How can we use genomics to improve cereals with rice as a reference genome?

Yunbi Xu[1], Susan R. McCouch[1] and Qifa Zhang[2,*]
[1]*Department of Plant Breeding, Cornell University, Ithaca, NY 14853-1901, USA;* [2]*National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, China(*author for correspondence: e-mail qifazh@mail.hzau.edu.cn)*

Abstract

Rice serves as a model crop for cereal genomics. The availability of complete genome sequences, together with various genomic resources available for both rice and *Arabidopsis*, have revolutionized our understanding of the genetic make-up of crop plants. Both macrocolinearity revealed by comparative mapping and microcolinearity revealed by sequence comparisons among the grasses indicate that sequencing and functional analysis of the rice genome will have a significant impact on other cereals in terms of both genomic studies and crop improvement. The availability of mutants, introgression libraries, and advanced transformation techniques make functional genomics in rice and other cereals more manageable than ever before. A wide array of genetic markers, including anchor markers for comparative mapping, SSRs and SNPs are widely used in genetic mapping, germplasm evaluation and marker assisted selection. An integrated database that combines genome information for rice and other cereals is key to the effective utilization of all genomics resources for cereal improvement. To maximize the potential of genomics for plant breeding, experiments must be further miniaturized and costs must be reduced. Many techniques, including targeted gene disruption or allele substitution, insertional mutagenesis, RNA interference and homologous recombination, need to be refined before they can be widely used in functional genomic analysis and plant breeding.

Abbreviations: AFLP, amplified fragment length polymorphism; BAC, bacterial artificial chromosome; COS, conserved orthologous sequence; DHPLC, denaturing high-pressure liquid chromatography; EST, expressed sequence tag; FM, functional marker; MAS, marker assisted selection; ORF, open reading frame; QTL, quantitative trait locus or loci; RM, random marker; RNAi, RNA interference; SAGE, serial analysis of gene expression; SNP, single nucleotide polymorphism; SSR, simple sequence repeat; TILLING, targeting induced local lesions in genomes

Introduction

Cereals serve as the main source of dietary calories for most human populations, either consumed directly, as is rice, or indirectly by way of animal feed. It is believed that the cereal species shared a common ancestor about 50–70 million years ago (Kellogg, 1998). In the past 5–10 thousand years, these species have been domesticated and appear to have experienced similar selection regimes associated with genetic improvement. The genomics revolution of the past decade has greatly improved our understanding of the genetic make up of a wide array of living organisms, including several plant species. Complete genome sequences of *Arabidopsis* (The *Arabidopsis* Genome Initiative,

2000) and rice (Feng *et al.*, 2002; Goff *et al.*, 2002; Sasaki *et al.*, 2002; Yu *et al.*, 2002; The Rice Chromosome 10 Sequencing Consortium, 2003), together with high-throughput technology for the analyses of transcripts, proteins, and mutants, provide the basis for understanding the relationship between genes, proteins and phenotypes. Characterization of plant genomes and the genes contained within them will aid geneticists and molecular biologists in their quest to understand cereal biology and help plant breeders in their goal of developing better products.

The rice genome sequence provides a platform for organizing information about diverse cereals, and together with genetic maps and partial sequence from other cereals is yielding new insights into both the shared and the independent dimensions of cereal evolution. Sequence capturing techniques promise to accelerate gene discovery in many large-genome cereals, and to better link the under-explored genomes of 'orphan' cereals with state-of-the-art knowledge (Paterson *et al.*, 2003). Genomics-based approaches are used to identify genes that have been involved in cereal improvement. Rice genome information will provide new innovations in rice research, as well as new knowledge, tools and opportunities for plant genome biology in other cereals. This article discusses how we can improve cereal crops using genomics approaches with rice as a reference genome.

Rice as reference genome for cereals

Rice was selected as a model for genome research in the cereals because of its small genome size (estimated at about 430 Mb) and global importance as a food crop. Today, it serves as a reference genome for molecular biological studies in the grasses for the following reasons: (1) availability of complete genome sequences of two divergent rice cultivars representing both the *indica* and *japonica* subspecies (Goff *et al.*, 2002; Yu *et al.*, 2002); (2) demonstrated conservation of gene content and relative conservation of gene order among the cereals; (3) large numbers of cultivated and wild germplasm resources and genetic stocks, including permanent mapping populations, introgression/ substitution lines, near-isogenic lines, and genetic mutants (i.e., knock-out and activation tagged lines) (Jeon *et al.*, 2000; Leung *et al.*, 2001; Xue

and Xu, 2002; An *et al.*, 2003; Hirochika, 2003; Hirochika *et al.*, 2004; Kolesnik *et al.*, 2004; Ryu *et al.*, 2004; Sallaud *et al.*, 2004; www.gramene. org; http://www.rgrc.dna.affrc.go.jp/index.html.en; http://www.icis.cgiar.org); (4) the availability of thousands of expressed sequence tags (ESTs) (http://www.ncbi. nlm.nih.gov/dbEST), full-length cDNAs (The Rice Full-Length cDNA Consortium, 2003), DNA chips (Lan *et al.,* 2004; http:// www.chem.aglient.com/scripts/PDS.asp?/page= 12133) and expression profiling techniques; (5) straight-forward transformation techniques (Hiei *et al.*, 1994; Komari *et al.*, 1998); (6) access to several genome databases that facilitate depositing, searching, querying and analyzing information about rice and other cereals in a comparative context (www.gramene.org; http://www. tigr.org.tdb/e2kl/osal/; http://rgp.dna.affrc.go.jp/; http://www.iris.irri.org/IRFGC/resources.shtml); (7) the exis-tence of a large international research community that is ready and able to make use of advances in genomic research.

Whole genome sequencing of rice, along with the extensive genetic and physical mapping efforts that preceded it provide a foundation for organizing information about diverse cereals, identifying orthologous genes, facilitating the genome sequencing of other cereals, and yielding new insights into cereal evolutionary history. The rice genome sequence provides a complete index of rice gene sequences but it does not tell us what the functions of these genes are or how they interact to give rise to desirable traits in cereal crops. Knowledge of the full set of genes in rice will permit comprehensive inspection of the gene complement in rice and related species to see which pathways are shared and which are unique, and how these pathways may have been modified. Using available rice databases to obtain a complete set of predicted and known peptides, those that are of most interest for further characterization can be identified. Finally, rice sequence information can be used to mine the rice genome and to understand how gene families are created, amplified, selectively eliminated, and how they diverge to create new biological activities and specificities.

In addition to the whole genome sequence, the availability of more than 300 000 public ESTs and a large set of full-length cDNA clones facilitates annotation of the rice genome. Information about

full-length cDNAs (The Rice Full-Length cDNA Consortium, 2003) will be extremely important for functional genomics and proteomics of rice. In addition, because little full-length cDNA data are yet available for other cereal genomes, it will have a great impact on future studies of plant genomics in general. Full-length cDNA clones are necessary to identify intron–exon boundaries, alternative splice sites and to clearly define gene-coding regions within the genomic sequence, as well as for comprehensive gene-function analyses at the transcriptional and translational levels. The availability of ESTs from a diverse set of cDNA libraries provides information on transcript abundance, tissue location, and developmental expression of genes.

Recently, an *Oryza* Map Alignment Project was initiated to produce comparative maps involving *Oryza sativa* and 12 of its wild *Oryza* relatives (www.omap.org). This effort represents the first attempt to develop a closed experimental system aimed at understanding the evolution, physiology and biochemistry of a single genus of plants or animals. Specifically, the project has constructed bacteria articial chromosome (BAC) libraries from 12 wild genomes, including diploids and tetraploids, that will be fingerprinted, and BAC-end sequenced and aligned to domesticated rice (*O. sativa*). Data will be integrated with other genomic resources and information through the Gramene database (www.gramene.org) for availability to the research community (Wing *et al.*, 2004). This information will provide genome-wide comparisons among *Oryza* species and will serve as an example of genus genomics for plants and animals.

Colinearity among rice and other cereals

Macrocolinearity

Significant genomic colinearity in plants has been revealed by comparative genetic mapping and genome sequencing, although plant genomes vary tremendously in genome size, chromosome number, and chromosome morphology (Devos and Gale, 2000; McCouch, 2001; Schmidt, 2002; Shimamoto and Kyozuka, 2002). Comparative mapping of cereal genomes using low copy number, cross-hybridizing genetic markers has provided compelling evidence for a high level of conservation of gene order across regions spanning many megabases (i.e., macrocolinearity). Initial studies of the organization of grass genomes indicated that individual rice chromosomes were largely collinear with those of several other grass species, and extensive work over the past decade has shown a remarkable conservation of large segments of linkage groups within rice, maize, sorghum, barley, wheat, rye, sugarcane, and other agriculturally important grasses (e.g., Ahn and Tanksley, 1993; van Deynze *et al.*, 1995; Gale and Devos, 1998; Wilson *et al.*, 1999). These studies led to the prediction that grasses could be studied as a single syntenic system. The macrocolinearity was originally summarized by Moore *et al.* (1995) for rice and several other cereals using what is now known as the 'Circle Diagram'. Further studies identified quantitative trait loci (QTL) controlling important agronomic traits that showed similarities in locations for the same or similar traits (i.e., Fatokun *et al.*, 1992; Lin *et al.*, 1995; Xiao *et al.*, 1996; for a review, see Xu, 1997). Shattering and plant height are examples that were also mapped to collinear regions among grass genomes (Paterson *et al.*, 1995; Peng *et al.*, 1999). More recently, Chen *et al.* (2003) identified four QTL for quantitative resistance to rice blast that showed corresponding map positions between rice and barley, two of which had completely conserved isolate specificity and the other two had partially conserved isolate specificity. Such corresponding locations and conserved specificity suggested a common origin and conserved functionality of the genes underlying the QTL for quantitative disease resistance. This suggests that comparative mapping may be used to discover genes, understand their function, and to help identify the evolutionary forces that determine the structure and the organization of grass genomes. Such findings have reinforced the utility of comparative mapping for understanding colinearity among cereal genomes.

The concept of a unified grass genome as a model for studying individual species has had a substantial impact upon plant biology, but has not yet lived up to its potential. There are some difficulties in evaluating synteny between genomes at the macro-level. First, genomic marker data are very incomplete, and genome sequence data are still largely lacking for grass species other than rice. Second, the data are sometimes biased

because the DNA probes used in comparative mapping are selected for polymorphism and single-copy hybridization patterns. Third, many genes are members of gene families and accordingly, it is often difficult to determine if a gene mapped in the second species is orthologous or paralogous to that in the first species. Fourth, the colinearity of gene order and content observed at the recombinational map level is often not observed at the level of local genome structure (Bennetzen and Ramakrishna, 2002; Feuillet and Keller, 2002). Finally, in most early studies, no statistical analysis was used to evaluate whether the presence of a few markers in the same order on two chromosomal segments in two species occurs by chance or is truly significant, although several computer programs, such as LineUp (Hampson et al., 2003) or ADHoRE (Vandepoele et al., 2002), have been developed to align genomes and to measure synteny.

Microcolinearity

Using the rice genome sequence as the reference to compare with molecular marker information of other cereals pointed to many more rearrangements than had been expected from Moore et al. (1995) and Gale and Devos (1998)'s concentric circles model. One of such comparisons involved more than 2600 mapped sequenced markers in maize among which only 656 putative orthologous genes could be identified (Salse et al., 2004). The comparison of the wheat genetic map with the rice sequence also points to numerous rearrangements between the two genomes, with a high frequency of breakdowns in colinearity (Sorrells et al., 2003). Extensive comparisons have also been made between sorghum and rice (Klein et al., 2003; The Rice Chromosome 10 Sequencing Consortium, 2003). An interesting new approach has been developed to align the sorghum physical map with the rice map (Klein et al., 2003). Sorghum BAC clones were selected from the minimum tiling path of chromosome 3. Unique partial sequences were obtained from each BAC clone and could be directly compared with the rice sequence. This approach revealed excellent conservation between the overall structure and gene order of sorghum chromosome 3 and rice chromosome 1 but also indicated several rearrangements: 50 of the 118 BACs examined did not show any sequence

similarity and five BACs showed better colinearity with rice BACs located elsewhere in the rice genome. Sixty-three BACs are collinear with rice BACs from chromosome 1, but four of them do not respect the rice minimum tiling path and therefore correspond to different positions in the two homologous chromosomes. Together, these studies point to a general conservation of large syntenic blocks within cereals, but with many more rearrangements and synteny breakdowns within these blocks than originally anticipated.

This trend is even more obvious when synteny is analyzed at the sequence level. The rearrangements that involve regions smaller than a few cM may occur and would be missed by most recombinational mapping studies. Comparative sequence analysis involving large genomic segments can detect these rearrangements. Such analyses reveal the composition, organization, and functional components of genomes and provide insight into regional differences in composition between related species. Recently, the sequencing of large genomic segments in numerous cereals made it possible to evaluate microcolinearity among genes or gene clusters. Sequencing of the domestication locus Q in *Triticum monococcum* revealed excellent colinearity with the bread wheat genetic map (Faris et al., 2003). Positional cloning of wheat vernalization genes *VRN1* and *VRN2* confirmed a high degree of microcolinearity in these two gene regions among cereals (Yan et al., 2003, 2004). Sequencing of the leaf-rust-resistance locus *Rph7* from barley demonstrated a more disrupted pattern. In barley, this locus is immediately flanked by two *HGA* genes. The orthologous locus on rice chromosome 1 contains five *HGA* genes. Looking again at barley, four of the five *HGA* genes are present, one is duplicated as a pseudogene and six additional genes have been inserted in between the *HGA* genes. These six genes have homologues on eight different rice chromosomes (Brunner et al., 2003). Striking rearrangements were also revealed by the comparison of 100 kb around the *Bronze* locus of two maize lines. Not only does the retrotransposon distribution differ between the two lines but the genes themselves could also be different (Fu and Dooner, 2002). Comparison of the low molecular weight glutenin locus between *Triticum monococcum* and *Triticum durum* also revealed dramatic rearrangements: more than 90% of the sequence diverged because of retroelement

insertions and because different genes are present at this locus (Wicker *et al.*, 2003). Therefore colinearity can be lost very rapidly, even within two genomes from the same species.

With the sequencing of long regions, several recent studies in cereals have demonstrated incomplete microcolinearity at the sequence level (Tarchini *et al.*, 2000; Dubcovsky *et al.*, 2001; SanMiguel *et al.*, 2002; Song *et al.*, 2002; Tikhonov *et al.*, 2000). Song *et al.* (2002) identified orthologous regions from maize, sorghum, and two subspecies of rice. It was found that gross macrocolinearity is maintained, but microcolinearity is incomplete among these cereals. Deviations from gene colinearity are attributable to micro-rearrangement or small-scale genomic changes, such as gene insertions, deletions, duplications, or inversions. In the region under study, the orthologous region was found to contain six genes in rice, 15 in sorghum, and 13 in maize. In maize and sorghum, gene amplification caused a local expansion of conserved genes but did not disrupt their order or orientation. As indicated by Bennetzen and Ma (2003), numerous local rearrangements differentiate the structures of different cereal genomes. On average, any comparison of a ten-gene segment between rice and a distant grass relative such as barley, maize, sorghum or wheat shows one or two rearrangements that involve genes. A simple extrapolation to the rice genome of about 40 000 genes (Goff *et al.*, 2002) suggests that about 6000 genic rearrangements occurred that differentiate rice from any of the other cereals. Most of these rearrangements appear to be tiny and thus would not interfere with the macrocolinearity observed by recombinational mapping. There are exceptions, however, which include chromosomal arm translocations and movements of single genes to different chromosomes (Bennetzen and Ma, 2003).

As expected, there is a high degree of gene conservation between the two shotgun-sequenced subspecies of rice, *japonica* and *indica*, which appeared to diverged about 0.44 mya (Ma and Bennetzen, 2004). On careful inspection, however, narrow regions of divergence can be found in these genomes (Song *et al.*, 2002; Han and Xue, 2003). These regions appear to correspond to areas of increased divergence among rice, sorghum and maize, suggesting that the alignment of the two rice subspecies might be useful for identifying regions of cereal genomes that are prone to rapid evolution. The deviations from colinearity are frequently owing to insertions or deletions. Intraspecific sequence polymorphisms commonly occur in both coding and non-coding regions. These variations often occur in promoter region and may affect gene structures, likely contributing to intraspecific phenotypic adaptations.

One of the standard and most powerful tools of molecular biology is the ability to efficiently compare the sequence of any gene of interest with the sequences of all previously discovered genes. Many of the first genes to be sequenced in rice and other grasses were represented by abundant mRNAs (e.g., those encoding storage proteins and photosynthetic proteins). Members of the same gene families (e.g., paralogs), including those that were mapped to the same genomic position and thus were derived by vertical descent from a common ancestral gene (i.e., orthologs), were often cloned and analyzed in multiple species (Bennetzen and Ma, 2003). Comparisons of gene family members within and between species yielded the expected result that the genes were most highly conserved between the most closely related species. Moreover, sequence conservation was greatest in the protein-coding portions of the exons.

Implications of genome colinearity

Genomics would be much simpler if the order of genes were common (syntenic) across major groups of plants or if we understood better why some regions appear to be highly conserved while others show rapid divergence. The usefulness of assessing colinearity between the genomes of model plants and important crops can be evaluated by the number of failures or successes in its exploitation.

Map-based cloning in plants with large genomes, such as barley, maize, and wheat, has been extremely difficult. However, it may be possible to use comparative maps to isolate a mapped gene from a large genome using a related plant with a small genome. Markers linked to the gene of interest and prior knowledge about colinearity of this region between large and small genomes are essential to isolate a gene using this approach. One of the first applications of colinearity between rice and another cereal species was the attempt to clone a

specific barley disease resistance gene by chromosome walking in rice (Kilian *et al.*, 1997). The colinearity provided numerous DNA markers from rice that facilitated the chromosome walk in barley, leading to the isolation of the desired stem-rust resistance gene *Rpg1*, although the synteny with rice failed to yield the gene because that particular gene does not seem to exist in rice (Brueggeman *et al.*, 2002). The application of colinearity among other cereals was also reported. The *Lr21* leaf-rust-resistance gene of bread wheat was successfully isolated using a strategy of shuttle-mapping between diploid wheat as a model and bread wheat (Huang *et al.*, 2003). However, this approach has potential pitfalls, especially with respect to some disease resistance genes, because resistance gene regions often undergo rapid rearrangement that results in a lack of micro-colinearity caused by deletion or translocation of the target loci. Peptide sequence comparison of dicot *R* genes and monocot *R*-like genes revealed shared motifs but provided no evidence for a monocot-specific signature. Interspecific analyses of *R*-like genes frequently revealed nonsyntenic map locations between the cereal species rice, barley, and foxtail millet (Leister *et al.*, 1998). Many breakages in microsynteny prevent the straightforward identification of a candidate gene by proxy. This was the case when attempts were made to isolate the leaf-rust-resistance gene *Rph7* (Brunner *et al.*, 2003) or the photoperiod response gene *Phd-H1* (Dunford *et al.*, 2002) from barley. A similar story was reported for the *Rfo* restorer genes isolated from radish: markers flanking these genes in radish are collinear with the *Arabidopsis* sequence, but the gene itself is not present in *Arabidopsis* although many homologues are present elsewhere in the *Arabidopsis* genome (Brown *et al.*, 2003; Desloire *et al.*, 2003). In conclusion, the use of a shuttle-mapping strategy has to be evaluated on a case-by-case basis, and even then, the numerous pitfalls of this approach must be kept in mind. The present information, from both successes and failures, strongly suggests that the development of efficient tools for isolating genes of agronomic importance within each important family should continue to be a priority, and that, as indicated by Delseny (2004), restricting ourselves to use the two existing model species, *Arabidopsis* and rice,

would be unwise, although colinearity has been useful in providing additional markers with which to saturate fine genetic and physical maps.

The cross-utilization of information from reference species, such as rice in the study of major cereal crops, requires a detailed understanding of the evolutionary history of plant genomes. To achieve a more complete picture, it is important to broaden our knowledge of diversity, and extend data from model plants across (and beyond) a few well studied genera. One important outcome of comparative gene analysis among rice and other cereals is the insight that is being gained into the evolutionary aspects of gene function in higher plants. Whole-genome duplication through polyploidization, segmental duplication, and local gene amplification increases the number of paralogous gene sequences found in plants. Together, these forms of duplication help explain why the numbers of genes in plant genomes tend to exceed those in other eukaryotic genomes, including the human genome. The use of all three duplication mechanisms in the evolution of grass genomes suggests that grasses, like other plants, may have evolved more rapidly than could be predicted by comparing only coding-sequence substitution rates. This rapid evolution may be one of the reasons for the degree of speciation within certain large tribes of the Gramineae family. The availability of new sequence information for additional cereal genomes will greatly facilitate ongoing efforts to understand the patterns of conservation and disruption in macro and micro-colinearity among the cereal genomes in light of the dynamic evolutionary history of plant genomes.

Where microcolinearity is broken and a gene that is present in one cereal is 'missing' from its homologous position in another, it may be possible to find a matching gene homologue in a non-orthologous location (Song *et al.*, 2002; Xu *et al.*, 2002). One putative mechanism for this phenomenon is an ancient gene duplication in the common ancestor followed by the loss of one gene copy in the first modern species and the loss of the other copy in the second species. Detailed analyses of the genomes of several model organisms revealed that large-scale gene or even entire genome duplications have played a prominent role in the evolutionary history of many eukaryotes. Recently, strong evidence has been presented that the genomic structure of the dicotyledonous model

plant species *Arabidopsis thaliana* is the result of multiple rounds of entire genome duplications (Simillion *et al.*, 2002). Segmental duplications were observed to occur commonly in the rice genome by using molecular marker analysis before the availability of the genome sequence (Wang *et al.*, 1999a, 2000). More recent analysis based on genome sequence revealed that a substantial fraction of rice genes (approximately 15%) were found in duplicated segments (Vandepoele *et al.*, 2003). Dating of these block duplications, their nonuniform distribution over the different rice chromosomes, and comparison with the duplication history of *Arabidopsis* suggest that rice is not an ancient polyploid, as suggested previously, but an ancient aneuploid that has experienced the duplication of one- or a large part of one-chromosome in its evolutionary past, approximately 70 million years ago. This predates the divergence of most of the cereals, and relative dating by phylogenetic analysis showed that this duplication event is shared by most, if not all, of them.

From sequences to molecular markers and marker-assisted selection

In plant breeding, the rice sequence can be directly used to acquire an almost unlimited number of DNA markers and genes for cereal improvement. A genome-wide rice DNA polymorphism database has been constructed using the genomes of Nipponbare and 93–11, which contains 1 703 176 single nucleotide polymorphisms (SNPs) and 479 406 insertions/deletions (indels), approximately one SNP every 268 bp and one indel every 953 bp in the rice genome (Shen *et al.*, 2004). Rice sequence information will help gain instantaneous access to the genes in breeding populations and will help evaluate the rich diversity of cereal germplasm through the use of molecular markers such as SNPs that can be increasingly identified within genes targeting functional nucleotide polymorphisms.

Functional markers (FMs)

Genetic diversity at or below the species level is mostly characterized by molecular markers that more or less randomly sample genetic variation in the genome. This type of 'neutral' or random marker (RM) is a very effective tool, amongst others, for the establishment of the breeding system, the study of gene flow among natural populations, and the determination of the genetic structure of genebank collections. RM systems are still the methods of choice for marker-assisted breeding (Dudley, 1993; Hospital and Charcosset, 1997; Stuber *et al.*, 1999; Xu, 2003). However, 'users' of biodiversity are often not interested in random variation but rather in variation that might affect the evolutionary potential of a species or the performance of an individual genotype. Such 'functional' variation can be tagged with neutral molecular markers using genetic and linkage disequilibrium mapping approaches. Alternatively, DNA-profiling techniques (Epplen and Lubjuhn, 1999; Henry, 2001) may be used that specifically target genetic variation in functional parts of the genome.

Different approaches (including association studies) have recently been adopted for the functional characterization of allelic variation in plants and the identification of sequence motifs affecting phenotypic variation. Andersen and Lübberstedt (2003) proposed the term 'FMs' for DNA markers derived from such functionally characterized sequence motifs. FM development requires allele sequences of functionally characterized genes from which polymorphic, functional motifs affecting plant phenotype can be identified. FMs are superior to RMs such as most RFLPs, SSRs and AFLPs owing to complete linkage with trait locus alleles and functional motifs. In contrast to RMs, FMs allow reliable application of markers in populations without prior mapping, the use of markers in mapped populations without risk of information loss owing to recombination, and better representation of genetic variation in natural or breeding populations. Once genetic effects have been assigned to functional sequence motifs, FMs derived from such motifs can be used to fix alleles (defined by one or several FM alleles) in several genetic backgrounds without additional calibration. This would be a major advance in marker applications, particularly in plant breeding, to select (for example) parental materials to build segregating populations, as well as subsequent selection of inbred lines (Andersen and Lübberstedt, 2003). An example of the application of FM in rice is the SSR marker residing in the *Waxy* gene (Bligh *et al.*, 1995; Ayres *et al.*, 1997) used to improve

Shanyou 63, the most widely cultivated hybrid in China. Its amylose content was modified by transferring the allele specifying a medium amylose content from the restorer line to replace the allele for high amylose content in the maintainer line, which simultaneously modified four quality related traits of this hybrid (Zhou *et al.*, 2003). Depending on the mode of FM characterization, FMs can also be used for the combination of target alleles in hybrid and synthetic breeding and variety testing based on the presence or absence of specific alleles at morphological trait loci. In population breeding and recurrent selection programs, FMs can be used to avoid genetic drift at characterized loci.

Conserved ortholog sequence markers

As a challenge to finding the manner in which map, sequence, and eventually functional genomic information from one species can be accessed, compared, and exploited across all plant species, it will require the identification of a subset of plant genes that have remained relatively stable in both sequence and copy number since the radiation of flowering plants from their last common ancestors. Identification of such a set of genes also would facilitate taxonomic and phylogenic studies in higher plants that are based at present on a very small set of highly conserved sequences, especially those of chloroplast and mitochondrial genes. Fulton *et al.* (2002) screened a large tomato EST database against the *Arabidopsis* genomic sequence and reported the identification of a set of 1025 genes (referred to as conserved orthologous sequences, COS) that are single or low copy in both genomes (as determined by computational screens and DNA gel blot hybridization) and that have remained relatively stable in sequence since the early radiation of dicotyledonous plants. Development of COS markers for cereal species that further target known functional domains within genes is practical now with the availability of complete rice sequences and a large number of ESTs from rice, maize and other cereals. Functional COS markers, identified computationally and experimentally, may further studies on comparative genomes and phylogenetics and elucidate the nature of genes conserved throughout plant evolution.

Single nucleotide polymorphism (SNP)

Comparative information about the chromosome organization of the two closely related rice subspecies has important implications for development of new molecular markers. The forward genetics approach for identifying functionally important genes derives from a known allelic difference conferring a different phenotype. In such an approach, the objective is to identify a sequence change that confers a different phenotype. Such a sequence change can then become the basis for a marker that is specific for that allele. These types of markers will always cosegregate with the trait of interest and should also be polymorphic in many crosses. Such a marker will often be based on SNP. By systematically searching for nucleotide differences, a complete set of markers that is based on SNPs or other sequence variations has been developed (Shen *et al.*, 2004). SNP analysis of rice germplasm using gel based or DNA-chip based methods can be used to identify existing genetic variation or multiple alleles for agronomically important traits. In the reverse genetics approach, SNPs or other types of sequence differences (e.g., indels) are sought in candidate genes to identify the phenotypic effects of genes. Known SNPs can be used to infer population structure and identify new candidate genes through association mapping (Buckler and Thornsberry, 2002; Rafalski, 2002a, b; Garris *et al.*, 2003). Phenotypic differences that correspond to particular SNPs may be the result of the sequence change. These SNPs can then be used for marker assisted selection (MAS) or screening germplasm and elite breeding lines. The nucleotide change that contributes to quantitative variation has been referred to as a quantitative trait nucleotide (QTN) (Lyman *et al.*, 1999). Fine-mapping combined with sequence analysis can narrow the chromosomal region associated with a quantitative trait locus (QTL) down to a specific nucleotide change (Fridman *et al.*, 2004), as suggested by association studies on major-gene controlled traits (Remington *et al.*, 2001; Thornsberry *et al.*, 2001; Garris *et al.*, 2003; Jung *et al.*, 2004).

Targeting induced local lesions in genomes (TILLING)

TILLING is a reverse genetic method that combines random chemical mutagenesis with

PCR-based screening of gene regions of interest (McCallum *et al.*, 2000). This identifies a range of allele types, including missense and knock-out mutations. By comparing isogenic genotypes differing in single sequence motifs and combined with phenotyping, TILLING provides direct proof of function of both induced and natural polymorphisms without involvement of transgenic modifications. Three SNP discovery methods have been adapted for TILLING: full sequencing, denaturing high-pressure liquid chromatography (DHPLC), and cleavage with the *cel1* enzyme followed by polyacrylamide gel electrophoretic analysis (Henikoff and Comai, 2003). High-throughout TILLING has been adopted by the *Arabidopsis* TILLING Project to provide allelic series of point mutations for the general *Arabidopsis* community (Till *et al.*, 2003). Based on TILLING, a strategy called 'Ecotilling' was developed for detecting multiple types of polymorphisms in natural populations (Comai *et al.*, 2004). Ecotilling can be used to effectively discover natural polymorphisms in a large set of individuals. The discovered polymorphisms can be confirmed by sequencing and include base-pair changes, small insertions and deletions, and variation in microsatellite repeat number. Ecotilling allows rapid detection of variation in many individuals and is cost effective because only one individual for each haplotype needs to be sequenced. The technology is applicable to any organism including those that are heterozygous or polyploid.

Germplasm evaluation

Evaluation of germplasm resources is required for the continuous improvement of crop plants. Vast genetic resources are available for rice and other cereals, but, to date, few of them have been characterized at the molecular level. Automated, high-throughput genotyping systems make large-scale marker-assisted germplasm evaluation possible. As summarized by Xu *et al.* (2003), molecular markers can be used in germplasm evaluation for (1) differentiating cultivars and constructing heterotic groups; (2) identifying germplasm redundancy, underrepresented alleles, and genetic gaps in germplasm collections; (3) monitoring genetic shifts that occur during germplasm storage, regeneration, domestication, and breeding; (4) screening germplasm for novel/

superior genes (alleles); and (5) constructing a representative subset or core collection. Some of these applications have been used in broadening the genetic base of rice in the U.S. (Xu *et al.*, 2004). The tools of genome research may finally unleash the genetic potential of our wild and cultivated germplasm resources for the benefit of society (Tanksley and McCouch, 1997).

Ideally a breeder would like to have information about all alleles of all genes across all available germplasm accessions. But even if all alleles were known, the value of an allele in terms of its contribution to an agronomic trait would usually not be good or bad in itself, but would have to be evaluated in the context of other alleles in a genomic network. Ultimately, one might want to design a variety of synthetic alleles that promised to be better than the ones provided by nature. These could be generated randomly (e.g., by gene shuffling) or via a targeted approach (e.g., by domain swapping). So genomics is going to bring to the breeder a better description (including allele-tagging) as well as an extension of his/her breeding material. The bottleneck is in the ability to assess the phenotypes caused by countless different alleles and allele combinations using available germplasm resources (Peerbolte, 2004). Molecular methods, can be used to help screen a large number of accessions through a pooling strategy. This can be used to screen germplasm collections for alleles of candidate genes that are involved in important processes of the plant, even though known variants for these genes have not yet been observed through genetic studies. A DNA bank is currently being developed as the basis for allele mining at the International Rice Research Institute to undertake allele mining (Mackill, 2003).

Marker-assisted selection (MAS)

MAS can be considered as one of the first direct benefits that breeders have obtained from genomics. With existing techniques, however, the use of molecular markers is still quite expensive for application on a large scale in many breeding programs. Many review articles are available for MAS (e.g., Dudley, 1993; Hospital and Charcosset, 1997; Stuber *et al.*, 1999; Xu, 2003), but the utility of molecular markers in plant breeding is best documented by highlighting their specific applications. As summarized by Xu (2002), there

are six situations most suitable for MAS with the level of current knowledge. These include selection without testcrossing or a progeny test; selection independent of environments; selection without laborious fieldwork or intensive laboratory work; selection at an earlier breeding stage; selection for multiple genes and/or multiple traits; and whole genome selection. Examples of where MAS would be advantageous include selecting for traits that are difficult or expensive to measure (e.g., resistance to a quarantine pest, restorer genes); pyramiding multiple genes that confer a similar or identical phenotype (e.g., multiple genes for disease resistance or quantitative traits); separating multiple QTL with similar effects or closely linked QTL; or selecting against the donor chromosomal segments in a backcrossing scheme (Young and Tanksley, 1989; Xu, 1997; Ribaut and Hoisington, 1998). Using rice and other cereal crops as examples, Xu (2003) provided a comprehensive review on MAS system, germplasm evaluation, hybrid prediction, and seed quality control, with emphasis on rice and other major cereals.

Wild relatives of cultivated plants harbor abundant genetic diversity and can be used to identify novel alleles for plant improvement. For example, wild relatives of rice represent a rich source of variation and can be used to discover agronomically important alleles for future breeding efforts in rice and other cereals as well. To fill the gulf between national research programs and breeding applications in developing countries, an international program called the 'Generation Challenge Program–Unlocking the genetic diversity in crops for resource poor' (www.cerealsgenomics.org), has been established to begin to characterize and utilize a wide spectrum of germplasm collections. Molecular markers have been proven particularly useful for accelerating the backcrossing of a gene or QTL from exotic cultivars or wild relatives into an elite cultivar or breeding line (Tanksley and Nelson, 1996). Favorable genes or alleles from wild species of rice have been detected after backcrossing to elite cultivars (Xiao *et al.*, 1998; Moncada *et al.*, 2001; Brondani *et al.*, 2002; Septiningsih *et al.*, 2003; Thomson *et al.*, 2003). This approach is thought to be promising in rice because a number of rice cultivars are widely grown for their adaptation, stable performance, and desirable grain quality. Chen *et al.* (2000) used such an approach to transfer the bacterial blight resistance gene *Xa21* into Minghui 63, a widely used parent for hybrid rice production in China. Ahmadi *et al.* (2001) used a similar approach to introgress two QTL controlling resistance to rice yellow mottle virus into the cultivar IR64. Such approaches, however, can only sample a small number of accessions.

From sequence to gene function

As we trek into the uncharted territories of the genomic era, there is urgency to develop approaches for assigning functions to the multitude of uncharacterized genes. Current progress in generating large numbers of molecular markers and near-saturation insertional mutant collections has immensely facilitated functional genomics studies in *Arabidopsis* (Østergaard and Yanofsky, 2004). These resources provide the essential blueprint for understanding the structure and function of many plant genes. Some of the insights emerging from whole-genome analysis include a surprising number of gene clusters, large-scale duplications of chromosome segments, and a high frequency (40%) of newly discovered genes of unknown function. In this section, several important issues related to functional analysis of genes will be discussed.

High-throughput techniques

Progress in functional genomics will depend on developing high-throughput technologies that can easily be used in both dedicated centers and individual laboratories to determine the function of the tens of thousands of plant genes. Many plant genes are novel showing no detectable homology to genes for which cellular roles have been identified in bacteria, yeast, or animals. DNA-based microarrays that detect the accumulation of transcripts from thousands of genes in a single hybridization experiment are one of the tools available to help decipher gene function. High-throughput techniques based on oligo chips, and various applications of serial analysis of gene expression (SAGE) are used for global gene expression analysis in particular tissue or at a particular stage. The gridding of thousands of unique DNA sequences on a chip provides a substrate that can be used to identify candidate genes that exert an influence at specific points in development. Common sources of

DNA for the arrays include cDNA, ESTs, subgenomic regions of specific chromosomes, or even an entire set of genes. Oligonucleotide based probe array (Gene Chip) technology, which has been commercially applied for *Arabidopsis* genes, has also been developed in rice and is now being applied in a number of studies. (Lan *et al.*, 2004; http://www.chem.aglient.com/scripts/PDS.asp?/Page = 12133). Availability of data for various gene expression studies will be essential for future studies on the functions of rice genes.

Map-based cloning

Although map-based cloning is feasible and genes have been cloned for several important rice traits such as disease resistance (Song *et al.*, 1995; Yoshimura *et al.*, 1998; Wang *et al.*, 1999b, Bryan *et al.*, 2000; Sun *et al.*, 2004), heading date (Yano *et al.*, 2000; Takahashi *et al.*, 2001; Kojima *et al.*, 2002), semidwarfism (Monna *et al.*, 2002; Spielmeyer *et al.*, 2002; Itoh *et al.*, 2004; Oikawa *et al.*, 2004), tillering (Li *et al.*, 2003a), and brittle culm (Li *et al.*, 2003b), it is generally regarded as time-consuming and laborious even for genes of typical Mendelian inheritance. In the last decade, large numbers of genes and QTL associated with traits of agronomic importance have been mapped and targeted for cloning in rice (Zhang and Yu, 1999; Li, 2001; Xu, 2002; www.gramene.org). The availability of whole genome sequence has greatly facilitated the process of positional cloning. It can be expected that a large number of genes will be isolated in the near future using this approach. The availability of genome sequence information is also reshaping the procedures of map-based cloning, by making each step more efficient (Jander *et al.*, 2002). With a larger number of sequence-based molecular markers available, low resolution genetic mapping makes it possible to quickly associate a target trait with a specific genomic region, and then to readily fine map the trait and narrow down the target region to one or more gene candidates.

Developments in high-throughput genomics are facilitating the process of dissecting the genetic basis of complex traits, including defining genetic intervals, identifying candidate genes and verifying an allele's contribution to a phenotype. With the availability of new technologies such as microarray,

fluorescence polarization, mass spectrometry and molecular barcodes, it is possible to achieve throughputs of 10 000's markers, which shows promise for testing the feasibility of high resolution association studies based on natural variation/natural populations and for high-throughput cloning strategies using reverse genetics.

Functional genomic approaches

One of the challenges in functional genomics is to understand how thousands of gene products interact with each other to control plant development and the plant's ability to respond to environment. Functional genomics has been broadly applied to include many endeavors aimed at determining the functions of genes on a genome-wide scale, such as sequence alignment-based comparisons to identify homologs between and within organisms; transcriptional profiling to determine gene expression patterns; and yeast two-hybrid and other interaction analyses to help identify pathways, networks, and protein complexes (Henikoff and Comai, 2003). In contrast to the previously prevalent gene-by-gene approaches, new high-throughput methods are being developed for expression analysis as well as for the recovery and identification of mutants. The experimental approach is consequently changing from hypothesis-driven to nonbiased data collection and an archiving methodology that makes these data available for analysis using bioinformatics tools. Reverse genetics (sequenced gene to mutant and function) approaches are bound to play a more prominent role in functional genomics studies in the future.

In functional genomics, libraries or populations of mutants that cover all possible genes become an increasingly important tool. Mutant libraries are being constructed in rice and other cereals using chemical and physical mutagenesis, T-DNA insertion, and transposon tagging (Jeon *et al.*, 2000; Leung *et al.*, 2001; Xue and Xu, 2002; An *et al.*, 2003; Hirochika, 2003; Hirochika *et al.*, 2004; Kolesnik *et al.*, 2004; Sallaud *et al.*, 2004). These libraries are primarily used for functional analysis based on loss-of-function analyses. Gain-of-function approaches such as T-DNA activation tagging and gene overexpression are powerful complements to insertional mutagenesis. A library of enhancer trap lines has

18

been developed which will facilitate the detection and isolation of regulatory elements (Wu *et al.*, 2003). With microarray techniques widely used in cDNA-based expression profiling, transcript profiling, and metabolite profiling, phenotypic profiling is also beginning to play an important role in functional analysis of plant genes (Kjemtrip *et al.*, 2003). Although currently available RNAi and other knock-out methodologies could be used for uncovering the function of newly discovered genes, the mixed outcomes of these approaches in down-regulating gene expression necessitate further development and optimization of this strategy for rice and other cereals.

The identification of genes by computational approaches is relatively straightforward for organisms with compact genomes (such as bacteria and yeast), because exons tend to be large, and the introns are either nonexistent or short. The challenge is much greater for larger genomes (such as those of cereals), because the exonic 'signal' is buried under nongenic 'noise.' Computational sequence analysis methods, which detect genes in genomic DNA, can be broadly classified into two main categories: homology-based methods, and *ab initio* methods (Davuluri and Zhang, 2003). Currently, computational methods are usually exploited as a complement to and component of other functional genomics approaches.

It is expected that the functional genomics of model plants will contribute to the understanding of basic plant biology as well as to the exploitation of genomic information for crop improvement. This is because a large number of gene functions are conserved across species. Perhaps the most exciting application of comparative cereal genomics will be the identification of different versions of rice genes from other species. Orthologous genes in other cereals will have sequence and function similar to those in rice but could result in markedly different phenotypes. These genes will be available for introduction into rice to produce new types of plants with many novel features.

Transformation

Transformation of allelic series into isogenic backgrounds can confirm the function of individual sequence motifs. However, current plant transformation protocols based on nonhomologous end joining result in random genomic integration of transgenic DNA, position effects, multiple insertions of the transgene and transgene alterations (Xu, 1997; Hanin and Paszkowski, 2003), obscuring quantitative phenotypic differences between alleles. This can be circumvented using homologous recombination-based, locus-targeted integration of alleles. In rice, 1% of insertion events were found to result from homologous recombination (Terada *et al.*, 2002). If the efficiency of this technique can be improved, it would revolutionize rice genomics-genetics. If the method can be applied to other species, a similar advance in genomics-based research in all plants would occur.

Sequence and bioinformatics

As -omic technologies, from genomics to phenomics, continue to grow, so does the need for integration and interrogation across the various types of data and scientific disciplines. Such integration enables the identification of genes and gene products, and can elucidate the functional relationships between genotype and observed phenotype, thereby permitting a system-wide analysis from genome to phenome (Edwards and Batley, 2004). The translation of complete-genome DNA sequence data into protein structures and predicted functions will provide a vital link between the genetics of an organism and its expressed phenotype. As genomic data will increase in the areas of genetic mapping, genome sequencing, gene expression monitoring, insertional mutagenesis, and map-based cloning, adequate tools for input, integration, and query become necessary. Eventually integrating information on rice structural and functional genomics will provide an overall view of the network of genes involved in complex biological responses.

Rice genome databases that evolve from rigorous and systematic sequencing efforts should not merely function as storehouses for thousands of nucleolide bases or amino acids. Of particular importance is the ability to attach substantial functional information to the sequence. These databases should therefore provide the framework to allow post-sequencing analysis such as identifying genes and predicting the proteins they encode, determining when and where the encoded proteins are expressed and how they interact, and how these expression and interaction profiles are

modified in response to environmental signals. One way to address this need is to interlink the various types of information such as genomic data, phenotypic or expression data, and genetic resources. For a given gene, the database horizontally links sequence, structure, and map position and connects related elements of the same type pertaining to the expression profile, protein, and phenotype (Antonio *et al.*, 2001; Ware *et al.*, 2002). What would be most useful to the breeding community will be for all of this information to be linked to the allelic profiles of specific germplasm accessions and genetic resources that are available for rice and other cereals. Careful curation will be required to clarify whether allelic variation at specific genetic loci is functionally related to an alteration in phenotype or whether it is simply neutral sequence variation. Neutral sequence variation is useful in tracing the inheritance of particular alleles, and might be functionally relevant in some genetic backgrounds but not others. Thus, annotating polymorphisms throughout the genome and correlating them with phenotypic variation in different genetic backgrounds will require a long-term and well coordinated effort to provide high quality sequence annotation using a set of carefully selected germplasm resources and a rigorous evaluation system to characterize diverse phenotypes under a wide range of environmental conditions. Logical connections to other information will enhance the intrinsic value of the genomic data to facilitate new biological discoveries and simulate approaches for an effective cereal improvement program.

Rice, serving as a model system for other grass crops, has established an informatics infrastructure that is designed to interlink database resources on rice and other cereals (www.gramene.org). Comparative bioinformatics offers possibilities to link various cereal crops through their genome map and sequences and to provide keys to understanding how genes and genomes are structured, how they function, and how they have evolved. To provide the cereal community with a resource for linking rice genome information to other species, the Gramene database develops curated and sequence-based correspondences between genetic, physical and sequence-based maps (Jaiswal *et al.*, 2002; Ware *et al.*, 2002; www.gramene.org). This resource includes curated correspondences among the Maize

Mapping Project BAC based maize fingerprint contig maps, genetic markers, and sequence tagged BAC clones. The correspondences between the maize and rice genomes are available as a graphical display as well as a downloadable tabular format. The creation of links between different databases will help foster interoperability and large comparative databases will provide a forum for coordinating and integrating that information. One of the most serious challenges facing of specialized or expert domain databases, such as the model organism databases, is to balance the needs of the broader scientific community and the specialized focused user-groups.

Perspectives

There is a gulf between genomics and its application to plant breeding although there are many opportunities available. Plant breeders have to work with a large number of agronomic traits (most are quantitatively inherited) that are expressed under a wide range of environments during a relatively short period. Plant breeders make selections with the aid of statistical analysis, phenotypic selection techniques and molecular tools (Xu and Zhu, 1994). As a result, genomics can be applied to plant breeding only when an integrated package becomes available that combines multiple components such as high-throughout techniques, cost-effective protocols, global integration of genetic and environmental factors, and precise determination of quantitative trait expression.

Functional genomics has become more practical because of advances in science and technology. Some examples are DNA-capturing techniques to facilitate gene hunting efforts in highly repetitive genomes, motif-directed profiling to specifically target genetic variation in functional parts of the genome, gene expression profiling to identify expression QTL (eQTL), and RNA interference (RNAi) to silence individual genes without altering the genome structure. Microarrays or other nongel systems may allow whole-genome analysis of large number of plants commonly grown in breeding programs. For example, proteome chips can be used to perform various biochemical analyses, protein–protein interaction assays, protein–DNA/RNA interactions, protein–phospholipids interac-

tions and the identification of substrates for kinases and other enzymes. On the other hand, the global analysis at the levels of whole genome (discovery of all genes), transcriptome (quantification of gene expression), proteome (cataloguing of all proteins) and metabolome (estimation of all types of metabolites) has become possible. To assign a cellular function to many novel genes which are predicted from the whole genome analysis, several high-throughput approaches, such as DNA chips, SAGE, RNA-mediated interference, gene traps, yeast two-hybrid screening and metabolite quantification can now be employed in functional genomics of rice and other cereals. Ultimately, the goal will be to rapidly assay the genetic makeup of individual plants or varieties in breeding populations and make accurate phenotypic predictions. Conversely, the knowledge can be used to design a genotype that is targeted to perform well in a given set of environmental conditions.

Many novel techniques need to be improved before they can be widely used in functional genomics analysis and plant breeding. For example, 2-D gel electrophoresis, which has been the method of choice for separation of complex protein mixtures for several decades, has inherent technical limitations, such as the limited ability to fractionate specific classes of proteins or to visualize low abundance proteins and the notorious difficulty in automation which limits throughput and results in greater experimental variability through manual intervention (Rose *et al.*, 2004). As discussed previously, gene function can be altered through gene deletions, insertional mutagenesis, RNAi or allele replacement. Insertional mutagenesis has been recognized as the potential to provide a vast catalog of knockout mutations for an organism. However, a significant insertion-site bias, plus a high level of background mutation, can confound subsequent phenotypic analysis. Also the efficacy of mutagenesis screens in identifying gene functions is limited because the majority of genes display no obvious phenotype as well as other methodological considerations. Insertions may not result in null alleles, depending on their position within the open reading frame (ORF), or intronic or untranslated regions, and effort is required to confirm that any resulting phenotype is not because of unlinked independent insertions. Because of genetic redundancy, a vast proportion of genes are silent when knocked out,

or might only show a subtle phenotype or one only revealed under extreme environmental conditions. A targeted deletion approach is also not without caveats – only annotated ORFs in a genome are generally targeted; the phenotypes of overlapping ORFs cannot be easily distinguished; transformation events might introduce secondary mutations; essential genes are inviable as homozygous deletions; and functionally redundant genes might not show a detectable phenotype when deleted (Steinmetz and Davis, 2004). The primary advantages of RNAi are the ease of generating short double-stranded RNAs that mediate RNAi and the flexibility of inhibition so that the user can spatially and temporally control the interference reaction. The disadvantage in systematic genome-wide application, however, is that the level of functional reduction is unpredictable and difficult to measure experimentally.

To maximize the potential of functional genomics for plant breeding, two goals should be met: experiments must be further miniaturized and costs must be lowered. Molecular barcoding is among the most promising techniques for miniaturizing high-throughput approaches into single tube assays. In a method termed 'molecular inversion probes,' genotyping can be performed directly on genomic DNA and thousands of SNPs can be analyzed in one reaction by taking advantage of the multiplexing capability of DNA barcoding (Hardenbol *et al.*, 2003). Through a combination of molecular barcodes and the microarray-based detection system, the yeast-deletion approach has achieved an unparalleled level of throughput (Steinmetz and Davis, 2004). The barcode concept can be applied to other knock-out approaches, such as insertional mutagenesis and RNAi, as well as to other biological assays.

As the sequence and annotation of the rice genome has been greatly accelerated by methodological improvements made in *Arabidopsis* (Rensink and Buell, 2004), the development of functional genomics resources and technologies in cereals has been, and will continue to be, more streamlined based on lessons learned from rice. With all the advances in functional genomics and genome sequencing, it is expected that crop plants will become more amenable to genetic manipulation at both the phenotypic and molecular levels with more directed breeding

strategies and objectives. As indicated by Mackill (2003), however, new tools and technologies developed in genomics are expected to greatly enhance, but not replace, the conventional breeding process.

References

Ahmadi, N., Albar, L., Pressoir, G., Pinel, A., Fargette, D. and Ghesquière, 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, S. and Tanksley, S.D. 1993. Comparative linkage maps of the rice and maize genomes. Proc. Natl. Acad. Sci. USA 90: 7980–7984.

An, G., Jeong, D.-H., An, S., Kang, H.-G., Moon, S., Han, J., Park, S., Lee, H.S. and An, K. 2003. Activation tagged mutants to discover novel rice genes. In: T.W. Mew, D.S. Brar, S. Peng, D. Dawe and B. Hardy (Eds.), Rice Science: Innovations and Impact for Livelihood. Proceedings of the International Rice Research Conference, 16–19 September 2002, Beijing, China, International Rice Research Institute, Chinese Academy of Engineering, and Chinese Academy of Agricultural Sciences, pp. 195–204.

Andersen, J.R. and Lübberstedt, T. 2003. Functional markers in plants. Trends Plant Sci. 8: 554–560.

Antonio, B.A., Sakata, K. and Sasak, T. 2001. Bioinformatics and the rice genome. In: G.S. Khush, D.S. Brar and B. Hardy (Eds.), Rice Genetics IV. Proceedings of the Fourth International Rice Genetics Symposium, 22–27 October 2000, Los Baños, Philippines, Science Publishers, Inc., New Delhi, India and International Rice Research Institute, Los Baños, Philippines, pp. 293–305.

Ayres, N.M., Mclung A.M., Larkin, P.D., Bligh, H.F.J., Jones, C.A. and Park, W.D. 1997. Microsatellites and a single-nucleotide polymorphism differentiate apparent amy-lose classes in an extended pedigree of US rice germ plasm. Theor. Appl. Genet. 94: 773–781.

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

Bennetzen, J.L. and 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.

Bligh, H.F.J., Till, R.I. and Jones, C.A. 1995. A microsatellite sequence closely linked to the waxy gene of Oryza sativa. Euphytica 86: 83–85.

Brondani, C., Rangel, N., Brondani, V. and Ferreira, E. 2002. QTL mapping and introgression of yield-related traits from Oryza glumaepatula to cultivated rice (Oryza sativa) using microsatellite markers. Theor. Appl. Genet. 104: 1192–1203.

Brown, G.G., Formanova, N., Jin, H., Wargachuk, R., Dondy, C., Patil, P., Laforest, M., Zhang, J., Cheung, W.Y. and Landry, B.S. 2003. The radish Rfo restorer gene of Ogura cytoplasmic mole sterility encodes a protein with multiple pentatricopeptide repeat. Plant J. 35: 262–272.

Brueggeman, R., Rostoks, N., Kudrna, D., Kilian, A., Han, F., Chen, J., Druka, A., Steffenson, B. and 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., Keller, B. and Feuillet, C. 2003. A large rearrangement involving genes and low copy DNA inter-rupts the micro-collinearity between rice and barley at the Rph7 locus. Genetics 164: 673–683.

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

Buckler, E.S. and Thornsberry, J.M. 2002. Plant molecular diversity and applications to genomics. Curr. Opin. Plant Biol. 5: 107–111.

Chen, S., Lin, X.H., Xu, C.G. and Zhang, Q.F. 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, H., Wang, S., Xing, Y., Xu, C., Hayes, P.M. and Zhang. Q. 2003. Comparative analyses of genomic locations and race specificities of loci for quantitative resistance to Pyric-ularia grisea in rice and barley. Proc. Natl. Acad. Sci. USA 100: 2544–2549.

Comai, L., Young, K., Till, B.J., Reynolds, S.H., Greene E.A., Codomo C.A., Enns L.C., Johnson J.E., Burtner, C., Odden, A.R. and Henikoff, S. 2004. Efficient discovery of DNA polymorphisms in natural populations by Ecotilling. Plant J. 37: 778–786.

Davuluri, R.V. and Zhang, M.Q. 2003. Computer software to find genes in plant genomic DNA. In: E. Grotewold (Ed.), Methods in Molecular Biology, Vol. 236. Plant Functional Genomics: Methods and Protocols, pp. 87–107.

Delseny, M. 2004. Re-evaluating the relevance of ancestral shared synteny as a tool for crop improvement. Curr. Opin. Plant Biol. 7: 126–131.

Desloire, S., Gherbi, H., Laloui, W., Marhadour, S., Clouet, V., Catholico, L., Falentin, C., Giancola, S., Renard, M., Budar, F., Small1, I., Caboche, M., Delourme R. and Bendahman, A. 2003. Identification of the fertility restoration locus Rfo in radish as a member of the pentatricopeptide repeat family. EMBO Rep. 4: 588–594.

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

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

Dudley, J.W. 1993. Molecular markers in plant improvement: manipulation of genes affecting quantitative traits. Crop Sci. 33: 660–668.

Dunford, R.P., Yano, M., Kurata, N., Sasaki, T., Huestis, G., Rocheford, T. and Laurie, D.A. 2002. Comparative mapping of the barley Phd-H1 photoperiod response gene region, which lies close to a junction between two rice linkage segments. Genetics 161: 825–834.

Edwards, D. and Batley, J. 2004. Plant bioinformatics: from genome to phenome. Trends Biotech. 22: 232–237.

Epplen, J.T. and Lubjuhn, T. (Eds.), 1999. DNA Profiling and DNA Fingerprinting. Birkhauser Verlag, Boston.

Faris, J.D., Fellers, J.P., Brooks, S.A. and Gill, B.S. 2003. A bacterial artificial chromosome contig spanning the major domestication locus Q in wheat and identification of a candidate gene. Genetics 164: 311–321.

Fatokun, C.A., Menancio-Hautea, D.I., Danesh, D. and Young, N.D. 1992. Evidence for orthologous seed weight genes in cowpea and mung bean based on RFLP mapping. Genetics 132: 841–846.

Feng, Q., Zhang, Y., Hao, P., Wang, S., Fu, G., Huang, Y., Li, Y., Zhu, J., Liu, Y., Hu, X., Jia, P., Zhang, Y., Zhao, Q., Ying, K., Yu, S., Tang, Y., Weng, Q., Zhang, L., Lu, Y., Mu, J., Lu, Y., Zhang, L.S., Yu, Z., Fan, D., Liu, X., Lu, T., Li, C., Wu, Y., Sun, T., Lei, H., Li, T., Hu, H., Guan, J., Wu, M., Zhang, R., Zhou, B., Chen, Z., Chen, L., Jin, Z., Wang, R., Yin, H., Cai, Z., Ren, S., Lv, G., Gu, W., Zhu, G., Tu, Y., Jia, J., Zhang, Y., Chen, J., Kang, H., Chen, X., Shao, C., Sun, Y., Hu, Q., Zhang, X., Zhang, W., Wang, L., Ding, C., Sheng, H., Gu, J., Chen, S., Ni, L., Zhu, F., Chen, W., Lan, L., Lai, Y., Cheng, Z., Gu, M., Jiang, J., Li, J., Hong, G., Xue, Y. and Han, B. 2002. Sequence and analysis of rice chromosome 4. Nature 420: 316–320.

Feuillet, C. and Keller, B. 1999. High gene density is conserved at syntenic loci of small and large grass genomes. Proc. Natl. Acad. Sci. USA 96: 8625–8270.

Fridman, E., Carrari, F., Liu, Y.-S. Fennie, A.R. and Zamir, D. 2004. Zooming in on a quantitative trait for tomato yield using interspecific introgressions. Science 305: 1786–1789.

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

Fulton, T.M., van der Hoeven, R., Eannetta. N.T. and Tanksley, S.D. 2002. Identification, analysis, and utilization of conserved ortholog set markers for comparative genomics in higher plants. Plant Cell 14: 1457–1467.

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

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

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

Hampson, S., McLysaght, A., Gaut, B. and Baldi, P. 2003. LineUp: statistical detection of chromosomal homology with application to plant comparative genomics. Genome Res. 13: 999–1010.

Han, B. and Xue, Y. 2003. Genome-wide intraspecific DNA-sequence variations in rice. Curr. Opin. Plant Biol. 6: 134–138.

Hanin, M. and Paszkowski, J. 2003. Plant genome modification by homologous recombination. Curr. Opin. Plant Biol. 6: 157–162.

Hardenbol, P., Banér, J., Jain, M., Nilsson, M., Namsaraev, E.A., Karlin-Neumann, G.A., Fakhral-Rad, H., Rongaghi, M., Willis, T.D., Landegren, U. and Davis, R.D. 2003. Multiplexing genotyping with sequence-tagged molecular inversion probes. Nat. Biotech. 21: 673–678.

Henikoff, S. and Comai, L. 2003. Single-nucleotide mutations for plant functional genomics. Ann. Rev. Plant Biol. 54: 375–401.

Henry, R.J. (Ed.), 2001. Plant Genotyping-the DNA Fingerprinting of Plants. CABI Publishing, Wallingford.

Hiei, Y., Ohta, S., Komari, T. and Kumashiro, T. 1994. Efficient transformation of rice (*Oryza sativa* L.) mediated by *Agrobacterium* and sequence analysis of the boundaries of the T-DNA. Plant J. 6: 271–282.

Hirochika, H. 2003. Insertional mutagenesis in rice using the endogenous retrotansposon. In: T.W. Mew, D.S. Brar, S. Peng, D. Dawe, and B. Hardy (Eds.), Rice Science: Innovations and Impact for Livelihood, Proceedings of the International Rice Research Conference, 16–19 September 2002, Beijing, China, International Rice Research Institute, Chinese Academy of Engineering, and Chinese Academy of Agricultural Sciences, pp. 205–212.

Hirochika, H., Guiderdoni, E., An, G., Hsing, Y.I., Eun, M.Y., Han, C.D., Upadhyaya, N., Ramachandran, S., Zhang, Q., Pereira, A., Sundaresan, V. and Leung, H. 2004. Rice mutant resources for gene discovery. Plant Mol. Biol. 54: 325–334.

Hospital, F. and Charcosset, A. 1997. Marker-assisted introgression of quantitative trait loci. Genetics 147: 1469–1485.

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

Itoh, H., Tatsumi, T., Sakamoto, T., Otomo, K., Toyomasu, T., Kitano, H., Ashikari, M., Ichihara, S. and Matsuoka, M. 2004. A rice semi-dwarf gene, *Tan-Ginbozu* (D35), encodes the gibberellin biosynthesis enzyme, *ent*-kaurene oxidase. Plant Mol. Biol. 54: 533–547.

Jaiswal, P., Ware, D., Ni, J., Chang, K., Zhao, W., Schmidt, S., Pan, X., Clark, K., Teytelman, L., Cartinhour, S., Stein, L. and McCouch, S. 2002. Gramene: development and integration of trait and gene ontologies for rice. Comp. Funct. Genomics 3: 132–136.

Jander, G., Norris, S.R., Rounsley, S.D., Bush, D.F., Levin, I.M. and Last, R.L. 2002. Arabidopsis map-based cloning in the post-genome era. Plant Physiol. 129: 440–450.

Jeon, J.S., Lee, S., Jung, K.H., Jun, S.H., Jeong, D.H., Lee, J., Kim, C., Jang, S., Lee, S., Yang, K., Nam, J., An, J., Han, M.J., Sung, R.J., Choi, H.S., Yu, J.H., Choi, J.H., Cho, S.Y., Cha, S.S., Kim, S.I. and An, G. 2000. T-DNA insertional mutagenesis for functional genomics of rice. Plant J. 22: 561–570.

Jung, M., Ching, A., Bhattramakki, D., Dolan, M., Tingey, S., Morgante, M. and Rafalski, A. 2004. Linkage disequilibrium and sequence diversity in a 500-kbp region around the *adh1* locus in elite maize germplasm. Theor. Appl. Genet. 109: 681–689.

Kellogg, E.A. 1998. Relationships of cereal crops and other grasses. Proc. Natl. Acad. Sci. USA 95: 2005–2010.

Kilian, A., Chen, J., Han, F., Steffenson, B. and Kleinhofs, A. 1997. Towards map-based cloning of the barley stem rust resistance genes *Rpg1* and *rpg4* using rice as an intergenomic cloning vehicle. Plant Mol. Biol. 35: 187–195.

Kjemtrip, S., Boyes, D.C., Christensen, C., McCaskill A.J., Hylton, M. and Davis K. 2003. Growth stage-based phenotypic profiling of plants. In: E. Grotewold (Ed.), Methods in

Molecular Biology, Vol. 236. Plant Functional Genomics: Methods and Protocols, pp. 427–441.

Klein, P.E., Klein, R.R., Vrebalov, J. and Mullet, J.E. 2003. Sequence-based alignment of sorghum chromosome 3 and rice chromosome 1 reveals extensive conservation of gene order and one major chromosomal rearrangement. Plant J. 34: 605–621.

Kojima, S., Takahashi, Y., Kobayashi, Y., Monna, L., Sasaki, T., Araki, T. and Yano, M. 2002. *Hd3a*, a rice ortholog of the *Arabidopsis FT* gene, promotes transition to flowering downstream of *Hd1* under short-day conditions. Plant Cell Physiol. 43: 1096–1105.

Kolesnik, T., Szeverenyi, I., Bachmann, D., Kumar, C.S., Jiang, S., Ramamoorthy, R., Cai, M., Ma, Z.G., Sundaresan, V. and Ramachandran, S. 2004. Establishing an efficient Ac/Ds tagging system in rice: large-scale analysis of Ds flanking sequences. Plant J. 37: 301–314.

Komari, T., Hiei, Y., Ishida, Y., Kumashiro, T. and Kubo, T. 1998. Advances in cereal gene transfer. Curr. Opin. Plant Biol. 1: 161–165.

Lan, L., Chen, W., Lai, Y., Suo, J., Kong, Z., Li, C., Lu, Y., Zhang, Y., Zhao, X., Zhang, X., Zhang, Y., Han, B., Cheng, J. and Xue, Y. 2004. Monitoring of gene expression profiles and isolation of candidate genes involved in pollination and fertilization in rice (*Oryza Sativa* L.) with a 10K cDNA microarray. Plant Mol. Biol. 54: 471–487.

Leister, D., Kurth, J., Laurie, D.A., Yano, M., Sasaki, T., Devos, K., Graner, A. and Schulze-Lefert, P. 1998. Rapid reorganization of resistance gene homologues in cereal genomes. Proc. Natl. Acad. Sci. USA 95: 370–375.

Leung, H., Wu, C., Baraoidan, M., Bordeos, A., Ramos, M., Madamba, S., Cabauatan, P., Vera Cruz, C., Portugal, A., Reyes, G., Bruskiewich, R., McLaren, G., Lafitte, R., Gregorio, G., Bennett, J., Brar, D., Khush, G., Schnable, P., Wang, G. and Leach, J. 2001. Deletion mutants for functional genomics: progress in phenotyping, sequence assignment, and database development. In: G.S. Khush, D.S. Brar and B. Hardy (Eds.), Rice Genetics IV. Proceedings of the Fourth International Rice Genetics Symposium, 22–27 October 2000, Los Baňos, Philippines, Science Publishers, Inc., New Delhi, India and International Rice Research Institute, Los Baňos, Philippines, pp. 239–251.

Li, Z. 2001. QTL mapping in rice. In: G.S. Khush, D.S. Brar and B. Hardy (Eds.), Rice Genetics IV. Proceedings of the Fourth International Rice Genetics Symposium, 22–27 October 2000, Los Baňos, Philippines, Science Publishers, Inc., New Delhi, India and International Rice Research Institute, Los Baňos, Philippines, pp. 153–171.

Li, X., Qian, Q., Fu, Z., Wang, Y., Xiong, G., Zeng, D., Wang, X., Liu, X., Teng, S., Hiroshi, F., Yuan, M., Luo, D., Han, B. and Li, J. 2003a. Control of tillering in rice. Nature 422: 618–621.

Li, Y., Qian, Q., Zhou, Y., Yan, M., Sun, L., Zhang, M., Fu, Z., Wang, Y., Han, B., Pang, X., Chen, M. and Li, J. 2003b. *BRITTLE CULM1*, which encodes a COBBA-like protein, affects the mechanical properties of rice plants. Plant Cell 15: 2020–2031.

Lin, Y.R., Schertz, K.F. and Paterson, A.H. 1995. Comparative analysis of QTLs affecting plant height and maturity across the Poaceae, in reference to an interspecific sorghum population. Genetics 140: 391–411.

Lyman, R.F., Lai, C. and Mackay, T.F. 1999. Linkage disequilibrium mapping of molecular polymorphisms at the scabrous locus associated with naturally occurring variation in bristle number in *Drosophila melanogaster*. Genet. Res. 74: 303–311.

Ma, J. and Bennetzen, J.L. 2004. Rapid recent growth and divergence of rice nuclear genomes. Proc. Natl. Acad. Sci. USA 101: 12404–12410.

McCallum, C.M., Comai, L., Green, E.A. and Henikoff. S. 2000. Targeting induced local lesions in genomes (TILLING) for plant functional genomics. Plant Physiol. 123: 439–442.

Mackill, D.J. 2003. Applications of genomics to rice breeding. Int. Rice Res. Note 28.1: 9–15.

McCouch, S.R. 2001. Genomics and synteny. Plant Physiol. 125: 152–155.

Moncada, P., Martinez, C.P., Borrero, J., Chatel, M., Gauch, H., Guimaraes, E., Tohme, J. and McCouch, S.R. 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.

Monna, L., Kitazawa, N., Yoshino, R., Suzuki, J., Masuda, H., Maehara, Y., Tanji, M., Sato, M., Nasu, S. and Minobe, Y. 2002. Positional cloning of rice semidwarfing gene, *sd-1*: rice 'green revolution gene' encodes a mutant enzyme involved in gibberellin synthesis. DNA Res 9: 11–17.

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

Oikawa, T., Koshioka, M., Kojima, K., Yoshida, H. and Motoshige, K. 2004. A role of *OsGA20ox1*, encoding an isoform of gibberellin 20-oxidase, for regulation of plant stature in rice. Plant Mol. Biol. 55: 687–700.

Østergaard, L. and Yanofsky, M.F. 2004. Establishing gene function by mutagenesis in *Arabidopsis thaliana*. Plant J. 39: 682–696.

Paterson, A.H., Bowers, J.E., Peterson, D.G., Estill, J.C. and Chapman, B.A. 2003. Structure and evolution of cereal genomes. Curr. Opin. Genet. Develop. 13: 1–7.

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

Peerbolte, R. 2004. Breeding by design: high throughput phenotyping. Plant & Animal Genome XII, Poster W265.

Peng, J., Richards, D.E., Hartley, N.M., Murphy, G.P., Devos, K.M., Flintham, J.E., Beales J., Fish, L.J., Wordland, A.J., Pelica, F., Sudhakar D., Christou, P., Snape, J.W., Gale, M.D. and Harberd, N.P. 1999. 'Green revolution' genes encode mutant gibberellin response modulators. Nature 400: 256–261.

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

Rafalski, J.A. 2002b. Novel genetic mapping tools in plants: SNPs and LD-based approaches. Plant Sci. 162: 329–333.

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

Rensink, W.A. and Buell, C.R. 2004. Arabidopsis to rice. Applying knowledge from a weed to enhance our understanding of a crop species. Plant Physiol. 135: 622–629.

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

Rose, J.K.C., Bashir, S., Giovannoni, J.J., Jahn, M.M. and Saravanan, R.S. 2004. Tackling the plant proteome:

practical approaches, hurdles and experimental tools. Plant J. 39: 715–733.

Ryu, C.-H., You, J.-H., Kang, H.-G., Hur, J., Kim, Y.-H., Han, M.-J., An, K., Chung, B.-C., Lee, C.-H. and An, G. 2004. Generation of T-DNA tagging lines with a bidirectional gene trap vector and the establishment of an insertion-site database. Plant Mol. Biol. 54: 489–502.

Salse, J., Piegu, B., Cooke, R. and Delseny, M. 2004. New *in silico* insight into the synteny between rice (*Oryza sativa* L.) and maize (*Zea mays* SL.) highlights reshuffling and identifies new duplications in the rice genome. Plant J. 38: 396–409.

Sallaud, C., Gay, C., Larmande, P., Bès, M., Piffanelli, P., Piégu, B., Droc, G., Regad, F., Bourgeois, E., Meynard, D., Périn, C., Sabau, X., Ghesquière, A., Glaszmann, J.C., Delseny, M. and Guiderdoni, E. 2004. High throughput T-DNA insertion mutagenesis in rice: a first step towards *in silico* reverse genetics. Plant J. 39: 450–464.

SanMiguel, P.J., Ramakrishna, W., Bennetzen, J.L., Busso, C.S. and Dubcovsky, J. 2002. Transposable elements, genes, and recombination in a 215-kb contig from wheat chromosome 5A(m). Funct. Integr. Genomics 2: 70–80.

Sasaki, T., Matsumoto,T., Yamamoto, K., Sakata, K., Baba, T., Katayose, Y., Wu, J., Niimura, Y., Cheng, Z., Nagamura, Y., Antonio, B.A., Kanamori, H., Hosokawa, S., Masukawa, M., Arikawa, K., Chiden, Y., Hayashi, M., Okamoto, M., Ando, T., Aoki, H., Arita, K., Hamada, M., Harada, C., Hijishita, S., Honda, M., Ichikawa, Y., Idonuma, A., Iijima, M., Ikeda, M., Ikeno, M., Ito, S., Ito, T., Ito, Y., Ito, Y., Iwabuchi, A., Kamiya, K., Karasawa, W., Katagiri, S., Kikuta, A., Kobayashi, N., Kono, I., Machita, K., Maehara, T., Mizuno, H., Mizubayashi, T., Mukai, Y., Nagasaki, H., Nakashima, M., Nakama, Y., Nakamichi, Y., Nakamura, M., Namiki, N., Negishi, M., Ohta, I., Ono, N., Saji, S., Sakai, K., Shibata, M., Shimokawa, T., Shomura, A., Song, J., Takazaki, Y., Terasawa, K., Tsuji, K., Waki, K., Yamagata, H., Yamane, H., Yoshiki, S., Yoshihara, R., Yukawa, K., Zhong, H., Iwama, H., Endo, T., Ito, H., Hahn, Jang H., Kim, H.-I., Eun, M.-Y., Yano, M., Jiang, J. and Gojobori, T. 2002. The genome sequence and structure of rice chromosome 1. Nature 420: 312–316.

Schmidt, R. 2002. Plant genome evolution: lessons from comparative genomics at the DNA level. Plant Mol. Biol. 48: 21–37.

Septiningsih, E.M., Prasetiyono, J., Lubis, E., Tai, T.H., Tjubaryat, T., Moeljopawiro, S. and McCouch, S.R. 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.

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

Shimamoto, K. and Kyozuka, J. 2002. Rice as a model for comparative genomics of plants. Annu. Rev. Plant Biol. 53: 399–419.

Simillion, C., Vandepoele, K., Van Montagu, M.C., Zabeau, M. and Van Depeer, Y. 2002. The hidden duplication past of Arabidopsis Thaliana. Proc. Natl. Acad. Sci. USA 99: 13627–13632.

Song, R., Llaca, V. and Messing, J. 2002. Mosaic organization of orthologous sequences in grass genome. Genome Res. 12: 1549–1555.

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

Sorrells, M.E., La Rota, M., Bermudez-Kandianis, C.E., Greene, R.A., Kentety, R., Munkvold, J.D., Miftahudin, Mahmoud, A., Ma, X.F., Gustafson, P.J., Qi, L.L., Echalier, B., Gill, B.S, Matthews, D.E., Lazo, G.R., Chao, S., Anderson, O.D., Edwards, H., Linkiewicz, A.M., Dubcovsky, J., Akhunov, E.D., Dvorak, J., Zhang, D., Nguyen, H.T., Peng, J., Lapitan, N.L.V., Gonzalez-Hernandez, J.L., Anderson, J.A., Hossain, K., Kalavacharla, V., Kianian, S.F., Choi, D.-W., Close, T.J., Dilbirligi, M., Gill, K.S., Steber, C., Walker-Simmons, M.K., McGuire, P.E. and Qualset, C.Q. 2003. Comparative DNA sequence analysis of wheat and rice genomes. Genome Res. 13: 1818–1827.

Spielmeyer, W., Ellis, M. and Chandler, P. 2002. Semidwarf (*sd-1*), green revolution rice, contains a defective gibberellin 20-oxidase gene. Proc. Natl. Acad. Sci. USA 99: 9043–9048.

Steinmetz, L.M. and Davis, R.W. 2004. Maximizing the potential of functional genomics. Nat. Rev. Genet. 5: 190–201.

Stuber, C.W., Polacco, M. and Senior, M.L. 1999. Synergy of empirical breeding, marker-assisted selection, and genomics to increase crop yield potential. Crop Sci. 39: 1571–1583.

Sun, X., Cao, Y., Yang, Z., Xu, C., Li, X., Wang, S. and Zhang, Q. 2004. *Xa26*, a gene conferring resistance to *Xanthomonas oryzae* pv. *oryzae* in rice, encodes a LRR receptor kinase-like protein. Plant J. 37: 517–527.

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

Tanksley, S.D. and McCouch, S.R. 1997. Seed banks and molecular maps: unlocking genetic potential from the wild. Science 277: 1063–1066.

Tanksley, S.D. and Nelson, J.C. 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.

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

Terada, R., Urawa, H., Yoshihsige, I., Thugane, K. and Lida, S. 2002. Efficient gene targeting by homologous recombination in rice. Nat. Biotech. 20: 1030–1034.

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

The Rice Chromosome 10 Sequencing Consortium. 2003. In-depth view of structure, activity, and evolution of rice chromosome 10. Science 300: 1566–1569.

The Rice Full-Length cDNA Consortium. 2003. Collection, mapping, and annotation of over 28 000 cDNA clones from *japonica* rice. Science 301: 376–379.

Thomson, M.J., Tai, T.H., McClung, A.M., Lai, X.H., Hinga, M.E., Lobos, K.B., Xu, Y., Martinez, C.P. and McCouch, S.R. 2003. Mapping quantitative trait loci for yield, yield components and morphological traits in an

advanced backcross population between *Oryza rufipogon* and the *Oryza sativa* cultivar Jefferson. Theor. Appl. Genet. 107: 479–493.

Thornsberry, J.M., Goodman, M.M., Doebley, J., Kresovich, S., Nielsen, D. and Buckler, E.S. IV. 2001. Dwarf8 polymorphisms associate with variation in flowering time. Nat. Genet. 28: 286–289.

Till, B.J., Colbert. T., Tompa, R., Enns, L.C., Codomo, C.A., Johnson, J.E., Reynolds, S.H., Henikoff. J.G., Greene, E.A., Steine, M.N., Comai, L. and Henikoff, S. 2003. High-throughput TILLING for functional genomics. In: E. Grotewold (Ed.), Methods in Molecular Biology, Vol. 236. Plant Functional Genomics: Methods and Protocols, pp. 205–220.

Tikhonov, A.P., Bennetzen, J.L. and Avramova, Z.V. 2000. Structure domains and matrix attachment regions along collinear chromosomal segments of maize and sorghum. Plant Cell 12: 249–264.

Vandepoele, K., Saeys, Y., Simillion, E., Raes, J. and Van de Peer, Y. 2002. The automatic detection of homologous regions (ADHoRE) and its application to micro-collinearity between *Arabidopsis* and rice. Genome Res. 12: 1792–1801.

Vandepoele, K., Simillion, C. and Van de Peer, Y. 2003. Evidence that rice and other cereals are ancient aneuploids. Plant Cell 15: 2192–2202.

van Deynze, A., Nelson, J.C., O'Donoughue, L.S., Ahn, S.N., Siripoonwiwat, W., Harrington, S.E., Yglesias, E.S., Braga, D.P., McCouch S.R. and Sorrells, M.E. 1995. Comparative mapping in grasses. Oat relationships. Mol. Gen. Genet. 249: 349–356.

Wang, S., Liu, N., Peng, K. and Zhang, Q. 1999a. The distribution and copy number of *copia*-like retrotransposons in rice (*Oryza sativa* L.) and their implications in the organization and evolution of the rice genome. Proc. Natl. Acad. Sci. USA 96: 6824–6828.

Wang, S., Liu, K. and Zhang, Q. 2000. Segmental duplications are common in the rice (*Oryza sativa* L.) genome. Acta Bot. Sin. 42: 1150–1155.

Wang, Z.X., Yano, M., Yamanouchi, U., Iwamoto, M., Monna, L., Hayasaka, H., Katayose, Y. and Sasaki, T.1999b. 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.

Ware, D., Jaiswal, P., Ni, J.; Yap, I., Pan, X., Clark, K., Teytelman, L., Schmidt, S.C., Zhao, W., Chang, K., Cartinhour, S., Stein, L.D. and McCouch, S.R. 2002. Gramene, a tool for grass genomics. Plant Physiol. 130: 1606–1613.

Wicker, T., Yahiaoui, N., Guyot, R., Schlagenhauf, E., Liu, Z.D., Dubcowski, J. and 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.

Wilson, W.A., Harrington, S.E., Woodman, W.L., Lee, M., Sorrells, M.E. and McCouch, S.R. 1999. Inferences on the genome structure of progenitor maize through comparative analysis of rice, maize and the domesticated panicoids. Genetics 153: 453–473.

Wing, R.A., Stein, L., Jackson, S., Kudrna, D.A., Yu, Y., SanMiguel, P., Butler, E., Yost, D., Goicoechea, J.L. and Kim, H.-R. 2004. The *Oryza* Map Alignment Project (OMAP): toward a closed experimental system for the Genus *Oryza*. Plant & Animal Genome XII, Poster P354.

Wu, C., Li, X.J., Yuan, W.Y., Chen, G.X., Kilian, A., Li, J., Xu, C., Li, X.H., Zhou, D.-X., Wang, S. and Zhang, Q. 2003. Development of enhancer trap lines for functional analysis of the rice genome. Plant J. 35: 418–427.

Xiao, J.H., Li, J.M., Grandillo, S., Ahn, S.N., Yuan, L.P., Tanksley, S.D. and McCouch, S.R. 1998. Identification of trait-improving quantitative trait loci alleles from a wild rice relative, *Oryza rufipogon*. Genetics 150: 899–909.

Xiao, J., Li, J., Yuan, L. and Tanksley, S.D. 1996. Identification of QTLs affecting traits of agronomic importance in a recombinant inbred population derived from a subspecific rice cross. Theor. Appl. Genet. 92: 230–244.

Xu, F., Lagudah, E.S., Moose, S.P. and Riechers, D.E. 2002. Tandemly duplicated safener-induced glutathione S-transferase genes from *Triticum tauschii* contribute to genome- and organ-specific expression in hexaploid wheat. Plant Physiol. 130: 362–373.

Xu, Y. 1997. Quantitative trait loci: separating, pyramiding, and cloning. Plant Breed. Rev. 15: 85–139.

Xu, Y. 2002. Global view of QTL: rice as a model. In: M.S. Kang (Ed.), Quantitative Genetics, Genomics and Plant Breeding, CABI Publishing, Wallingford, UK, pp. 109–134.

Xu, Y. 2003. Developing marker-assisted selection strategies for breeding hybrid rice. Plant Breed. Rev. 23: 73–174.

Xu, Y., Beachell, H. and McCouch, S.R. 2004. A marker-based approach to broadening the genetic base of rice (*Oryza sativa* L.) in the U.S. Crop Sci. 44: 1947–1959.

Xu, Y., Ishii, T. and McCouch, S.R. 2003. Marker-assisted evaluation of germplasm resources for plant breeding. In: T.W. Mew, D.S. Brar, S. Peng, D. Dawe, and B. Hardy (Eds.), Rice Science: Innovations and Impact for Livelihood, Proceedings of the International Rice Research Conference, 16–19 September 2002, Beijing, China, International Rice Research Institute, Chinese Academy of Engineering, and Chinese Academy of Agricultural Sciences, pp. 213–229.

Xu, Y. and Zhu, L. 1994. Molecular Quantitative Genetics. China Agriculture Press, Beijing, China. 291 p.

Xue, Y. and Xu, Z. 2002. An introduction to the China Rice Functional Genomics Program. Comp. Funct. Genomics 3: 161–163.

Yan, L., Loukoiannov, A., Blechl, A., Tranquilli, G., Ramakrishna, W., SanMiguel, P., Bennetzen, J.L., Echenique, V. and Dubcovsky, J. 2004. The wheat *VRN2* gene is a flowering repressor down-regulated by vernalization. Science 303: 1640–1644.

Yan, L., Loukoiannov, A., Tranquilli, G., Helguera, M., Fahima, T. and Dubcovsky, J. 2003. Positional cloning of the wheat vernalization gene *VRN1*. Proc. Natl. Acad. Sci. USA 100: 6263–6268.

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

Yoshimura, S., Yamanouchi, U, Katayose, Y., Toki, S., Wang, Z.X., Kono, I., Kurata, N., Yano, M., Iwata, N. and 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.

Young, N.D. and Tanksley, S.D. 1989. Restriction fragment length polymorphism maps and the concept of graphical genotypes. Theor. Appl. Genet. 77: 95–101.

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

Zhang, Q. and Yu, S. 1999. Molecular marker-based gene tagging and its impact on rice improvement. In: J.S. Nanda (Ed.), Rice Breeding and Genetics -Research Priorities and Challenges, Science Publishers Inc., Enfield, New Hampshire, pp. 241–270.

Zhou, P.H., Tan, Y.F., He, Y.Q., Xu, C.G. and Zhang, Q. 2003. Simultaneous improvement for four quality traits of Zhenshan 97, an elite parent of hybrid rice, by molecular marker-assisted selection. Theor. Appl. Genet. 106: 326–331.