninsula was the center of origin of all these mutations which go back as early as in the third and the fourth centuries (Cho et al., 1993).

Other height reducing genes are known: the group of *Rht* genes may include up to 20 genes. In addition, genes were identified that reduce plant height without affecting early growth, or coleoptile length and vigour as do the *Rht* genes (Ellis et al., 2004). In a recent study these genes were mapped to different wheat chromosomes, thus widening their exploitation in plant breeding (Ellis et al., 2005). Wheat breeders have already looked at mutants of the duplicated gibberellin insensitivity loci to design varieties with specific heights (Peng et al., 1999).

The *Rht* (of the subgroups B and D) gene encodes a repressor of GA signalling. The gene is the orthologous of the Arabidopsis *GAI*, maize *dwarf8 (d8)*, and barley *Slender1 (Sln1)* genes. Mutations in these genes, which result in GA-insensitive dwarfs are commonly observed (Peng et al, 1997; Peng et al., 1999; Chandler et al., 2002). Genes belonging to this class -also present in rice (Ikeda et al., 2001)- encode GRAS transcription factors characteristic for the presence of the DELLA domain in their encoded protein product. Deletions of this domain render the protein insensitive to GA-dependent mRNA degradation and result in dwarfing alleles (Peng et al., 1997; Peng et al., 1999). In wheat, the *Rht* genes have pleiotropic effects on flowering time and tillering (Khush, 2001). Interestingly, *d8* of maize is a target of selection and adaptation to various flowering times (Thornsberry *et al*., 2001). Pleiotropic effects are not surprising for genes controlling hormone action and may be a common occurrence for the traits targeted by domestication and breeding (Cai and Morishima, 2002; Salamini, 2002).

4. GENOMICS OF SEED-RELATED YIELD COMPONENTS

The first steps in domestication concentrated on genes that facilitate harvesting and enable colonization of new environments. Soon after, yield must have assumed priority, as to minimize labour input and land needs (Buckler et al., 2001).

4.1. Seed Size

A transition from small-seeded plants with natural seed dispersal to larger-seeded non-shattering plants is evident for all seed crops. Seed size was positively selected in all domesticated cereals. The genetic control of seed size in domesticated versus wild tetraplolid wheats was studied by generating *T. dicoccoides* substitution lines in *T. durum* (Elias et al., 1996). Kernel size is under complex polygenic control, and alleles contributing to an increase and to a decrease in kernel size have been mapped on chromosomes 1A, 2A, 3A, 4A, 7A, 5B, and 7B. In a recent experiment (Peng et al., 2003) eight QTL for grain weight were mapped. Tetraploid wheat and rice were compared based on functional genomics. The study indicates that all detected seed-size QTLs in *T. dicoccoides* correspond to their rice counterparts. QTLs for *T. aestivum* yield and for other important agronomic traits were also mapped to chromosomes 3A, 4A, and 5A (Shah et al., 1999; Campbell et al., 2003; Araki et al., 1999; Kato et al., 2000). Additional QTL controlling plant adaptation and morphology traits were reported, including heading date, plant height (Cadalen

et al., 1998), lodging, lead rust reaction, and spike morphology (Sourdille et al., 2002). In the work of Marza et al. (2006), in addition to grain yield, QTLs for physiological maturity date, plant height, shattering, and lodging were assigned to linkage groups 4B, 5A, 6A, 6B, 7A, and 7DL (inconsistency among environments and moderate QTL effect were, nevertheless, noted). Grain yield QTLs were assigned to chromosomes 5A and 4B, while shattering to 6B. Lodging QTLs were mapped to 1B, 4AL, and 5A. In a recent study (Breseghello and Sorrells, 2006) the association of 36 unlinked simple-sequence repeat markers with kernel size and milling quality was analyzed in a collection of modern cultivars of soft winter wheat. Authors concentrated on chromosomes 2D and 5A/5B , where they suspected the presence of QTLs for kernel size, and evaluated linkage disequilibrium (LD) observed among linked markers. A consistent LD for markers of chromosome 2D extended on linkage intervals shorter than 1 cM, while in the centromeric region of 5A LD was significant for ca. 5 cM. A refinement of their mapping, based on further 62 linked SSS_2, revealed on the three chromosomes tested markers with a significant LD linked to kernel size. Candidate genes mapping in the regions hosting the QTLs are described elsewhere in this text: *Rht8* (reduced height); *Ppd1* (response to photoperiod), *B1* (awnedness inhibitor), *Rht12* (reduced height), and *Vrn1* (response to vernalization).

4.2. Genomics of Grain Hardness

Grain hardness (i.e. whether the endosperm is physically hard or soft) is a major property of the grain of *T. aestivum* (for example, it controls the susceptibility to damage during milling and, thus, the amount of water uptake during baking). Compared to *T. aestivum*, *T. durum* is mainly used for pasta production because of its hard endosperm.

Already in the '60s, Symes (1965), using the particle size index as a measure for grain hardness, demonstrated the segregation of a major gene for this trait in the progenies derived from the cross of the wheat cultivars Heron and Falcon. Later, he confirmed this result using 7 additional crosses between hard and soft wheats. The gene influencing hardness (i.e. conferring the "soft" phenotype) named *Ha* was subsequently mapped to the short arm of chromosome 5D (Sourdille et al., 1996). The *Ha* locus encodes friabilins, consisting of three lipid binding proteins: puroindoline a (encoded by *Pina*), puroindoline b (encoded by *Pinb*) and Grain Softness Protein (encoded by *Gsp-1*) (Gautier et al., 1994; Rahman et al., 1994). The three genes are tightly linked and define the *Ha* locus on chromosome 5D (Sourdille et al. 1996; Giroux and Morris, 1998; Giroux et al. 2000).

To understand the molecular nature of the locus, a bacterial artificial chromosome (BAC) covering that region was analysed. The three genes were found in the order (from 5' to 3' ends) *Pinb*, *Pina*, and *Gsp-1* within a 60-kb fragment of the D genome deriving from *Ae. tauschii* (Turnbull et al. 2003). Interestingly, the same genearrangement has been observed in *T. monococcum* (Chantret et al. 2005), a wheat with a genome closely related to the A^u genome donor of the A genome

of wheat. Hard wheats have a mutation in either *Pina* or *Pinb* (but not in *Gsp-1*). Transformation experiments (Beecher et al., 2002) proved the direct functional role of the *Pin* genes: transformed rice and wheat had an endosperm more soft than untransformed controls. These results still do not contribute to explain the biochemical basis of the action of the *PIN* proteins.

Transcript abundance of *Pinb* do not vary significantly between soft and hard seeds in any of the wheats tested (Beecher et al., 2002). In addition, other genetic factors unlinked to the *Ha* locus may contribute to modify grain hardness (mapping on chromosomes 1A and 6D; Perretant et al. 2000). In general it can be concluded that the *Pin* genes are the primary cause of grain softness and that at least two puroindoline genes should be present for the grain texture to be soft (Giroux and Morris, 1998). In rice the expression of *Pinb* alone is sufficient to increase grain softness (Krishnamurthy and Giroux 2001). A more recent study (Capparelli et al., 2003) indicates that the level of friabilin associated with starch is regulated by the level of expression of the *Pina* transcripts, and that for a fully functional friabilin by both PINA and PINB gene products are required. Studies conducted using a microarray based approach demonstrate that the major difference in *Pina* gene expression is necessary to determine the hardness of the grain (Clarke and Rahman, 2005).

The *Ha* locus (ca. 80 kb) represents a classical example of a locus controlling a trait whose variation arose as a result of gene elimination after polyploidization. The domestication history of the locus is complex: it arose in diploid wheats, it was lost in tetraploids, and was later on introduced in the hexaploids via the D genome. The D-version of the locus present in the wild ancestor *Ae. tauschii* is 29 kb larger than in hexaploid wheats (Giroux and Morris, 1998). *Pina* and *Pinb* gene loss from the *Ha* locus of tetraploids was caused by a large genomic deletion that probably occurred independently in the A and B progenitor genomes of the tetraploid wheats. Duplication of puroindoline genes in the *Pooideae* has shaped the *Ha* locus (Chantret et al., 2005). After the acquisition of *Ha* by hexaploid wheats via the D genome (at the time of the emergence of allohexaploid bread wheat), the spreading of the species to northern cultivation areas has forced breeders to select hard wheat hexaploid cultivars, and thus the selection pressure on *Ha* has increased. This resulted in mutations or genomic rearrangements at the locus.

4.3. Tillering

When studied, grasses and particularly cereals have revealed the existence of several mutants with an altered pattern of axillary bud development. In maize, the *teosinte branched1* (*tb1*) mutation causes a complete loss of apical dominance, associated to axillary buds grow, a phenotype mimicking the teosinte phenotype (Doebley et al., 1997). Tillering and branching mutations have been described also in barley, where the recessive *uniculm2* mutation on chromosome 6H fail to produce tillers (Babb and Muehlbauer 2003). The *moc1* mutation of rice, on chromosome 6, causes lack of axillary buds (Li et al., 2003). Additional recessive tillering mutants

with reduced culm number (*rcn1* to *rcn5*) have been reported in rice (Gramene, http://www.gramene.org/index.html). The genetic variation for the capacity of tillers to exist has been assessed for the wheat gene pool (Atsmon and Jacobs 1977): low tillering genotypes frequently have a uniculm phenotype, enlarged spike and modified leaf morphology. Some wild diploid grasses belonging to the *Aegilops* section *Sitopsis* (like *Ae. speltoides*, the progenitor of wild tetraploid emmer), have an ear morphology mimicking the teosinte ear. This specific ear shape is even more similar to teosinte than current domesticated wheats. It is concluded that current wheats have already undergone changes leading to a maize-cob like morphology. In wheat, the uniculm line "492" was studied by Richards (1988), and a single recessive gene (*tin*) located on chromosome 1AS was found to control tiller number (Spielmeyer and Richards, 2004). The absence, in *tin*, of tiller development is due to an altered pattern of axillary bud formation and outgrowth (W. Spielmeyer, unpublished). In *tin* populations, low-tillering lines have a greater harvest index, fewer and sterile tillers, and a larger grain size, demonstrating the agronomic potential of the *tin* gene in wheat (Richards 1988; Duggan et al., 2002). Comparative genomics analyses lead to the conclusion that *tin* and rice reduced tillering mutations and the barley *uniculm2* mutant map to non-syntenic chromosomes. A wheat sequence that was closely related to the gene that encodes the *tb1* mutation in maize hybridises to a single RFLP band on chromosome group 4 of wheat (W. Spielmeyer, unpublished).

5. GENOMICS OF SEED TRAITS RELATED TO HARVESTABILITY

5.1. Free-Threshing

At the start of wheat cultivation, varieties were characterized by hulled seeds. The harvested spikelets were stored in barns and, when needed, they were parched to liberate the seeds from the chaff. This operation of parching is often reported in the literature on hulled wheats. Nesbitt and Samuel (1995) list additional reasons for the use of heat in wheat harvest processing. Whatever was the significance of spikelet parching, the end results is that in excavated archaeological layer charred seeds are present which allow paleobotanical conclusions.

A relevant agricultural improvement was the introduction of free-threshing varieties. Species characterized by a low degree of glume tenacity and by fragile rachis and free-threshing habit were selected by the farmers and the harvest of grains became efficient. Free-threshing wheats have thinner glumes and pales which allow an early release of naked kernels. After threshing, free grains are winnowed and stored ready for milling. Free-threshing varieties – tetraploid hard wheat and hexaploid bread wheat – represent the final steps of *Triticum* domestication.

The appearance of the free-threshing phenotype can be discussed considering the comparative genetics of two hexaploid wheat types, one hulled (called spelt), and the other free-threshing. Both have the AABBDD hexaploid genome. Major

and minor mutations have been considered to expalin the evolution of the free-threshing habit in this type of wheat (McKey, 1966). The major gene components is usually discussed starting with the gene or gene complex *Q*, located on the arm of chromosome 5A. In *T. vulgare*, the *Q* allele supports the formation of square-headed ears with good threshability, besides inducing softening of the glumes, reduction of ear length, more spikelets per ear, and toughness of the rachis (Snape et al., 1985; Kato et al., 1998; Kato et al., 2003). Mutagenic disruption of the *Q* gene in *QQ* hexaploid free-threshing wheat generates a *q* mutant phenotype, known as speltoid because *q* mutants have a morphology similar to that of spelt (i.e. tenacious glumes), with a *qq* genotype. Bread wheat lines which are *Qq* have intermediate phenotypes. Because spelt has the *q* allele, Sears was confident that the large *Q* allele of free-threshing wheats derives from a form of *q* that retains some activity. As a matter of fact, Muramatsu (1963) showed that *q* is active by creating genotypes with 1–5 doses of either *Q* or *q* alleles. He showed that a square-headed hexaploid ear derives from either two doses of *Q* or five doses of *q*. A second major gene (*Tg*) is associated with the D genome. Two further major genes are known: *C* on the long arm of chromosome 2D (Rao 1972) and *S-D1* on the long arm of chromosome 3D (Rao 1977). In hexaploid wheats the polygenic component is scattered throughout all three A, B, and D genomes. In tetraploid wheats, QTL studies identified a total of four putative loci for threshability (Simonetti et al., 1999), located on chromosomes 2B, 5A and 6A. Two of these QTLs correspond, in position, to the already mentioned genes *Q* and *Tg*. Sourdille et al. (2000) explored the mapping of QTL involved in ear morphology (spike length, number of spikelets and compactness). For each trait, 4 to 6 QTLs were identified with individual effects ranging between 7% and 22% or total phenotypic variation. Several QTLs were detected that affect more than one trait. In general, these studies revealed the existence of additional QTLs, but their number and location varied according to the population under study.

5.2. Brittle-Rachis

In a tough, non-brittle rachis the formation of fracture rachis zones is suppressed, or the collapse of the rachis is delayed until mature spikes are harvested. In virtue of this characteristic *T. dicoccum*, the domesticated emmer wheat, diffused also in Mediterranean countries during the Bronze Age. The same is true for einkorn (*T. monococcum*, non-brittle) compared to its wild progenitor *T. boeoticum*. Thus the brittle rachis trait is at the origins of agriculture and sedentary societies.

In *T. turgidum*, two dominant genes, mapping to chromosomes 3A and 3B are responsible for the brittle trait (Watanabe et al., 2002; Watanabe et al., 2005; summary of genes involved and on the related terminology in Salamini et al., 2002). The trait, in einkorn, is under the control of two genes that segregate 15 brittle to 1 tough rachis in the F2 progeny of wild x domesticated crosses (Sharma and Waines, 1980). In *T. aestivum* the group of Riera-Lizarazu (Jantasuriyarat et al., 2004) detected several QTL involved in free-threshing and spike compactness. Four

QTLs for glume tenacity were detected; six for threshability, which is related to non-brittle rachis; four for spike length; five for number of spikelets per spike and five for spike compactness. Some of the identified QTLs probably coincide with known genes, like *Tg*, or *Q*. Cao et al. (1997) identified a single dominant gene, *Br-A1*, as the determinant of rachis fragility in a feral form of *T. aestivum* from Tibet. The gene was later localized on chromosome 3D (Chen et al., 1998), as supported also by studies on T.*dicoccoides* x T.*aestivum* crosses (Rong et al., 2000).

To summarize, studies like those cited point to the existence of different loci controlling rachis characteristics but also glumes and ear rachis traits (Salamini et al., 2002), suggesting that: i) multiple genetic pathways are involved in controlling the trait(s); ii) the donors of the genomes that made up polyploid wheats were different at the locus controlling shattering, a conclusion that would support a diphyletic origin of tetraploid wheats. These considerations, combined with the mapping of shattering QTLs, allows the study of microsyntenous relationship of the trait in barley, rice, *T. turgidum* and *T. aestivum*, sorghum, and maize, pointing to a conservation of the identified region.

The rice shattering QTL (*qSHT-1*) is located in region syntenous with regions on chromosomes 3A, 3B and 3H (respectively of wheat genomes A and B, and of barley), even if comparative mapping indicates that the genes may not be orthologous (Nalam et al., 2006; Li et al., 2006). The *qSH*1 gene has been recently cloned and shown to encode a BELL1-type homeobox gene (Konishi et al., 2006). Similar consideration can be made in the case of barley: the *Btr1* and *Btr2* barley loci controlling shattering (Komatsuda et al., 2004) are mapping in a region not syntenous to wheat. In *T. dicoccoides,* Peng et al. (2003) identified a gene controlling shattering and mapped it on the long arm of chromosome 2A. In the same region maps one of the QTL involved in 4n wheat domestication. Additional QTLs controlling the domestication syndrome are tightly linked in the same species to the *Br* gene and/or were the result of a pleiotropic effect of *Br*.

The *Q* and *Tg* genes affect, besides free-threshing, also the toughness of the rachis. For this reason, the *Q* gene is agronomically important. Upon cloning, *Q* and *q* alleles were found to differ both in structure and in the level of transcription, and this can be associated to phenotypic differences due to dosage effects observed in transgenic plants. High resolution mapping of the gene (Faris and Gill, 2002) was instrumental for the molecular identification of this gene which shows DNA sequence similarity to the Arabidopsis *APETALA2* gene (Faris et al., 2003). Further studies confirmed the association (Simons et al., 2006) and demonstrated that ectopic expression of *Q* in transgenic plants mimicked dosage and pleiotropic effects of *Q*. Increased transcription of *Q* was associated with spike compactness and reduced plant height. Previous research suggested that *Q* might have arisen from a duplication of *q* (Kuckuck, 1959), as it was observed that 5 doses of *q* conferred the same phenotype as 2 doses of *Q*. Simons et al. (2006) contradict this hypothesis, showing that most probably *Q* arose trough a gain of function mutation.

Synthetic hexaploids deriving from free-threshing tetraploids crossed to hulled *Ae. tauschii* allowed the identification of a second genetic system that controls the

speltoid phenotype. The hybrids were hulled, even in the presence of the dominant *Q* allele. The suppression of the free-threshing character was thought to be due to a partially dominant *Tg* allele, con chromosome 2D of *Ae. tauschii*, with *Tg* inhibiting *Q* and thus leading to tenacious glumes. The conclusion is that free-threshing hexaploids have the genotype *tgtg*, *QQ*, whereas all tested *Ae. tauschii* lines have the dominant *Tg* allele (Kerber and Rowland, 1974). Villareal et al. (1996) found that, in synthetic hexaploids, two genes (rather than *Tg* alone) affect glume tenacity, both of which are contributed by the D genome. Rachis fragility and square-headed ears in free-threshing hexaploid wheats can assort independently. For instance, in experimental crosses with *T. spelta*, three genes segregate for rachis fragility, but only one, *q*, for tenacious glumes. Moreover, *q* and *Q* interact with genes on other chromosomes that also govern glume tenacity and fragility (Cao et al., 1997; Luo et al., 2000). A further gene, located at 32 cM from *Tg*, on chromosome 2D, belongs to the system.

When the free-threshing habit was studied in a *T. durum* x *T. dicoccoides* cross, the F2 population showed a continuous distribution for the trait (Simonetti et al., 1999). F3 lines were fingerprinted and four major QTL were located on three chromosomes. Two of these corresponds to the homologous genes *Tg* and *Q* of hexaploid wheats, whereby the *Tg*-like gene *Tg2* is located on chromosome 2. It is concluded that the polygenic control of the free-threshing character in tetraploid wheats is based on the selection on new loci in addition to *Q* and *Tg*.

5.3. Control of Awning

Awns are important for their contribution to yield under drought conditions; as such the trait was most probably subjected to domestication. The genetic control of awn presence/absence is simple: few genes control the differences between awned and awnless varieties. Three dominant inhibitors are described: *Hd* (Hooded), *B1* and *B2* (*tipped 1* and *2*) (McIntosh et al., 1998). Wheat genotypes homozygous for three recessive alleles are fully awned. Using aneuploid lines, the loci were mapped using on chromosome 4A, 5Am and 6B respectively, as recently reported also by Sourdille et al. (2002) in a QTL experiment.

6. WHEAT GENETIC DIVERSITY TRENDS DURING DOMESTICATION AND BREEDING

Crops seem molecularly less variably than wild relatives. This is explained by the bottlenecks that a species has experienced during domestication and breeding (Tanksley and McCouch, 1997; Wright and Gaut, 2005; Dvorak and Akhunov, 2005; Kilian et al., 2006). When nucleotide diversity was evaluated at 21 loci in 4 taxa, representing key steps in the evolution of wheats, the results indicated a single domestication event of the wild tetraploid and a possible diphyletic origin of bread wheat Loss of molecular diversity in *T. dicoccu*m (durum wheat), compared to *T. dicoccoides* was 84%, and 69% in *T. aestivum* (David J., pers. comm.). Similar

data have been provided by a study of barley, carried out analyzing seven genes (Kilian et al., 2006) and comparing wild to domesticated barley, where the number of haplotypes (70 *vs*. 17), Π values (7.7 *vs*. 2.8 x 10^{-3}) and Wattersons's theta (10 *vs*. 2.8 x 10^{-3}) were significantly reduced in domesticated lines.

It has to be noted that in the cited cases the loss of diversity was calculated from data concerning several genes, and each of them had a different degree of nucleotide variation. For instance, Ravel et al. (2006) noted that in 26 bread wheat lines SNPs frequency was relatively low (1 every 335 bp), and essentially represented by the variation at only 2 out of 21 genes. This and similar findings may explain the contrasting data available from literature survey. For example Thuillet et al. (2005) found that the molecular diversity in domesticated tetraploid wheats was similar to the one of wilds. Dvorak et al. (2006) have reasoned that mutations-supported nucleotide diversity in domesticated 4n wheats had no relevance in the generation of present diversity. They concluded that the diversity in today cultivated 4n wheats is due to gene flow from wild lines (Luo et al., 2006). Gene flow has been reported also from wild emmer to *T. aestivum* in the case of the ABCT-A1 locus (Dvorak et al., 2006).

In general, it seems safe to summarize that the relatively young history of the wheat crop, the presumably small founder population of this gene pool, and the intensive long-term selection for agronomic traits did set the basis for a reduced genetic variability in the genus (Reif et al., 2005). Part of the reduction in genetic diversity in modern wheats (Lubbers et al., 1991; Lelley et al., 2000) is probably the product of genetic drift and selection, both unintentional and intentional, carried out by early farmers. Ninety-seven percent of land races were rapidly replaced, during the last century, by new wheat cultivars (Smale et al., 2002), which have been bred starting with a limited number of land races. The extended period of the last 50 years of intensive selection may have further reduced genetic diversity among cultivars, narrowing the germplasm base available for future breeding (Tanksley and McCouch, 1997). A risk is that mutation in pest populations, the widespread use of monocultures and the changes in environmental conditions introduce new biotic and abiotic stresses that the current cultivars can not cope with and, therefore, lead to severe crop losses. For example, in India during the 70ies a severe epidemics of shoot fly and karnal bunt have heavily affected wheat yield.

Where and how do we find a safe and sustainable approach to modern wheat breeding? Introgression of novel germplasm is one option. Reif et al. (2005) used CYMMIT germplasm and reported a loss of genetic diversity in spring wheat during domestication followed by a change from traditional land races to modern breeding varieties during the last 50 years of intentional breeding. The second step accounted for most of the reduction in genetic variability. This trend is supported by other and various surveys, pointing to the potential of land races stored in gene banks as a source of useful allelic variation (Cox et al., 1992; Villareal et al., 1995). In modern varieties, moreover, intraspecific relatedness of germplasm increased in time periods, as also evident for the stabilization and fixation of certain well defined wheat plant ideotypes, a kind of strong selective drift. Of course, the

green revolution has contributed (Evenson and Gollin, 2003): the high-yielding new semi-dwarf cultivars introduced by the green revolution were based on a limited number of key parents and their genes are now dominating the wheat germplasm base. Roussel et al. (2004) studied 559 French bread wheat accessions and found a significant reduction of genetic diversity. However, they also observed a recent increase in diversity which could be explained by a change in the late 70ies in the breeding strategy at CIMMYT. The CIMMYT's wheat breeding program is now increasing wheat genetic diversity on a large scale to respond to the need for biological diversification, environmental sustainability, and durable resistance to pathogens.

REFERENCES

Akhunov ED, Goodyear JA, Geng S, Qi L-L, Echalier B et al (2003) The organization and rate of evolution of the wheat genomes are correlated with recombination rates along chromosome arms. Genome Res 13:753–763

Akhunov ED, Akhunova AR, Anderson OD, Anderson JA, Blake N et al (2006) SNP discovery and deployment in polyploid wheat. Paper presented at the Plant Animal Genome XIV, San. Diego, CA

Aragon-Alcaide L, reader S, Miller T, Moore G (1997) Centromeric behaviour in wheat with high and low homologous chromosomal pairing. Chromosoma 106:327–333

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

Atsmon D, Jacobs E (1977) A newly bred 'Gigas' form of bread wheat (*Triticum aestivum* L.): morphological features and thermo-photoperiodic responses. Crop Sci 17:31–35

Babb S, Muehlbauer GJ (2003) Genetic and morphological characterization of the barley uniculm2 (cul2) mutant. Theor Appl Genet 106:846–857

Barton NH, Keightley PD (2002) Understanding quantitative genetic variation. Nat Rev Genet 3:11–21

Beecher M, Bettege A, Smidansky E, Giroux MJ (2002) Expression of wild type *pinB* sequence in transgenic wheat complements a hard phenotype. Theor Appl Genet 105:870–877

Belayev A, Raskina O, Korol A, Nevo E (2000) Coevolution of A and B genomes in allotetraploid *Triticum dicoccoides*. Genome 43:1021–1026

Borlaug N (1968) Wheat breeding and its impact on world food supply. Canberra, Australia: Proceeding III International Wheat Genetics Symposium

Borner A, Korzun V, Worland AJ (1998) Comparative genetic mapping of loci affecting plant height and development in cereals. Euphytica 100:245–248

Brandolini A, Vaccino P, Boggini G, Ozkan H, Kilian B, Salamini F (2006) Quantification of genetic relationships among A genomes of wheats. Genome 49:297–305

Breseghello F, Sorrels ME (2006) Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. Genetics 172:1165–1177

Brunner S, Fengler K, Morgante M, Tingey S, Rafalski A (2005) Evolution of DNA sequence non homologies among maize inbreds. Plant Cell 17:343–360

Buchner P, Prosser IM, Hawkesford, MJ (2004) Phylogeny and expression of paralogous and orthologous sulphate transporter genes in diploid and hexaploid wheats. Genome 47:526–534

Buckler E, Thornsberry JM, Kresovich S (2001) Molecular diversity, structure and domestication of grasses. Genet Res 77:213–218

Bullrich L, Appendino ML, Tranquilli G, Lewis S, Dubcovski J (2002) Mapping of a thermo-sensitive earliness *per se* gene on *Triticum monococcum* chromosome $1A^m$. Theor Appl Genet 105:585–593

Cadalen T, Sourdille P, Charmet G, Tixier MH, Gay G, Boeuf C, Bernard S, Leroy P, Bernard M (1998) Molecular markers linked to genes affecting plant height in wheat using a double-haploid population. Theor Appl Genet 96:933–940

Cai W, Morishima H (2002) QTL clusters reflect charcater associations in wild and cultivated rice. Theor Appl Genet 104:1217–1228

Caldwell KS, Dvorak J, Lagudh ES, Akhunov E, Luo MC et al (2004) Sequence polymorphism in polyploid wheat and their D genome diploid ancestor. Genetics 167:941–947

Campbell B, Baenziger PS, Gill KS, Eskridge KM, Budak H, Erayman M, Dweikat I, Yen Y (2003) Identification of QTLs and environmental interactions associated with agronomic traits on chromosome 3A of wheat. Crop Sci 43:1493–1505

Cao W, Scoles GJ, Hucl P (1997) The genetics of rachis fragility and glume tenacity in semi-wild wheat. Euphytica 94:119–124

Capparelli R, Boriello G, Giroux MJ, Amoroso MG (2003) Puroindoline A-gene expression is involved in association of puroindolines to starch. Theor Appl Genet 107:1463–1468

Chandler PM, Marion-Poll A, Ellis M, Gubler F (2002) Mutants at the *Slender1* locus of barley cv. Himalaya. Molecular and physiological characterization. Plant Physiol 129:181–190

Chantret N, Salse J, Sabot F, Rahman S, Bellec A, Laubin B, et al (2005) Molecular basis of evolutionary events that shaped the *Hardness* locus in diploid and polyploid wheat species (*Triticum* and *Aegilops*). Plant Cell 17:1033–1045

Chen Q-F, Yen C, Yang J-L (1998) Chromosome location of the gene for *brittle rachis* in the Tibetan weedrace of common wheat. Genet Res Crop Evol 45:21–25

Cho C, Kyu HO, Lee SH (1993) Origin, dissemination and utilization of semi-dwarf genes in Korea. In Miller T, Koebner RMD (eds) Proceedings VII International Wheat Genetic Symposium Bath, Bath Press, pp 223–231

Clarke B, Rahman S (2005) A microarray analysis of wheat grain hardness. Theor Appl Genet 110:1259–1267

Cox TS (1998) Deepening the wheat gene pool. J Crop Prod 1:1–25

Cox T, Wilson WJ, Gill DL, Leath S, Bockus WW, Browder LE (1992) Resistance to foliar diseases in a collection of *Triticum tauschii* germplasm. Plant Dis 76:1061–1064

Diamond J (1997) Guns, germs and steel. Random House, London

Doebley J, Stec A (1993) Genetic analysis of the morphological differences between maize and teosinte. Genetics 141:333–346

Doebley J, Stec A, Hubbard L (1997) The evolution of apical dominance in maize. Nature 386:485–488

Driscoll, C (1972) Genetic suppression of homoeologous chromosome pairing in hexaploid wheat. Can J Genet Cytol 14:39–42

Dubcovsky, J, Loukoianov A, Fu D, Valarik M, Sanchez A, Yan L (2006) Effect of photoperiod on the regulation of wheat vernalization genes VRN1 and VRN2. Plant Mol Biol 60:469–480

Duggan B, Richards RA, Tsuyuzaki H (2002) Environmental effects on stunting and the expression of the tiller inhibition (*tin*) gene in wheat. Funct Plant Biol 29:45–53

Dvorak J, Akhunov ED (2005) Tempos of deletions and duplications of gene loci in relation to recombination rate during diploid and polyploid evolution in the *Aegilops–Triticum* alliance. Genetics 171:323–332

Dvorak J, Luo MC, Yang Zl, Zhang HB (1998) The structure of the *Aegilops tauschii* genepool and the evolution of hexaploid wheat. Theor Appl Genet 67:657–670

Dvorak J, Akhunov ED, Akhunov AR, Deal KR, Luo MC (2006) Molecular characterization of a diagnostic DNA marker for domesticated tetraploid wheat provides evidence for gene flow from wild tetraploid wheat to hexaploid wheat. Mol Biol Evol 23:1386–1396

Elias EM, Steiger KD, Cantrell RG (1996) Evaluation of lines derived from wild emmer chromosome substitutions II. Agronomic traits. Crop Sci 36:228–233

Ellis M, Rebetzke GJ, Chandler P, Bonnett D, Spielmeyer W, Richards RA (2004) The effect of different height reducing genes on early growth characteristics of wheat. Funct Plant Biol 31:583–589

Ellis M, Rebetzke GJ, Azanza F, Richards RA, Spielmeyer W (2005) Molecular mapping of gibberellin-responsive dwarfing genes in bread wheat. Theor Appl Genet 111:423–430

Evenson R, Gollin D (2003) Assessing the impact of the green revolution, 1960–2000. Science 300: 758–762

Faris J, Gill BS (2002) Genomic targeting and high resolution mapping of the domestication gene *Q* in wheat

Faris J, Fellers JP, Brooks SA, Gill BS (2003) A bacterial artificial chromosome contig spanning the major domestication locus *Q* in wheat and identification of a candidate gene. Genetics 164:311–321

Feldman M (2001) Origin of cultivated wheat. In: The World Wheat Book: A history of wheat breeding. Lavoisier Publications, Paris

Feldman M, Liu B, Segal G, Abbo S, Levy AA, Vega JM (1997) Rapid elimination of low-copy DNA sequences in polyploid wheat: a possible mechanism for differentiation of homoeologous chromosomes. Genetics 147:1381–1387

Freeling M, Thomas BC (2006) Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. Genome 16:805–814

Fu D, Szucs P, Yan L, Helguera M, Skinner JS, von Zitzewitz J, Hayes PM, Dubcovsky J (2005) Large deletions within the first intron in *VRN1* are associated with spring growth habit in barley and wheat. Mol Gen Genomics 273:54–65

Gautier M, Aleman ME, Guirao A, Marion D, Joudrier P (1994) *Triticum aestivum* puroindolines, two basic cystine-rich seed proteins: cDNA sequence analysis and developmental gene expression. Plant Mol Biol 25:43–57

Giroux M, Morris CG (1998) Wheat grain hardness results from highly conserved mutations in the friabilin components *puroindoline a* and *b*. Proc Natl Acad Sci USA 95:6262–6266

Giroux MJ, Talbert L, Habernicht DK, Lanning S, Hempill A, Martin JM (2000) Association of puroindoline sequence type and grain hardness in hard red spring wheat. Crop Sci 40:370–374

Griffiths S, Sharp R, Foote TN, Bertin I, Wanous M, Reader S, Colas I, Moore G (2006) Molecular characterization of *Ph1* as a major chromosome pairing locus in polyploid wheat. Nature 439:749–752

Gu Y, Coleman-Derr D, Kong X, Anderson OD (2004) Rapid genome evolution revealed by comparative sequence analysis of orthologous regions from four triticeae genomes. Plant Physiol 135: 459–470

He P, Friebe BR, Gill BS, Zhou J-M (2003) Allopolyploidy alters gene expression in the highly stable hexaploid wheat. Plant Mol Biol 52:401–414

Hedden P (2003) The genes of the Green Revolution. Trends Genet 19:5–9

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

Huang S, Sirikhachornikit A, Su X, Faris J, Gill B, Haselkorn R, Gornicki P (2002) Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase ot the *Triticum/Aegilops* complex and the evolutionary history of polyploid wheat. Proc Natl Acad Sci USA 99:8133–8138

Ikeda A, Ueguchi-Tanaka M, Sonoda Y, Kitano H, Koshioka M, Futsuhara Y, Matsuoka M, Yamaguchi J (2001) *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 13: 999–1010

Jantasuriyarat C, Vales MI, Watson CJW, Riera-Lizarazu O (2004) Identification and mapping of genetic loci affecting the free-threshing habit and spike compactness in wheat. Theor Appl Genet 108:261–273

Kashkush K, Feldman M, Levy AA (2002) Gene loss, silencing and activation in a newly synthesized wheat allotetraploid. Genetics 160:1651–1659

Kato K, Miura H, Akiyama M, Kuroshima M, Sawada S (1998) RFLP mapping of the three major genes, *Vrn1*, *Q* and *B1*, on the long arm of chromosome 5A of wheat. Euphytica 101:91–95

Kato K, Miura H, Sawada S (2000) Mapping QTLs controlling grain yield and its compoenents on chromosome 5A of wheat. Theor Appl Genet 101:933–943

Kato K, Sonokawa R, Miura H, Sawada S (2003) Dwarfing effect associated with the threshability gene *Q* on wheat chromosome 5A. Plant Breed 122:489–492

Kerber E, Rowland GG (1974) Origin of the free threshing character in hexaploid wheat. Can J Genet Cytol 16:145–154

Khush GS (2001) Green revolution: the way forward. Nature Rev Genet 2:815–822

Kihara H (1944) Discovery of the DD-analyser, of the ancestors of *Triticum vulgare*. Agric Hort 19:13–14

Kilian B, Ozkan H, Deusch O, Effgen S, Brandolini A, Kohl J, Martin W, Salamini F (2006) Independent wheat B and G genome origins in outcrossing *Aegilops* progenitor haplotypes. Mol Biol Evol 24:217–227

Koebner R, Shepherd KW (1985) Induction of recombination between rye chromosome 1RL and wheat chromosomes. Theor Appl Genet 71:208–215

Kojima S, Takahashi Y, Kobayashi Y, Monna L, Sasaki T, Araki T, Yano M (2002) *Hd3* a rice ortholog of the *Arabidopsis FT* gene, promotes transition to flowering downstream of *Hd1* under short-day conditions. Plant Cell Physiol 43:1096–1105

Komatsuda T, Maxim P, Senthil N, Mano Y (2004) High-density AFLP map of *nonbrittle rachis 1* (*Btr1*) and 2 (*Btr2*) genes in barley (*Hordeum vulgare* L.). Theor Appl Genet 109:989–995

Kong X-Y, Gu YQ, You FM, Dubcovsky J, Anderson OD (2004) Dynamics of the evolution of orthologous and paralogous portions of a complex locus region in two genomes of allopolyploid wheat. Plant Mol Biol 54:55–69

Konishi S, Izawa T, Lin SY, Ebana K, Fukuta Y, Sasaki T, Yano M (2006) An SNP caused loss of seed shattering during rice domestication. Science 312:1392–1396

Kowalski S, Lan T-H, Feldamnn K, Paterson A (1994) Comparative mapping of *Arabidopsis thaliana* and *Brassica oleracea* chromosomes reveals islands of conserved gene order. Genetics 138: 499–510

Kuckuck H (1959) Neuere arbeiten zur entsheung der hexaploiden Kulturweize. Z Pflanzenzucht 41: 205–226

Law C, Worland AJ (1997) Genetic analysis of some flowering time and adaptive traits in wheat. New Phytol 137:19–28

Le Thierry D'Ennequin MLT, Toupance B, Robert T, Godele B, Gouyon P (1999) Plant domestication: a model for studying the selection of linkage. J Evol Biol 12:1138–1147

Leicht IJ, Bennett MD (1997) Polyploidy in angiosperms. Trends Plant Sci 2:470–475

Lelley T, Stachel M, Grausgruber H, Vollmann J (2000) Analysis of relationships between *Aegilops tauschii* and the D genome of wheat utilizing microsatellites. Genome 43:661–668

Levy AL, Feldman, M (2002) The impact of polyploidy on grass genome evolution. Plant Physiol 130:1587–1593

Li Z, Pinson SRM, Stansel JW, Park WD (1995) Identification of quantitative trait loci (QTLs) for heading date and plant height in cultivated rice (*Oryza sativa* L.). Theor Appl Genet 91: 374–381

Li W, Zhang P, Fellers JP, Friebe B, Gill BS (2004) Sequence composition, organization, and evolution of the core *Triticeae* genome. Plant J 40:500–511

Li C, Zhou A, Sang T (2006) Rice domestication by reducing shattering. Science 311:1936–1939

Lin Y-Y, Shertz KF, Paterson AH (1995) Comparative analysis of QTLs affecting plant height and maturity across the *Poaceae*, in reference to an interspecific sorghum population. Genetics 141: 391–411

Lubbers E, Gill KS, Cox TS, Gill BS (1991) Variation of molecular markers among geographically diverse accessions of *Triticum tauschii*. Genome 34:354–361

Luo M, Yang ZL, Kota RS, Dvorak J (2000) Recombination of chromosomes 3A(m) and 5A(m) of wheat: the distribution of recombination across chromosomes. Genetics 154:1301–1308

Luo M, Young ZL, Kawahara T, You F, Dvorak J (2006) The structure of wild and domesticated emmer wheat populations, gene flow between them, and the site of emmer domestication. Theor Appl Genet, in press

Marza F, Bai G-H, Carver BF, Zhou W-C (2006) Quantitative trait loci for yield and related traits in the wheat population Ning7840 × Clark. Theor Appl Genet 112:688–698

Mc Key J (1966) Species relationships in *Triticum*. Hereditas 2:237–276

McFadden E, Sears ER (1946) The origin of *Triticum spelta* and its free-threshing hexaploid relative. J Hered 37:81–89

McIntosh R, Hart GE Devos KM, Gale D, Rogers WJ (1998) Catalogue of gene symbols for wheat. Proc 9th Int Wheat Genet Symp 5:235

Mello-Sampayo T (1971) Genetic regulation of meiotic chromosome pairing by chromosome 3D of *Triticum aestivum*. Nat New Biol 230:23–24

Mikhailova E, Naranjo T, Shepherd K, Eden JW, Heyting C, de Jong JH (1998) The effect of the wheat *Ph1* locus on chromatin organization and meiotic chromosome pairing analysed by genome painting. Chromosoma 107:339–350

Morgante M, Salamini F (2003) From plant genomics to breeding practice. Curr Opin Biotechnol 14: 214–219

Mori N, Ishi T, Ishido T, Hirosawa S, Watatani H et al (2003) Origins of domesticated emmer and common wheat inferred from chloroplast DNA fingerprinting. In Pogna NE, Romano M, Pogna EA, Galtiero G (eds) 10th International Wheat Genetic Symposium. Istituto Sperimentale per la Cerealicultura Rome, Italy pp 25–28

Muramatsu M (1963) Dosage effect of the *spelta* gene *q* of hexaploid wheeat. Genetics 48:469–482

Nalam V, Vales MI, Watson CJW, Kianian SF, RieraLizarazu O (2006) Map-based analysis of genes affecting the brittle rachis character in tetraploid wheat (*Triticum turgidum*). Theor Appl Genet 112:373–381

Nesbitt M, Samuel D (1995) Promoting the conservation and use of underutilized and neglected crops. Paper presented at the First International. Workshop on hulled wheats, Castelvecchio Pascoli, Italy

Ozkan H, Levy AA, Feldman M (2001) Allopolyploidy-induced rapid genome evolution in the wheat (*Aegilops-Triticum*) group. Plant Cell 13:1735–1747

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

Ozkan H, Brandolini A, Pozzi C, Effeng S, Wunder J et al (2005) A reconsideration of the domestication geography of tetraploid wheat. Theor Appl Genet 110:1052–1060

Paterson AH (2002) What has QTL mapping taught us about plant domestication? New Phytol 154: 591–608

Paterson A, Lin Y-R, Li Z, Schertz KF, Doebley JF et al (1995) Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. Science 269:1714–1718

Paterson A, Bowers J, Burow M, Draye X, Eslik C, Jinag C, Katsar C, Lan T, LinY, Ming R, Wright R (2000) Comparative genomics of plant chromosomes. Plant Cell 12:1523–1539

Peng J, Carol P, Richards DE, King KE, Cowling RJ, Murphy GP, Harberd NP (1997) The Arabidopsis GAI gene defines a signalling pathway that negatively regulates gibberellin responses. Gen Dev 11:3194–3205

Peng J, Richards DE, Hartley NM, Murphy GP, Devos KM, Flintham JE, Beales J, Fish LJ et al (1999) "Green revolution" genes encode mutant gibberellin response modulators. Nature 400:256–261

Peng J, Ronin Y, Fahima T, Roder MS, Li Y, Nevo E, Korol A (2003) Domestication quantitative trait loci in *Triticum dicoccoides* the progenitor of wheat. Proc Natl Acad Sci USA 10:2489–2494

Peng JH, Zadeh H, Lazo GR, Gustafson JP, Chao S et al (2004) Chromosome bin map of expressed sequence tags in homoeologous group 1 of hexaploid wheat and homoeology with rice and *Arabidopsis.* Genetics 168:609–623

Perretant M, Cadalen T, Charmet G, Sourdille P, Nicolas P, Boeuf C et al (2000) QTL analysis of bread making quality in wheat using a doubled haploid population. Theor Appl Genet 100:1167–1175

Petersen G, Seberg O, Merete Y, Kasper B (2006) Phylogenetic relationships of *Triticum* and *Aegilops* and evidence for the origin of the A, B and D genomes of common wheat (*T. aestivum*). Mol Phylogenet Evol 39:70–82

Pozzi C, Rossini L, Vecchietti A, Salamini F (2004) Gene and genome changes during domestication of cereals. In: Gupta P, Varsheny RK (eds) Cereal genomics, Kluwer Academic Press, Netherlands

Provan J, Wolters P, Caldwell KH, Powell W (2004) High resolution organellar genome analysis of triticum and *Aegilops* sheds new light on cytoplasm evolution in wheat. Theor Appl Genet 108: 1182–1190

Rahman S, Jolly JC, Skerritt JH, Wallosheck A (1994) Cloning of a wheat 15 kDA grain softness protein (GSP). GSP is a mixture of puroindoline-like polypetides. Eur J Biochem 223:917–925

Rao M (1972) Mapping of the *compactum* gene *C* on chromosome 2D of wheat. Wheat Inf Serv 35:9

Rao M (1977) Mapping of the sphaerococcum gene "s" on chromosome 3D of wheat. Cereal Res Commun 5:15–17

Ravel C, Nagy IJ, Martre P, Sourdille P, Dardevet M, Balfourier F, Pont C, Giancola S, Praud S, Charmet G (2006) Single nucleotide polymorphism, genetic mapping, and expression of genes coding for the DOF wheat prolamin-box binding factor. Funct Integr Genomics 6:310–321

Reif J, Zhang P, Dreisigacker S, Warburton ML, van Ginkel M, Hoisington D, Bohn M, Melchinger AE (2005) Wheat genetic diversity trends during domestication and breeding. Theor Appl Genet 110:859–864

Richards, R (1988) A tiller inhibitor gene in wheat and its effect on plant growth. Aust J Agric Res 39:749–757

Rodriguez S, Maestra B, Perera B, Diez M, Naranjo T (2000) Pairing affinities of the B- and G-genome chromosomes of polyploid wheats with those of *Aegilops speltoides*. Genome 43:814–819

Rong J, Millet E, Manisterski J, Feldman M (2000) A new powdery mildew resistance gene: introgression from wild emmer into common wheat and RFLP-based mapping. Euphytica 115:121–126

Roussel V, Koenig J, Beckert M, Balfourier F (2004) Molecular diversity in French bread wheat accessions related to temporal trends and breeding programs. Theor Appl Genet 108:920–930

Salamini F (2003) Hormones and the green revolution. Science 302:71–72

Salamini F, Ozkan H, Brandolini A, Schaefer-Pregl R, Martin W (2002) Genetics and geography of wild cereal domestication in the near east. Nat Rev Genet 3:429–441

Sallares R, Brown TA (2004) Phylogenetic analysis of complete 5′ external transcribed spacers of the 18S ribosomal RNA genes of diploid *Aegilops* and related species (*Triticeae*, *Poaceae*). Genet Resour Crop Evol 51:701–712

Salvi S, Tuberosa R (2005) To clone or not to clone plant QTLs: present and future challenges. Trends Plant Sci 10:297–304

Sears E (1976) Genetic control of chromosome pairing in wheat. Annu Rev Genet 10:31–51

Shah M, Gill KS, Bezinger PS, Yen Y, Kaeppler SM, Ariyarathne HM (1999) Molecular mapping of loci for agronomic traits on chromosome 3A of bread wheat. Crop Sci 39:1728–1732

Shaked H, Kashkush K, Ozkan H, Feldman M, Levy AA (2001) Sequence elimination and cytosine methylation are rapid and reproducible responses of the genome to wide hybridiazion and allopolyploidy in wheat. Plant Cell 13:1749–1759

Sharma H, Waynes J (1980) Inheritance of tough rachis in crosses of *Triticum monococcum* and *Triticum boeoticum*. J Hered 7:214–216

Shindo C, Tsujimoto H, Sasakuma T (2003) Segregation analysis of heading traits in hexaploid wheat utilizing recombinant inbred lines. Heredity 90:56–93

Simonetti M, Bellomo MP, Laghetti G, Perrino P, Simeone R, Blanco A (1999) Quantitative trait loci influencing free-threshing habit in tetraploid wheats. Gen Res Crop Evol 46:267–271

Simons K, Fellers JP, Trik HN, Zhang Z, Tai Y-S, Gill BS, Faris JD (2006) Molecular characterization of the major wheat domestication gene *Q*. Genetics 172:547–555

Slageren MW (1994) Wild wheats: a monograph of *Aegilops* L. and *Amblyopyrum* (Jaub Abd Spach). Wageningen Agric Univ Press, Wageningen

Smale M, Reynolds MP, Warburton M, Skovmand B, Trethowan R, Singh RP, Ortiz-Monasterio I, Crossa J (2002) Dimensions of diversity in modern spring bread wheat in developing countries from 1965. Crop Sci 42:1766–1779

Snape J, Law W, Parker CN, Worland BB, Worland AJ (1985) Genetical analysis of chromosome 5A of wheat and its influence on important agronomic characters. Theor Appl Genet 71:518–526

Soltis PS (2005) Ancient and recent polyploidy in the angiosperms. New Phytol 166:5–8

Sourdille P, Perretnat MR, Charmet G, Leory P, Gautire MF, Joudrier P, Nelson JC, Sorrells ME, Bernard M (1996) Linkage between RFLP markers and genes affecting kernel hardness in wheat. Theor Appl Genet 93:580–586

Sourdille P, Tixier MH, Charmet G, Gay G, Cadalen T, Bernard S, Bernard M (2000) Location of genes involved in ear compactness in wheat (*Triticum aestivum*) by means of molecular markers. Mol Breed 6:247–255

Sourdille P, Cadalen T, Gay G, Gill B, Bernard M (2002) Molecular and physical mapping of genes affecting awning in wheat. Plant Breed 121:320–324

Spielmeyer W, Richards RA (2004) Comparative mapping of wheat chromosome 1AS which contains the tiller inhibition gene (*tin*) with rice chromosome 5S. Theor Appl Genet 109:1303–1310

Stelmack A (1990) Geographic distribution of *Vrn*-genes in landraces and improved varieties of spring bread wheat. Euphytica 45:113–118

Sutton T, Whitford R, Baumann Y, Dong C (2003) The *Ph2* pairing homoeologous locus of wheat (*Triticum aestivum*): identification of candidate meiotic genes using a comparative genetics approach. Plant J 36:443–456

Symes KJ (1965) The inheritance of grain hardness in wheat as measured by the particle size index. Aust J Agric Res 16:113–123

Tajima F (1989) Statistical methods to test for nucleotide mutation hypothesis by DNA polymorphism. Genetics 123:585–595

Talbert L, Smith LY, Blake NK (1998) More than one origin of hexaploid wheat is indicated by sequence comparison of low-copy DNA. Genome 41:402–407

Tanksley SD, McCouch SR (1997) Seed banks and molecular maps: unlocking genetic potential from the wild. Science 277:1063–1066

Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ES (2001) *Dwarf8* polymorphisms associate with variation in flowering time. Nat Genet 28:286–289

Thuillet AC, Bru D, David J, Roumet P, Santomi S, et al (2005) Direct estimation of mutation rate for 10 microsatellites loci in durum wheat, *T. turgidum*. Mol Biol Evol 19:122–125

Tranquilli G, Dubcovsky J (2000) Epistatic interactions between vernalization genes *VrnAm1* and *VrnAm2* in diploid wheat. J Hered 91:304–306

Turnbull K, Turner M, Mukai Y, Yamamoto M, Morell MK, Appels R, Rahman S (2003) The organization of genes tightly linked to the *Ha* locus in *Ae. tauschii*, the D genome donor of wheat. Genome 46:330–336

Vega J, Feldman M (1998) Effect of the pairing gene *Ph1* on centromere misdivision in common wheat. Genetics 148:1285–1294

Villareal R, Davila GF, Kazi AM (1995) Synthetic hexaploids *Triticum aestivum* advanced derivatives resistant to karnal bunt (*Tilletia indica* Mitra). Cereal Res Commun 23:127–132

Villareal R, Mujeeb-Kazi A, Rajaram S (1996) Inherintance of threshability in synthetic hexaploid by *T. aestivum* crosses. Plant Breed 115:407–409

Vision T, Brown D, Tanksley S (2000) The origins of genomic duplications in *Arabidopsis*. Science 290:2114–2117

Wanlong L, Gill BS (2006) Multiple genetic pathways for seed shattering in the grasses. Funct Integr Genomics 6:300–309

Watanabe N, Sogiyama K, Yamagashi Y, Skata Y (2002) Comparative telosomic mapping of homoeologous genes for brittle rachis in tetraploid and hexaploid wheats. Hereditas 137:180–185

Watanabe N, Takesada N, Fujii Y, Martinek P (2005) Comparative mapping of genes for brittle rachis in *Triticum* and *Aegilops*. Czech J Genet Plant Breed 41:39–44

Wendel J (2000) Genome evolution in polyploids. Plant Mol Biol 42:225–249

Wicker T, Yahiaoui N, Guyot R, Schlagenhauf E, Liu ZD, Dubcovsky J, Keller B (2003) Rapid genome divergence at orthologous low molecular weight glutenin loci of the A and A(m) genomes of wheat. Plant Cell 15:1186–1197

Worland A (1996) The influence of flowering time genes on environmental adaptability in European wheats. Euphytica 89:49–57

Wright SI, Gaut, BS (2005) Molecular population genetics and the search for adaptive evolution in plants. Mol Biol Evol 22:506–519

Wright SI, Vroh I, Schroeder SG, Yamasaki M, Doebley JF, McMullen MD, Gaut BS (2005) The effects of artificial selection on the maize genome. Science 308:1310–1314

Yan L, Loukoianov A, Tranquilli G, Helguera M, Fahima T, Dubcovsky J (2003) Positional cloning of the wheat vernalization gene *VRN1*. Proc Natl Acad Sci USA 100:6263–6268

Yan L, Helguera M, Kato K, Fukuyama S, Sherman J, Dubcosky J (2004) Allelic variation at the *VRN*-1 promoter region in polyploid wheat. Theor Appl Genet 109:1677–1686

Yan L, Loukoianov A, Blech A, Tranquilli G, Ramakrishna W, SanMiguel P, Bennetzen JL, Echenique V, Dubcovsky J (2006a) The wheat *VRN2* gene is a flowering repressor down-regulated by vernalization. Science 303:1640–1644

Yan L, Fu D, Li C, Blechl A, Tranquilli G, Bonafede M, Sanchez A, Valarik M, Yasuda S, Dubcovsky J (2006b) The wheat and barley vernalization gene *VRN3* is an orthologue of *FT*. Proc Natl Acad Sci USA 103:19581–19586

Yano M, Katayose Y, Ahikari M, Yamanouchi U, Monna L et al (2000) *Hd1*, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the *Arabidopsis* flowering time gene CONSTANS. Plant Cell 12:2473–2484

Zhang W, Qu L-J, Gu H, Gao W, Liu M, Chen J, Chen Z (2002) Studies on the origin and evolution of tetraploid wheats based on the internal transcribed spacer (ITS) sequences of nuclear ribosomal DNA. Theor Appl Genet 104:1099–1106

CHAPTER 18

TRANSCRIPTOME ANALYSIS OF THE SUGARCANE GENOME FOR CROP IMPROVEMENT

PAULO ARRUDA[1,2,*] AND THAÍS REZENDE SILVA[1,2]

[1]*Centro de Biologia Molecular e Engenharia Genética (CBMEG), Universidade Estadual de Campinas (UNICAMP), 13083-970, Campinas, SP, Brazil*

[2]*Departamento de Genética e Evolução, Instituto de Biologia, Universidade Estadual de Campinas (UNICAMP), 13083-970, Campinas, SP, Brazil*

Abstract: Sugarcane is being considered one the most important crops to meet the demand of the world bionergy needs. However the productivity in commercial plantations around the world is far way from its potential of about 300 tons/ha. Sugarcane breeding did not take advantage yet of the best plant breeding technologies, mainly because constrains imposed by the high polyploid nature of its genome. Even transgenic technologies would face difficulties because of the complex behaviour of introduced genes regarding the chromosome where the gene is inserted. One important resource to overcome at least part of these difficulties is the availability of a large collection of sugarcane Expressed Sequence Tags. The transcriptome information has allowed the identification of genes involved in biotic and abiotic stress response, disease resistance and sucrose accumulation. In additioncomparative mapping has allowed the identification in sugarcane disease resistance genes already mapped in sorghum and maize. In this chapter we discuss the use of transcriptome resources for sugarcane improvement.

1. INTRODUCTION

1.1. The Sugarcane Crop

Sugarcane is one of the most photosynthetically efficient cultivated crops, with a potential productivity of 300 tons of biomass per hectare under optimized agricultural conditions (Alexander, 1985). However, the average productivity of commercial plantations around the world is far below the full crop potential. Two factors affect sugarcane productivity, namely, the quality of agricultural

*Author for correspondence: parruda@unicamp.br

R.K. Varshney and R. Tuberosa (eds.), Genomics Assisted Crop Improvement: Vol. 2: Genomics Applications in Crops, 483–494.

management, ding fertilization conditions and irrigation, and the availability of genetically improved varieties able to withstand the limited growing conditions imposed by biotic and abiotic stress. With regard to genetic improvement, a major constraint is imposed by the highly polyploid nature of the sugarcane genome, which is not amenable to most of the breeding techniques developed for diploid species. Despite this limitation, improved varieties with yields of over 120 ton/ha on a few commercial plantations in some countries have been developed in the last 30 years (: http://www.fao.org/). However, such a high yield is not the reality for most of the 20 million hectares scattered throughout the 90 sugarcane-cultivating countries, for which the average productivity is around 40–70 tons/ha (: http://www.fao.org/).

The first cultivated sugarcane varieties were basically clones of *Saccharum officinarum*, a species with a high sugar content domesticated from the wild species *S. robustum*. The yield increased substantially a century ago when breeders crossed *S. officinarum* with *S. spontaneum*, a disease-resistant and vigorous relative (Berding and Roach, 1987). Yields were subsequently improved by backcrossing the hybrids to *S. officinarum* to produce the basic germplasm source of the modern sugarcane cultivars. Over the centuries, sugarcane has been cultivated almost solely as a source of sucrose, which accumulates at high concentrations in the stem internodes. Sucrose is extracted and purified in large mill factories, and is widely used as a raw material in the food industry. A major breakthrough in the last 30 years in Brazil has been the development of technology for the large-scale fermentation of sucrose to produce fuel ethanol. In the last few years, more than 17 billion liters of sugarcane ethanol have been used as car fuel each year, with an unprecedented environmental benefit in reducing CO_2 emission. Based on the Brazilian example, sugarcane is believed to become one of the most important crops for renewable energy production in the near future.

In this scenario, genomics will provide important contributions for improving the sugarcane crop. The most obvious genomic tool is the transcriptome data produced by sequencing Expressed Sequence Tags (ESTs). Large EST collections have been produced in recent years and are available in public databases (Carson and Botha, 2000; Arruda, 2001; Vettore et al. 2003; Casu et al. 2004). ESTs are the primary source material for accessing the expression of thousands of genes through technologies such as microarrays (Figure 1). However, to fully explore these tools it is necessary to understand the structure of the sugarcane genome.

1.2. The Sugarcane Genome

Sugarcane has a complex genomic architecture characterized by a high level of ploidy (Sreenivasan et al. 1987). The cultivated sugarcane varieties have 2n = 100 to 130 chromosomes derived from crosses between *S. officinarum* (2n = 80 chromosomes) and *S. spontaneum* (2n = 40 to 128 chromosomes) (Butterfield et al. 2001; D'Hont and Glaszmann, 2001). *Saccharum officinarum* has a basic chromosome number of x=10, whereas *S. spontaneum* has a basic chromosome number of x=8 (Grivet and Arruda, 2002). The different chromosome numbers of

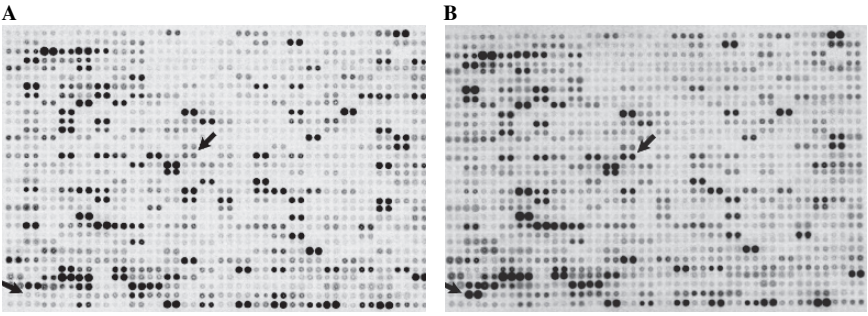


Figure 1. Example of a high-density membrane array used to analyze gene expression. Filters containing sugarcane cDNAs were probed with [^{33}P]cDNA reverse-transcribed from total RNA of control (A) and cold-treated (48 h at 4°C) (B) sugarcane plantlets. The arrows indicate examples of cold-inducible sugarcane ESTs

these two species allow the coexistence of distinct chromosomal organizations in the hybrid, with the possibility of some of the chromosomes being inherited entirely from *S. spontaneum* or from *S. officinarum* and some being recombination products between chromosomes of the two ancestral species (Grivet and Arruda, 2002). Despite the high level of ploidy, the monoploid genome sizes of *S. officinarum* and *S. spontaneum* are estimated to be around 930 Mbp and 750 Mbp, respectively, which are comparable to that of sorghum (~760 Mbp) and about twice that of rice (~390 Mbp). The modern sugarcane cultivars have around 120 chromosomes and a genome size of 10,000 Mb (D'Hont, 2005).

As in other grasses, the sugarcane genome contains an array of transposable elements (TEs), some of which are located between coding genes (Janoo et al. 2003). A large number of TEs is differentially expressed in sugarcane tissues and may affect the genomic stability (Araujo et al. 2005). These expressed TEs include homologs of the maize mutator (Mu) and the mudrA-like transposase. Sequences encoding the TEs found in the sugarcane EST project (SUCEST) database (Vettore et al. 2001) have also been found in *S. officinarum* and *S. spontaneum*, which suggests that these TEs were present in the ancestral Andropogoneae species common to sorghum, maize and sugarcane (Rossi et al. 2004a).

1.3. Synteny and Co-Linearity in Sugarcane and Other Grasses

The chromosome organization in grasses is highly conserved (Bennetzen and Freeling, 1993, 1997; Ming et al. 1998), possibly as a consequence of the short period that the species belonging to this family have evolved independently from a common ancestor. Comparative mapping has shown that the sugarcane and sorghum chromosomes are highly syntenic, which suggests that the order of the DNA sequences in their chromosomes has been conserved (Dufour et al. 1997; Glaszmann et al. 1997, Guimarães et al. 1997).

Sequence analysis of genomic DNA surrounding the alcohol dehydrogenase genes in sugarcane and sorghum has revealed a high colinearity in gene order, a conserved gene structure and a nucleotide similarity close to 95% (Janoo et al. 2003). These findings agree with those for comparative genetic mapping and reinforce the notion that sorghum has the simplest syntenic relationship with sugarcane, and makes sorghum the species of choice for modelling sugarcane (Dufour et al. 1997; Glaszmann et al. 1997; Guimarães et al. 1997; Ming et al. 1998; Asnaghi et al. 2000). Consequently, access to the sorghum genome and to the largest dataset of the sugarcane transcriptome will be of considerable use in the genetic improvement of sugarcane.

2. THE SUGARCANE TRANSCRIPTOME

2.1. Transcriptome Sequence Information

Large-scale EST sequencing has been used to quickly access gene sequence information. Currently, ESTs from over 1,000 species comprise the EST (dbEST) database at NCBI (http://www.ncbi.nlm.nih.gov/dbEST). Plants represent ~18% of all ESTs in the dbEST, with half of these sequences being from rice, wheat, maize, sugarcane, sorghum and barley. This huge amount of transcript information from grasses, especially sugarcane, sorghum and maize, provides a good opportunity for comparison of the expression profiles among these species. Traits associated with metabolism, development and responses to biotic and abiotic stress can be assessed by extensive comparative analyses of the gene expression among these related species.

The functional categorization of the transcriptome in these species has mapped the expression of genes implicated in major aspects of plant biology. When ESTs are derived from non-normalized cDNA libraries, as in the case of the SUCEST project (Vettore et al. 2001; http://sucest.lad.ic.unicamp.br/en/), a snapshot of the processes and pathways that are active in plant growth, development and metabolism can be inferred (Vettore et al. 2003). The expression analysis of SUCEST libraries by virtual northern blotting and array hybridization has shown that EST abundance can be used to assess transcript expression levels among tissues (Arruda, 2001; Nogueira et al. 2003; Vettore et al. 2003; de Rosa et al. 2005).

Analysis of large EST dataset depends on the accurate clustering of the set of messages arising from the same transcribed gene. This allows estimating the number and the expression profile of the genes in the collection. Clustering of sugarcane ESTs in the SUCEST project was performed using the CAP3 fragment assembler (Huang and Madan, 1999) for the whole set of ESTs (Telles and da Silva, 2001). The assembly produced was based on the result of the internal and external consistency tests, where the CAP3 assembly outperformed both the phrap-hs and phrap-d assemblies. This resulted in an overall EST clustering with low incidence of discrepant reads and few redundant clusters, *i.e.* two or more clusters that probably should be condensed to a single cluster.

2.2. Functional Profiling of the Sugarcane Transcriptome

The sugarcane EST project, SUCEST (Arruda, 2001), analyzed a collection of ~240,000 ESTs generated from 26 cDNA libraries constructed from different sugarcane organs/tissues sampled at various developmental stages (Vettore et al. 2001). These ESTs were assembled into contigs, denominated sugarcane assembled sequences (SAS), that allowed the targeting of over 30,000 sugarcane genes (Vettore et al. 2003). Comparison of the sugarcane transcriptome with monocots and dicots revealed that genes shared by these plants have been highly conserved, despite the long period of independent evolution (Vincentz et al. 2004). Thus, around 71% of the sugarcane and *Arabidopsis* proteins are similar, whereas 82% of the sugarcane proteins have a match in the rice genome. Likewise, approximately 11% of the SASs may correspond to genes restricted to monocots while around 18% of the SASs may represent proteins restricted to sugarcane (Vincentz et al. 2004).

Analysis of the sugarcane SUCEST dataset has revealed genes encoding a vast array of functions associated with useful agronomic traits (Vettore et al. 2003). SASs were functionally categorized and grouped into 18 broad categories of biological roles assigned according to their involvement in a cellular process or pathway. Almost 50% of all SASs was associated with five broad categories: cellular dynamics, stress response, protein metabolism, bioenergetics, and cellular communication/signal transduction. An "unable to classify" category comprising 17% of all categorized SASs correspond to genes whose roles were unknown or could not be assigned with confidence. Proteins were also categorized according to the presence and number of conserved domains. Most of the SASs identified in the SUCEST database coded for proteins containing one domain, but a significant number contained two or three domains. The most repeated domain found was the nucleoporin FG domain of the nuclear pore complex protein. The next most repeated domains were the leucine rich repeat (LRR) and the HEAT repeat (related to armadillo/_-catenin-like repeats), both of which are involved in protein–protein interactions. Also common were the M repeats, commonly found in bacterial M proteins, and the PPR repeat, of unknown function but frequently found in plants.

The SUCEST dataset allowed a preliminary analysis of the tissue-enriched expression profile. A SAS was considered tissue-enriched if it contained at least three ESTs originating exclusively from a single sugarcane tissue (Vettore et al. 2003). Around 2,8% of all SASs presented a tissue-enriched expression. Most of the SASs presenting tissue-enriched expression was found in developing seeds that contained 33% of the total tissue-enriched SASs. Among these SASs it were found six putative new genes with a high level of expression in developing seeds.

Another interesting feature of the SUCEST database was the presence of a large group of SASs encoding proteins containing the C3H-type zinc-finger domains C-x8-C-x5-C-x3-H, which are not as numerous in *Arabidopsis* (Riechmann et al. 2000). In general, 60% of the transcription factors found in *Arabidopsis* or rice were also present in the SUCEST database (Vettore et al. 2003).

Additional information was obtained by analyzing the expression profile of genes in plants submitted to biotic and/or abiotic stress. Figure 1 illustrates the use of

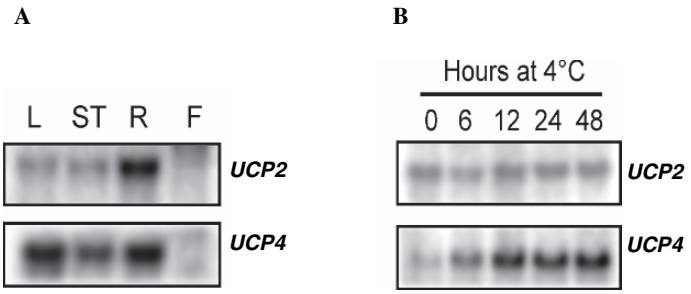


Figure 2. Accumulation of sugarcane UCP transcripts in plant organs. (A) RNA-blot analysis of *UCP2* and *UCP4* in control plants. F - flowers, L - leaves, ST- stem and R - roots. (B) Effect of exposure to cold on the accumulation of *UCP2* and *UCP4* transcripts

high-density membranes spotted with random sugarcane cDNA clones and probed with ^{33}P-labeled total RNA samples from plants submitted to cold-stress. Comparing the membranes hybridized with RNA probes from plants submitted to 4°C for 48 hrs with membranes hybridized with RNA probes from control plants it is possible to identify the cDNA clones that are up- or down-regulated by the cold treatment. The up-regulation can be further be confirmed by conventional RNA gel blot analysis as exemplified in Figure 2. Using this approach, the expression profiling of sugarcane in response to low temperature revealed 34 cold-inducible genes, including cellulose synthase, ABI3-interacting protein 2, a negative transcription regulator, a phosphate transporter, and several genes encoding proteins of unknown function (Nogueira et al. 2003). Of particular interest was the transcription factor known as SsNAC23. This protein belongs to the NAC family and, in addition to be strongly induced by chilling stress, is also induced by water stress and herbivory (Nogueira et al. 2005). SsNAC23 may be involved in a general responsive process that protects sugarcane plants against biotic and abiotic stress. This process may be linked to a general regulatory network that can be activated to protect plants against oxidative stress. This hypothesis was raised by the finding that genes encoding plant mitochondrial uncoupling proteins (UCPs) were also induced by exposure to cold (Figure 2) (Nogueira et al. 2003). UCPs are a family of membrane proteins that mediate purine nucleotide-sensitive free fatty acid-activated H^+ flux through the inner mitochondrial membrane (Vercesi et al. 2006). Five members of the UCP family were identified in sugarcane (Borecký et al. 2006). Tissue-enriched expression profiling revealed that these UCPs were ubiquitously expressed, although differential expression patterns were observed during chilling stress. UCPs may help sugarcane plants respond more flexibly to biotic and abiotic stress by modulating the electrochemical proton potential ($\Delta\mu H^+$) that is directly involved in regulating the production of reactive oxygen species (Vercesi et al. 2006). Other stress-associated genes have also been identified by expression profiling analysis of sugarcane plants treated with methyl jasmonate (Bower et al. 2005; de Rosa et al. 2005).

The reaction to stress involves signalling pathways that mediate short-to-middle term responses. Over 3,500 genes involved in several aspects of signal transduction, transcription, development, cell cycle, stress responses and interaction with pathogens were identified in the SUCEST database (Souza et al. 2001). Microarray expression profiling of these signalling components in field cultivated plants revealed the expression pattern of distinct signalling proteins in flowers, roots, leaves, lateral buds, and internodes (Papini-Terzi et al. 2005). These genes may be candidate targets for controlling traits associated with the responses to biotic and abiotic stress.

Other stress related proteins identified in the SUCEST database were phytocystatins, a class of cysteine proteinase inhibitors implicated in the endogenous regulation of protein turnover, programmed cell death, and in defense mechanisms against pathogens. Two cystatin proteins were identified in the SUCEST database and named CaneCPI-2 and CaneCPI-3. These canecystatins were shown to inhibit papain, a cysteine proteinase from Carica papaya and shown to be equally expressed in leaf, meristem and root. (Gianotti et al. 2006).

Other sequencing projects have contributed to unraveling the sugarcane transcriptome. Ma et al. (2004) sequenced 9,216 ESTs from three cDNA libraries (apex, leaf and mature internode). About 57% of these sequences have significant similarity to previously characterized proteins while 28% corresponded matched proteins with unknown function. Sequence comparisons of this sugarcane ESTs with sorghum ESTs revealed similar compositions of expressed genes between tissues. Further, some of this new set of sugarcane ESTs showed no matches in SUCEST database, suggesting possible transcriptome differences between these two projects that may be attributed to germplasm and/or environmental factors affecting gene expression.

2.3. Expression Profiling of Genes Related to Sugar Accumulation

Enhancing the accumulation of sucrose is one of the major priorities of sugarcane breeders. In many sugarcane varieties, sucrose can account for up to 50–60% of the stem dry weight (Bull and Glasziou, 1963), with the limiting step in sucrose accumulation occurring during the elongation phase. This phase involves complex metabolic and physiological processes partitioned along the internodes of the cane (Whittaker and Botha, 1997; Botha and Black, 2000). Sucrose accumulation also depends on processes involved in stem maturation (Moore, 1995; Walsh et al. 2005).

The pattern of gene expression associated with sugar metabolism has been assessed by analyzing the transcriptome expression profile of stems during the elongation and/or maturation phase (Casu et al. 2003, 2004, 2005). Sugar transporters and enzymes involved in sucrose metabolism were differentially expressed in stem internodes during the elongation phase. Genes associated with stem maturation were also identified, including enzymes involved in sucrose synthesis and cleavage, in glycolysis and in the pentose phosphate pathway (Casu et al. 2003). Genes

involved in fiber and lignin biosynthesis and stress-induced responses were also up-regulated during the elongation phase (Casu et al. 2003). Of particular interest was the finding of a sugar transporter designated as PST type 2a that was abundantly expressed in the maturing stem. PST type 2a may be involved in sugar translocation in the phloem companion cells and associated parenchyma in the maturing stem (Casu et al. 2003).

3. MOLECULAR MARKERS FROM THE TRANSCRIPTOME

3.1. ESTs as a Source for Detecting Polymorphisms

Polymorphism analysis of 178 cDNAs encoding alcohol dehydrogenases (Adh) in four sugarcane cultivars revealed 37 SNPs (single nucleotide polymorphisms) in the coding and untranslated regions of three Adh genes (Grivet et al. 2003). Overall analysis revealed one SNP occurring for every 122 base pairs in the coding regions whereas insertions and deletions were limited to untranslated regions. There is considerable variation in base compositions at SNP loci between sugarcane genotypes, which can be used to fingerprint and identify individual genotypes (Cordeiro et al. 2006). This base composition variation segregates in progeny of mapping populations. Single dose SNPs appear to occur in sugarcane ESTs in low frequency.

A survey of the SUCEST database revealed over 2,000 SASs containing dinucleotide, trinucleotide and tetranucleotide simple sequence repeats (SSRs) (Pinto et al. 2004). Twenty-three out of 30 tested SSRs produced scorable polymorphisms in 18 commercial sugarcane clones. Testing of these SSRs in F_1 individuals from a cross between two commercial sugarcane cultivars produced 52 segregating markers that were potentially useful for mapping important traits since these EST-derived SSRs encoded proteins involved in sugar biosynthesis (Pinto et al. 2004). EST-derived RFLP markers have also been developed to target the sugar content (Da Silva and Bressiani, 2005).

Although the data are still limited, the EST collection now available for sugarcane is very rich in information on sequence polymorphisms. Polymorphism revealed by ESTs is particularly easy to detect in genes that are highly expressed, i.e., those for which several tens of ESTs are available.

3.2. Genes Associated with Disease Resistance

Analysis of the SUCEST database revealed a vast array of genes encoding proteins similar to the NBARC domain and NBS/LRR, LRR or S/T KINASE domains which are characteristic of the major classes of disease resistance genes (known as resistance gene analogs or RGAs) (Rossi et al. 2003). Since RGAs have been associated with disease resistance loci or quantitative trait loci (QTL) (Graham et al. 2000; Wang et al. 2001), this is potentially useful for sugarcane since these RGAs could be used to search the sorghum genome for genes or QTLs associated with disease resistance.

Fifty sugarcane RGAs revealed restriction fragment length polymorphisms (RFLP) that could be located on the sugarcane reference genetic map. One cluster

of two LRR-like loci mapped close to the gene conferring resistance to common rust (Rossi et al. 2004b). Similar results were obtained by McIntyre et al. (2005), who identified RGAs linked to major disease-resistance loci in sugarcane. The mapping of 31 of these RGAs revealed a significant association with QTLs for brown rust resistance. The presence of a brown rust resistance gene (Bru1) in sugarcane was confirmed by analyzing segregation of the disease in a population of 658 individuals, derived from selfing of clone R570 (Asnagui et al. 2004). Several of these RGAs were also mapped in sorghum in a region previously shown to contain a major rust-resistance QTL (McIntyre et al. 2004).

Resistance genes frequently occur in clusters (Michelmore and Meyers, 1998). In sugarcane, 16 of the 50 mapped RGA loci are organized in four clusters containing 3–6 RGAs, while 12 are in pairs (Rossi et al. 2004b). Sequence comparisons of these RGAs with the respective references in rice or maize revealed that the members of a given sugarcane cluster were more similar to the alien reference (RPR1 or Rp1-D) than to members of the other sugarcane NBS/LRR locus. This finding suggested the existence of a common ancestral gene for rice RPR1 and the sugarcane RPR1-like cluster, and for maize Rp1-D and the sugarcane Rp1-D-like cluster (Rossi et al. 2004b).

4. CONCLUDING REMARKS

Improvements in sugarcane productivity have not accompanied the progress observed for other grasses. This can be attributed to the highly polyploid nature of its genome that makes sugarcane not amenable to most of the breeding techniques developed for diploid species. Despite this limitation, recent progress in sugarcane genomics has provided a better understanding of its evolutionary origin and genomic structure. An important achievement in this regard was the sequencing of a large collection of sugarcane ESTs. The creation of large sugarcane transcriptome databases has allowed the identification of genes involved in sucrose metabolism and accumulation, disease resistance and responses to biotic and abiotic stress. Because of the high synteny of sugarcane and sorghum, genes discovered and mapped in sorghum have been useful in identifying their orthologs in sugarcane. In addition, the expression profiling of thousands of genes in plants subjected to biotic and abiotic stress has shed light on how sugarcane responds and adapts to different conditions. However, exploration of the information contained in the sugarcane transcriptome is still at an initial stage. Since the complete genome sequence of sugarcane has yet to be determined, the large collection of ESTs remains one of the most valuable resources for gene discovery and functional analyses in sugarcane.

REFERENCES

Alexander AG (1985) The energy cane alternative. Elsevier, New York

Araujo PG, Rossi M, de Jesus EM, Saccaro NL Jr, Kajihara D, Massa R, de Felix JM, Drummond RD, Falco MC, Chabregas SM, Ulian EC, Menossi M, Van Sluys MA (2005) Transcriptionally active transposable elements in recent hybrid sugarcane. Plant J 44:707–717

Arruda P (2001) Sugarcane transcriptome: a landmark in plant genomics in the tropics. Genet Mol Biol 24:1–296

Asnaghi C, Paulete F, Kaye C, Grivet L, Deu M, Glaszmann J-C, D'Hont A (2000) Application of synteny across Poaceae to determine the map location of a sugarcane rust resistance gene. Theor Appl Genet 101:962–969

Asnaghi C, Roques D, Ruffel S, Kaye C, Hoarau JY, Telismart H, Girard JC, Raboin LM, Risterucci AM, Grivet L, D'Hont A (2004) Targeted mapping of a sugarcane rust resistance gene (*Bru1*) using bulked segregant analysis and AFLP markers. Theor Appl Genet 108:759–64

Bennetzen JL, Freeling M (1993) Grasses as a single genetic system: genome composition, collinearity and compatibility. Trends Genet 9:259–261

Bennetzen JL, Freeling M (1997) The unified grass genome: synergy in synteny. Genome Res 7:301–306

Berding N, Roach BT (1987) Germplasm collection, maintenance, and use. In: Heinz DJ (ed) Sugarcane improvement through breeding. Elsevier, New York, pp 143–210

Borecký J, Nogueira FTS, de Oliveira KAP, Maia IG, Vercesi AE, Arruda P (2006) The plant energy-dissipating mitochondrial systems: depicting the genomic structure and the expression profiles of the gene families of uncoupling protein and alternative oxidase in monocots and dicots. J Exp Bot 57:849–864

Botha FC, Black KG (2000) Sucrose phosphate synthase and sucrose synthase activity during maturation of internodal tissue in sugarcane. Aust J Plant Physiol 27:81–85

Bower NI, Casu RE, Maclean DJ, Reverter A, Chapmann SC, Manners JM (2005) Transcriptional response of sugarcane roots to methyl jasmonate. Plant Sci 168:761–772

Bull TA, Glasziou KT (1963) The evolutionary significance of sugar accumulation in *Saccharum.* Aust J Biol Sci 16:737–742

Butterfield MK, D'Hont A, Berding N (2001) The sugarcane genome: a synthesis of current understanding, and lessons for breeding and biotechnology. Proc S Afr Sug Technol Ass 75:1–5

Carson DL, Botha FC (2000) Preliminary analysis of expressed sequence tags for sugarcane. Crop Sci 40:1769–1779

Casu RE, Grof CPL, Rae AL, McIntyre CL, Dimmock CM, Manners JM (2003) Identification of a novel sugar transporter homologue strongly expressed in maturing stem vascular tissues of sugarcane by expressed sequence tag and microarray analysis. Plant Mol Biol 52:371–386

Casu RE, Dimmock CM, Chapman SC, Grof CPL, McIntyre CL, Bonnett GD, Manners JM (2004) Identification of differentially expressed transcripts from maturing stem of sugarcane by in silico analysis of stem expressed sequence tags and gene expression profiling. Plant Mol Biol 54:503–517

Casu RE, Manners JM, Bonnett GD, Jackson PA, McIntyre CL, Dunne R, Chapman SC, Rae AL, Grof CPL (2005) Genomics approaches for the identification of genes determining important traits in sugarcane. Field Crop Res 92:137–147

Cordeiro GM, Eliott F, McIntyre CL, Casu RE, Henry RJ (2006) Characterisation of single nucleotide polymorphisms in sugarcane ESTs. Theor Appl Genet 113:331–343

Da Silva JA, Bressiani JA (2005) Sucrose synthase molecular marker associated with sugar content in elite sugarcane progeny. Genet Mol Biol 28:294–298

de Rosa VE Jr, Nogueira FTS, Menossi M, Ulian EC, Arruda P (2005) Identification of methyl jasmonate-responsive genes in sugarcane using cDNA arrays. Braz J Plant Physiol 17:173–180

D'Hont A (2005) Unravelling the genome structure of polyploids using FISH and GISH; examples of sugarcane and banana. Cytogent Genomics Res 109:27–33

D'Hont A, Glaszmann JC (2001) Sugarcane genome analysis with molecular markers, a first decade of research. Proc Int Soc Sugarcane Technol 24:556–559

Dufour P, Deu M, Grivet L, D'Hont A, Paulet F, Bouet A, Lanaud C, Glaszmann JC, Hamon P (1997) Construction of a composite sorghum genome map and comparison with sugarcane, a related complex polyploid. Theor Appl Genet 94:409–418

Gianotti A, Rios WM, Soares-Costa A, Nogaroto V, Carmona AK, Oliva ML, Andrade SS, Henrique-Silva F (2006) Recombinant expression, purification, and functional analysis of two novel cystatins from sugarcane (*Saccharum officinarum*). Protein Expr Purif 47:483–489

Glaszmann JC, Dufour P, Grivet L, D'Hont A, Deu M, Paulet F, Hamon P (1997) Comparative genome analysis between several tropical grasses. Euphytica 96:13–21

Graham MA, Marek LF, Lohnes D, Cregan P, Schoemaker RC (2000) Expression and genome organization of resistance gene analogs in soybean. Genome 43:86–93

Grivet L, Arruda P (2002) Sugarcane genomics: depicting the complex genome of an important tropical crop. Curr Opin Plant Biol 5:122–127

Grivet L, Glaszmann JC, Vincentz M, da Silva FR, Arruda P (2003) ESTs as a source for sequence polymorphism discovery in sugarcane: example of the *Adh* genes. Theor Appl Genet 106:190–197

Guimarães CT, Sills GR, Sobral BWS (1997) Comparative mapping of Andropogoneae: *Saccharum* L. (sugarcane) and its relation to sorghum and maize. Proc Natl Acad Sci USA 94:14261–14266

Huang, X, Madan, A (1999). CAP3: a DNA sequence assembly program. Genome Res 9:868–877

Jannoo N, Grivet L, D'Hont A, Arruda P (2003) Genomic sequencing in sugarcane: first insight into the physical organization of the genome and microsynteny with other grasses. Abstracts of PAG XI Conference, San Diego, Abstract W182. URL: http://www.intl-pag.org// 11/abstracts/W26_W182_XI.html

Ma HM, Schulze S, Lee S, Yang M, Mirkov E, Irvine J, Moore P, Paterson A (2004) An EST survey of the sugarcane transcriptome. Theor Appl Genet 108:851–63

McIntyre CL (2004) Homologues of the maize rust resistance gene *Rp1-D* are genetically associated with a major rust resistance QTL in sorghum. Theor Appl Genet 109:875–883

McIntyre CL, Casu RE, Drenth J, Knight D, Whan VA, Croft BJ, Jordan DR, Manners JM (2005) Resistance gene analogues in sugarcane and sorghum and their association with quantitative trait loci for rust resistance. Genome 48:391–400

Michelmore RW, Meyers BC (1998) Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. Genome Res 8:1113–1130

Ming R et al (1998) Alignment of the *Sorghum* and *Saccharum* chromosomes: comparative genome organization and evolution of a polysomic polyploid genus and its diploid cousin. Genetics 150: 1663–1682

Moore P (1995) Temporal and spatial regulation of sucrose accumulation in the sugarcane stem. Aust J Plant Physiol 22:661–679

Nogueira FTS, de Rosa VE, Menossi M, Ulian EC, Arruda P (2003) RNA expression profiles and data mining of sugarcane response to low temperature. Plant Physiol 132:1811–1824

Nogueira FTS, Schlogl PS, Camargo SR, Fernandez JH, de Rosa VE, Pompermayer P, Arruda P (2005) SsNAC23, a member of the NAC domain protein family, is associated with cold, herbivory and water stress in sugarcane. Plant Sci 169:93–106

Papini-Terzi FS, Rocha FR, Vêncio RZN, Oliveira KC, Felix JM, Vicentini R, Rocha CS, Simões ACQ, Ulian EC, di Mauro SMZ, da Silva AM, Pereira CAB, Menossi M, Souza GM (2005) Transcription profiling of signal transduction-related genes in sugarcane tissues. DNA Res 12:27–38

Pinto LR, Oliveira KM, Ulian EC, Garcia AAF, de Souza AP (2004) Survey in the sugarcane expressed sequence tag database (SUCEST) for simple sequence repeats. Genome 47:795–804

Riechmann JL, Heard J, Martin G, Reuber L, Jiang C, Keddie J, Adam L, Pineda O, Ratcliffe OJ, Samaha RR et al (2000) *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. Science 290:2105–2110

Rossi M, Araujo PG, Paulet F, Garsmeur O, Dias VM, Chen H, Van Sluys MA, D'Hont A (2004b) Genomic distribution and characterization of EST-derived resistance gene analogs (RGAs) in sugarcane. Mol Genet Genomics 269:406–419

Rossi M, Araujo PG, de Jesus EM, Varani AM, Van Sluys MA (2004a) Comparative analysis of *Mutator*-like transposases in sugarcane. Mol Genet Genomics 272:194–203

Rossi M, Araujo PG, Paulet F, Garsmeur O, Dias V, Hui C, Van Sluys MA, D'Hont A (2003) Genome distribution and characterization of EST derived sugarcane resistance gene analogs. Mol Genet Genomics 269:406–419

Souza GM et al (2001) The sugarcane signal transduction (SUCAST) catalogue: prospecting signal transduction in sugarcane. Genet Mol Biol 24:25–34

Sreenivasan TV, Ahloowalia BS, Heinz DJ (1987) Cytogenetics. In: Heinz DJ (ed) Sugarcane improvement through breeding. Elsevier, New York, pp 211–253

Telles GP, da Silva FR (2001) Trimming and clustering sugarcane ESTs. Genet Mol Biol 24:17–23

Vercesi AE, Borecký J, Maia IG, Arruda P, Cuccovia IM, Chaimovich H (2006) Plant uncoupling mitochondrial proteins. Annu Rev Plant Biol 57:383–404

Vettore AL, da Silva FR, Kemper EL, Arruda P (2001) The libraries that made SUCEST. Genet Mol *Biol 24:1–7*

Vettore AL et al (2003) Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane. Genome Res 13:2725–2735

Vincentz M et al (2004) Evaluation of monocot and eudicot divergence using the sugarcane transcriptome. Plant Physiol 134:951–959

Walsh KB, Sky RB, Brown SM (2005) The anatomy of the pathway of sucrose unloading within the sugarcane stem. Funct Plant Biol 32:367–374

Wang M-B, Wesley SV, Finnegan EJ, Smith NA, Waterhouse PM (2001) Replicating satellite RNA induces sequence-specific DNA methylation and truncated transcripts in plants. RNA 7:16–28

Whittaker A, Botha FC (1997) Carbon partitioning during sucrose accumulation in sugarcane internodal tissue. Plant Physiol 115:1651–1659

APPENDIX I

LIST OF CONTRIBUTORS

Ramesh K. Aggarwal Centre for Cellular and Molecular Biology (CCMB), Uppal Road, Hyderabad– 500 007, India

Paulo Arruda Centro de Biologia Molecular e Engenharia Genética (CBMEG)/Departamento de Genética e Evolução, Universidade Estadual de Campinas (UNICAMP), 13083-970, Campinas, SP, Brazil

Rupakula Aruna International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru-502 324, India

Pere Arús IRTA (Institut de Recerca i Tecnologia Agroalimentàries), Carretera de Cabrils Km.2, E-08348 Cabrils, Spain

Michael Baum Integrated Gene Management Mega-Project, International Center for Agricultural Research in the Dry Areas, Aleppo, Syria

Sabhyata Bhatia National Institute for plant Genome Research (NIPGR), Aruna Asaf Ali Marg, New Delhi-110 067, India

Stéphane Bieri Institute of Plant Biology, University of Zürich, Zollikerstrasse 107, 8008 Zürich, Switzerland

Andreas Börner Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Correnstrasse 3, D–06466 Gatersleben, Germany

Eligio Bossolini Institute of Plant Biology, University of Zürich, Zollikerstrasse 107, 8008 Zürich, Switzerland

Edward S. Buckler Institute for Genomic Diversity, Cornell University, Ithaca, NY, 14853, USA USDA-ARS, Ithaca, NY, 14853, USA

Salvatore Ceccarelli Integrated Gene Management Mega-Project, International Center for Agricultural Research in the Dry Areas, Aleppo, Syria

Wafa Choumane Integrated Gene Management Mega-Project, International Center for Agricultural Research in the Dry Areas, Aleppo, Syria

Luca Comai University of California Davis Genome Center, Davis, California 95616, USA

Mark Cooper Pioneer Hi-Bred International. 7250 NW 62nd Ave, Johnston, IA 50131-0552, USA

Enzo DeAmbrogio Società Produttori Sementi Bologna, Divisione Ricerca, Argelato (BO), Italy

Rebecca W. Doerge Departments of Statistics and Agronomy, Purdue University; 150 North University Street, West Lafayette, IN 47907, USA

Gebisa Ejeta Department of Agronomy, Purdue University, 915 W. State Street, West Lafayette, IN 47907, USA

Elhan S. Ersoz Institute for Genomic Diversity, Cornell University, Ithaca, NY, 14853, USA

Catherine Feuillet UMR INRA-UBP 1095, Amélioration et Santé des Plantes, Domaine de Crouelle 234, Avenue du Brézet, 63100 Clermont-Ferrand, France

Vidal Fey Centre for Biotechnology, Tykistökatu 6, FIN-20521 Turku, Finland

Majid R. Foolad Department of Horticulture, The Pennsylvania State University, University Park, PA 16802, USA

Wolfgang Friedt Institute of Crop Science and Plant Breeding I, Justus-Liebig-University, Heinrich-Buff-Ring 26-32, 35392 Giessen, Germany

Martin W. Ganal TraitGenetics GmbH, Am Schwabeplan 1b, D-06466 Gatersleben, Germany

Susan Gardiner HortResearch, Private Bag 11030, Palmerston North, New Zealand

Pooran M. Gaur International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru-502 324, India

Marie-Véronique Gentil Laboratoire de Biologie des Ligneux et des Grandes Cultures, UPRES EA 1207, rue de Chartres. BP 6759. Faculté des sciences, Université d'Orléans. 45067 Orléans Cedex 2 France

M. V. Channabyre Gowda University of Agricultural Sciences (UAS), Dharwad-500 006, India

Silvana Grandillo CNR-IGV, Institute of Plant Genetics, Portici Research Division, Via Università 133, 80055 - Portici (NA), Italy

Stefania Grando Integrated Gene Management Mega-Project, International Center for Agricultural Research in the Dry Areas, Aleppo, Syria

Anil Grover Department of Plant Molecular Biology, University of Delhi South Campus, Benito Juarez Road, New Delhi 110 021, India

Peiguo Guo Integrated Gene Management Mega-Project, International Center for Agricultural Research in the Dry Areas, Aleppo, Syria College of Life Science, Guangzhou University, Guangzhou 510006, China

Prasad S. Hendre Centre for Cellular and Molecular Biology, Uppal Road, Hyderabad-500 007, India

Steven Henikoff Howard Hughes Medical Institute, 4000 Jones Bridge Road Chevy Chase, MD 20815-6789, USA

David A. Hoisington International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru-502 324, India

Mukesh Jain Interdisciplinary Centre for Plant Genomics and Department of Plant Molecular Biology, University of Delhi South Campus, Benito Juarez road, New Delhi-110021, India

Sini Junttila Centre for Biotechnology, Tykistökatu 6, FIN-20521 Turku, Finland

Beat Keller Institute of Plant Biology, University of Zürich, Zollikerstrasse 107, 8008 Zürich, Switzerland

Jitendra P. Khurana Interdisciplinary Centre for Plant Genomics and Department of Plant Molecular Biology, University of Delhi South Campus, Benito Juarez road, New Delhi-110021, India

Moon Young Kim Department of Plant Science/Research Institute for Agriculture and Life Sciences, Seoul National University, Seoul 151-921, The Republic of Korea

Matias Kirst School of Forest Resources and Conservation, University of Florida, PO Box 110410, Gainesville, FL 32611, USA

Joseph E. Knoll Department of Agronomy, Purdue University, 915 W. State Street, West Lafayette, IN 47907, USA

Maria von Korff Integrated Gene Management Mega-Project, International Center for Agricultural Research in the Dry Areas, Aleppo, Syria

Domenico Lafiandra Department of Agrobiology and Agrochemistry, University of Tuscia, Viterbo, Italy

Berhane Lakew Integrated Gene Management Mega-Project, International Center for Agricultural Research in the Dry Areas, Aleppo, Syria

Thomas Lübberstedt University of Århus, Faculty of Agricultural Sciences, Research Centre Flakkebjerg, 4200 Slagelse, Denmark

Michael Lee Iowa State University, Ames, IA 50011, Iowa USA

Suk-Ha Lee Department of Plant Science/Research Institute for Agriculture and Life Sciences, Seoul National University, Seoul 151-921, The Republic of Korea

Kean Jin Lim Centre for Biotechnology, Tykistökatu 6, FIN-20521 Turku, Finland

Lang Luo Pioneer Hi-Bred International. 7250 NW 62nd Ave, Johnston, IA 50131-0552, USA

Marco Maccaferri Department of Agroenvironmental Sciences and Technology, University of Bologna, Viale Fanin 44, 40127 Bologna, Italy

David J. Mackill International Rice Research Institute, DAPO Box 7777, Metro Manila, Philippines

Thudi Mahender International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru- 502324, India

Hideo Matsumura Iwate Biotechnology Research Center, 22-174-4, Narita, Kitakami, Iwate 024-0003, Japan

Stéphane Maury Laboratoire de Biologie des Ligneux et des Grandes Cultures, UPRES EA 1207, rue de Chartres. BP 6759. Faculté des sciences, Université d'Orléans. 45067 Orléans Cedex 2 France

Trilochan Mohapatra National Research Centre on Plant Biotechnology (NRCPB), Indian Agricultural Research Institute (IARI), New Delhi-110 012, India

Henry T. Nguyen National Center for Soybean Biotechnology and Division of Plant Sciences, University of Missouri, Columbia, MO 65211, USA

Shyam N. Nigam International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru-502 324, India

Frank Ordon Institute of Epidemiology and Resistance Resources, Federal Centre for Breeding Research on Cultivated Plants, Erwin-Baur-Str.37 06484 Quedlinburg

Ashwani Pareek Stress Physiology and Molecular Biology Laboratory, School of Life Sciences, Jawaharlal Nehru University, New Delhi 110067, India

Md S. Pathan National Center for Soybean Biotechnology and Division of Plant Sciences, University of Missouri, Columbia, MO 65211, USA

Johan Peleman Nunhems B.V., P.O. Box 4005, 6080 AA Haelen, The Netherlands

Dean W. Podlich Pioneer Hi-Bred International. 7250 NW 62nd Ave, Johnston, IA 50131-0552, USA

Carlo Pozzi Fondazione Parco Tecnologico Padano, via Einstein Loc. C.na Codazza, 26900 Lodi, Italy

Thaís Rezende Silva Centro de Biologia Molecular e Engenharia Genética (CBMEG)/Departamento de Genética e Evolução, Universidade Estadual de Campinas (UNICAMP), 13083-970, Campinas, SP, Brazil

Marion S. Röder Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Corrensstr. 3, D-06466 Gatersleben, Germany

Stephen Rudd Centre for Biotechnology, Tykistökatu 6, FIN-20521 Turku, Finland

Francesco Salamini Fondazione Parco Tecnologico Padano, via Einstein Loc. C.na Codazza, 26900 Lodi, Italy Università degli Studi di Milano, Dip. Produzione Vegetale, via Celoria 2, 20133 Milano, Italy

Jérôme Salse UMR INRA-UBP 1095, Amélioration et Santé des Plantes, Domaine de Crouelle 234, Avenue du Brézet, 63100 Clermont-Ferrand, France

Silvio Salvi Department of Agroenvironmental Sciences and Technology, University of Bologna, Viale Fanin 44, 40127 Bologna, Italy

Maria Corinna Sanguineti Department of Agroenvironmental Sciences and Technology, University of Bologna, Viale Fanin 44, 40127 Bologna, Italy

Kulbhushan Saxena International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru-502 324, India

Haitham Sayed Integrated Gene Management Mega-Project, International Center for Agricultural Research in the Dry Areas, Aleppo, Syria

Niroj K. Sethy National Institute for Plant Genome Research (NCPGR), Asaf Ali Marg, New Delhi-110 067, India Present address: Defence Institute of Physiology and Allied Sciences (DIPAS), Defence Research and Development Organization (DRDO), Timarpur, Delhi-110 054, India

Prakash C. Sharma University School of Biotechnology, Guru Gobind Singh Indraprastha University, Delhi 110006, India

Nagendra K. Singh National Research Centre on Plant Biotechnology (NRCPB), Indian Agricultural Research Institute (IARI), New Delhi-110 012, India

Sneh L. Singla-Pareek Plant Molecular Biology, International Centre for Genetic Engineering and Biotechnology, New Delhi 100067, India

Sudhir K. Sopory Plant Molecular Biology, International Centre for Genetic Engineering and Biotechnology, New Delhi 100067, India

Anker P. Sørensen Keygene N.V., P.O. Box 216, 6700 AE, Wageningen, The Netherlands

Jeroen Stuurman Keygene N.V., P.O. Box 216, 6700 AE, Wageningen, The Netherlands

Steve D. Tanksley Department of Plant Breeding and Department of Plant Biology, Cornell University, Ithaca, NY 14853, USA

Ryohei Terauchi Iwate Biotechnology Research Center, 22-174-4, Narita, Kitakami, Iwate 024-0003, Japan

Bradley J. Till Basic Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA

Roberto Tuberosa Department of Agroenvironmental Sciences and Technology, University of Bologna, Viale Fanin 44, 40127 Bologna, Italy

Akhilesh K. Tyagi Interdisciplinary Centre for Plant Genomics and Department of Plant Molecular Biology, University of Delhi South Campus, Benito Juarez road, New Delhi-110021, India

Sripada M. Udupa Integrated Gene Management Mega-Project, International Center for Agricultural Research in the Dry Areas, Aleppo, Syria

Hari D. Upadhyaya International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru-502 324, India

Vincent Vadez International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru-502 324, India

Babu Valliyodan National Center for Soybean Biotechnology and Division of Plant Sciences, University of Missouri, Columbia, MO 65211, USA

Kyujung Van Department of Plant Science, Seoul National University, Seoul 151-921, The Republic of Korea

Jeroen Rouppe van der Voort Enza Zaden B.V, P.O. Box 7, 1600 AA Enkhnizen, The Netherlands

Rajeev K. Varshney International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru-502 324, India

Tri D. Vuong National Center for Soybean Biotechnology and Division of Plant Sciences, University of Missouri, Columbia, MO 65211, USA

Xiaolei Wu National Center for Soybean Biotechnology and Division of Plant Sciences, University of Missouri, Columbia, MO 65211, USA

Nabila Yahiaoui Institute of Plant Biology, University of Zürich, Zollikerstrasse 107, 8008 Zürich, Switzerland

Jianming Yu Institute for Genomic Diversity, Cornell University, Ithaca, NY, 14853, USA

Qibin Yu School of Forest Resources and Conservation, University of Florida, PO Box 110410, Gainesville, FL 32611, USA

Dani Zamir Faculty of Agriculture, The Hebrew University of Jerusalem, PO Box 12, Rehovot 76100, Israel

APPENDIX II

LIST OF REVIEWERS

Amalia Barone University of Naples "Federico II", Portici, Italy

Alain Charcosset UMR, INRA UPS-XI INAPG CNRS Génétique Végétale, Gif-sur-Yvette, France

Ayman A. Diab Agricultural Genetic Engineering Research Institute (AGERI), Modern Science and Art University (MSA), Egypt

Ivo Grosse Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany

B Jayashree International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, India

Guenter Kahl University of Frankfurt, Frankfurt am Main, Germany

Beat Keller University of Zurich, Switzerland

Takao Komatsuda National Institute of Agrobiological Sciences, Tsukuba, Japan

Thomas Lübberstedt Árhus University, Research Centre Flakkebjerg, Slagelse, Denmark

Jianxin Ma Purdue University, West Lafayette, USA

Tudi Mahender International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, India

Henry Nguyen National Centre for Soybean Research, University of Missouri, Columbia, USA

Antoni Rafalski DuPont Co. Crop Genetics and University of Delaware, Newark, USA

Saurabh Raghuvanshi University of Delhi South Campus, New Delhi, India

Ajoy K Roy Indian Grassland and Fodder Research Institute (IGFRI), Jhansi, India

Joy K Roy University of Minnesota, Minneapolis, USA

Nese Sreenivasulu Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben, Germany

Brian Steffenson University of Minnesota, Minneapolis, USA

WT Bill Thomas SCRI, Dundee, Scotland

Roberto Tuberosa University of Bologna, Italy

Vincent Vadez International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, India

Rajeev K Varshney International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, India

Desh Pal S. Verma The Ohio State University, Columbus, USA

San Ming Wang Center for Functional Genomics, Northwestern University, Evanston, USA

Clifford Weil Purdue University, West Lafayette, USA

Ramakrishna Wusirika Michigan Technological University, Houghton, USA

INDEX