REVIEW ARTICLE

Structural and functional analysis of rice genome

AKHILESH K. TYAGI*, JITENDRA P. KHURANA, PARAMJIT KHURANA, SAURABH RAGHUVANSHI, ANUPAMA GAUR, ANITA KAPUR, VIKRANT GUPTA, DIBYENDU KUMAR, V. RAVI, SHUBHA VIJ, PARUL KHURANA and SULABHA SHARMA

Department of Plant Molecular Biology, University of Delhi South Campus, Benito Juarez Road, New Delhi 110 021, India

Abstract

Rice is an excellent system for plant genomics as it represents a modest size genome of 430 Mb. It feeds more than half the population of the world. Draft sequences of the rice genome, derived by whole-genome shotgun approach at relatively low coverage (4–6 X), were published and the International Rice Genome Sequencing Project (IRGSP) declared high quality (>10 X), genetically anchored, phase 2 level sequence in 2002. In addition, phase 3 level finished sequence of chromosomes 1, 4 and 10 (out of 12 chromosomes of rice) has already been reported by scientists from IRGSP consortium. Various estimates of genes in rice place the number at >50,000. Already, over 28,000 full-length cDNAs have been sequenced, most of which map to genetically anchored genome sequence. Such information is very useful in revealing novel features of macro- and micro-level synteny of rice genome with other cereals. Microarray analysis is unraveling the identity of rice genes expressing in temporal and spatial manner and should help target candidate genes useful for improving traits of agronomic importance. Simultaneously, functional analysis of rice genome has been initiated by marker-based characterization of useful genes and employing functional knock-outs created by mutation or gene tagging. Integration of this enormous information is expected to catalyze tremendous activity on basic and applied aspects of rice genomics.

[Tyagi A. K., Khurana J. P., Khurana P., Raghuvanshi S., Gaur A., Kapur A., Gupta V., Kumar D., Ravi V., Vij S., Khurana P. and Sharma S. 2004 Structural and functional analysis of rice genome. *J. Genet.* **83**, 79–99]

Introduction

Rice is one of the most important crops for mankind. It feeds nearly half the world's population and accounts for more than 50% of their daily calorie intake (Maclean *et al*. 2002). Although, in the past 30 years, world rice production has doubled due to the introduction of new high yielding varieties and improved cultivation practices, it is still insufficient to cope up with the ever-increasing global demands (Fischer *et al*. 2000; Sasaki and Burr 2000). It is expected that the demand for rice in the world would increase at about 1% per annum from 2001 to 2025, implying that the current average yield of 3.9 tons/hectare has to be raised considerably in order to meet the growing needs (Maclean *et al*. 2002; USDA Foreign Agricultural Service official estimates, http://www.fas.usda.gov/psd/complete_tables/GF-table 13-141htm). This is not an easy task in view of the fact that the land available for cultivation is decreasing due to continuous urbanization and inappropriate land use (Khush 1997; Fischer *et al*. 2000). It necessitates the development of high yielding varieties and also minimizing yield loss due to disease and abiotic stresses (such as drought and salinization). Engineering of plants for traits like yield or resistance to various stresses (biotic and abiotic) requires a thorough understanding of the cellular and functional aspects of the plant, which is dictated by its genetic make up. Thus, it is essential to identify all the

*For correspondence. E-mail: akhilesh@genomeindia.org.

Keywords. annotation; comparative genomics; functional genomics; genetic mapping; genome sequencing; physical mapping; rice.

genes and understand their function as well as networking. This idea invoked the interest of researchers who made a contribution in setting the pace for studies in rice genomics.

Beyond its importance as the world's premier crop, rice is also an excellent model plant for genomics, second only to *Arabidopsis* (Izawa and Shimamoto 1996; Tyagi *et al*. 2003). Among cereals, rice has the smallest genome with an estimated size of 430 megabase pairs (Mbp) as compared to the significantly large genome sizes of sorghum, maize, barley, and wheat (about 750, 3000, 5000, and 16000 Mbp, respectively). This also gives it a relatively higher gene density. According to some earlier estimates, if the number of genes in each of the cereal genomes is considered to be about 30,000, rice will have an average of approximately one gene every 15 kilobase pairs (kbp), while maize and wheat will have one gene every 100 and 500 kbp, respectively (Goff 1999). However, recent data place the number of genes to about 50,000 (Goff *et al* 2002; Yu *et al*. 2002), reflecting even higher gene density. Moreover, rice contains relatively less repetitive DNA (Moore *et al*. 1995; Gale and Devos 1998a). Other factors that aid in the use of rice as a model plant species include the fact that it can be regenerated from protoplasts and can also be transformed by exogenous DNA, employing methods like *Agrobacterium tumefaciens*-mediated transformation or particle bombardment, making it an easy target for genetic manipulation among cereals (Tyagi and Mohanty 2000). This tool permits geneticists to complement mutations or to confer dominant phenotypes to verify gene function. In addition, rice has a vast germplasm of cultivated and wild species (Nakagahra *et al*. 1997; Vaughan *et al*. 2003).

With the widespread studies being performed on rice, a number of genetic markers have been mapped aiding in the development of a comprehensive genetic and physical map of rice (Chen *et al*. 2002). Such information has also been used for extensive comparative mapping studies, which have established that the gene order is significantly conserved between rice chromosomes and other cultivated cereals (Schmidt 2000), thus suggesting that rice could provide a road map for the characterization of larger genomes like that of maize, barley and wheat. Hence, rice has been put forward as a model for crop plants, allowing valuable comparisons to a model dicotyledon (*Arabidopsis thaliana*) and the most important monocotyledons like maize, wheat and barley (Schmidt 2000; Bennetzen 2002).

Furthermore, the release of four versions of the genome sequence from two subspecies of *Oryza sativa* over a short span of time (Buell 2002b; Delseny 2003) has brought rice to the forefront of all genomic studies. The importance of rice as a food crop and as a model plant especially due to its syntenic association with other cereal crops, as already mentioned, had a profound effect on rice genomics. The vast amount of structural and functional data of the rice genome generated earlier (Sasaki 1998; Sasaki and Burr 2000) provided a strong backbone for its sequencing efforts (Delseny *et al*. 2001; Sasaki 2002). One of the projects aimed at the latter was the International Rice Genome Sequencing Project (IRGSP) that was started in 1998 (Sasaki and Burr 2000). It was able to release the sequence data up to the Phase 2 level (ordered regions of sequences, which can contain gaps) in December 2002 (http://rgp.dna.affrc.go.jp/rgp/Dec18_NEWS.html; http://www.tigr.org/new/press_ release_12-18-02.shtml). The other rice genome projects included those being carried out by Monsanto (St Louis, MO, USA; Barry 2001) and Syngenta (Torrey Mesa Research Institute, San Diego, USA; Goff *et al*. 2002). These agribusinesses and the IRGSP focused on the sequencing of the *japonica* cultivar Nipponbare with its already available molecular and genetic resources. The fourth project, that of the Beijing Genomics Institute (BGI; http://btn.genomics.org.cn/rice/; Yu *et al*. 2002), on the other hand, had taken up the sequencing of the tropical *indica* cultivar 93-11.

The sequencing of the rice genome alone, however, is not the ultimate goal of genomic research. It is important to understand the involvement of the genome in the functioning of the organism. Functional genomics comes into play to shed light on this aspect (Arber 2002). Genomic DNA sequences can be searched for the identification of open reading frames or signals for their regulation. To elaborate on their function, however, both classical and, more importantly, reverse genetics studies are required. The results can then be compared to other organisms by searching for sequence homologies, followed by appropriate function and expression analyses (Harris 2002). The latter information is an important prerequisite for biotechnological applications, particularly for the improvement or engineering of crops.

This review presents an overview of the scientific progress made in the field of rice genome analysis, both at the structural and the functional levels.

Mapping of the rice genome

Cytogenetic Mapping

Genome mapping is the appropriate method for defining the relative positions of different features of genome. This can be done in two ways – by scrutinizing the chromosome visually or by analysing through experimentation. Conventionally, cytogenetic maps are solely based on photomicroscopic examination of chromosomes. *In situ* hybridization is one of the most effective methods to analyse the essential characteristics of genomes. This is mainly based on the structural constituents of the genome, which are sometimes over-looked by biological means. The development of new techniques like fluorescence *in situ* hybridization (FISH) and genomic *in situ* hybridization (GISH) has made it possible to generate a quality rice chromosome map (Heng *et al*. 1997). FISH facilitates the mapping of Restriction

fragment length polymorphism (RFLP) markers, rDNA loci, Yeast artificial chromosome (YAC) and Bacterial artificial chromosome (BAC) clones on rice chromosomes (Jiang *et al*. 1995; Fukui and Ohmido 2000), while GISH provides a unique means to identify or paint a chromosome complement belonging to a specific genome. The latter has been applied for the determination of the D genome chromosomes in rice (Fukui *et al*. 1997; Fukui and Ohmido 2000), as well as to identify the distribution of the A, B and C genomes in rice somatic hybrids (Apisitwanich 1999; Shishido *et al*. 1998) and also to determine the genomic constitution of some rice tetraploid species (Li *et al*. 2000).

An important modification of the FISH technique involves the use of highly decondensed nuclei or straightened DNA fibers for FISH (de Jong *et al*. 1999). This method provides a high-resolution map as depicted by studies done on the molecular organization of telomere ends (Ohmido *et al*. 2001), and in the physical mapping of unique nucleotide sequences or BACs on the rice genome (Jiang *et al*. 1995; Ohmido *et al*. 1998). However, most of these experiments involved the use of mitotic metaphase chromosomes. Cheng *et al*. (2001a, b) on the other hand utilized meiotic pachytene chromosome-based karyotypes for both cytological and physical mapping of the rice genome. FISH and Fiber-FISH have also been used to dissect anatomical features of rice chromosome centromeres and they were found to be marked by a satellite repeat and a centromere-specific retrotransposon (Cheng *et al*. 2002).

Molecular Genetic Mapping

Linkage study is a fundamental aspect for plant genetics. Several rice linkage maps based on genes for morphological and physiological traits have been constructed in the past half-century after the twelve linkage groups corresponding to the haploid number of rice chromosomes were proposed (*see* Khush and Brar 2001). The linkage groups have been related to the rice chromosomes by primary trisomics and reciprocal translocations (Iwata 1986; Khush and Singh 1985; Khush and Brar 2001). The centromere positions and the orientation of these linkage groups have also been determined using secondary and telotrisomics, thus giving the correct orientation of the rice linkage map leading to a useful resource for genetic studies (Khush *et al*. 1996; Singh *et al*. 1996). Moreover, various studies have been carried out to establish a successful linear correspondence between the cytological, classical and molecular linkage markers for the development of a more comprehensive molecular linkage map of rice (Khush *et al*. 1996), which could be advantageous to investigations related to synteny amongst the cereals and, in turn, to breeding of improved crops.

Extensive mapping studies have been carried out in rice over the years with the advent of molecular genetic markers such as restriction fragment length polymorphism (RFLP), amplified fragment length polymorphism (AFLP), random amplified polymorphic DNA (RAPD), microsatellite or simple sequence length polymorphism (SSLP), and cleaved amplified polymorphic sequence (CAPS) markers (*see* Mohan *et al*. 1997). Several genetic maps of rice have been generated and integrated with phenotypic markers (Nagamura *et al*. 1997; Yoshimura *et al*. 1997). The first RFLP map with 135 loci covering 1389 cM was prepared using an F2 population (50 individuals) derived from a cross between an *indica* (IR34583) and a *javanica* (Bulu Dalam) variety (McCouch *et al*. 1988). Another RFLP map from 144 F2 plants resulting from a cross between an *indica* (Kasalath) and a *japonica* (F1134) variety covered 1836 cM with 347 loci (Nagamura *et al*. 1997). Further improvement resulted in a map of 1491 cM, providing one marker per 2.1 cM, based on 726 markers including 11 microsatellite markers, 3 telomere markers, 11 isozymes, 26 cloned genes and 47 morphological mutant markers. For this purpose, 113 plants derived from a backcross between cultivated rice, *O. sativa*, and wild African rice species, *O. longistaminata*, were used.

A population of 186 F2 individuals from a cross between a *japonica* (Nipponbare) and an *indica* variety (Kasalath) has also been used to construct genetic maps of high quality. At first 1383 DNA markers at an average distance of 1.1 cM (~300 kb) were aligned along 1575 cM on 12 linkage groups (Kurata *et al*. 1994). Most of the mapped clones were sequenced and about 260 rice genes, including 57 putative ribosomal protein genes (Wu *et al*. 1995), were identified by homology search against the Protein Identification Resources (PIR) database. Similarly, most of the random genomic markers and RAPDs have been converted into Sequence Tag Sites (STS) (Monna *et al*. 1994; Inoue *et al*. 1994; Fukuoka *et al*. 1994). In addition, four DNA clones having a telomere-associated sequence were also mapped and located on both distal ends of chromosome 11 and one end of both chromosome 5 and 12 (Ashikawa *et al*. 1994). The next genetic map harbored 2275 markers spanning 1521.6 cM (Harushima *et al*. 1998). To analyse centromeric locations on 12 linkage groups, dosage analysis of secondary and telotrisomics using >130 DNA markers was carried out and markers were located on respective chromosome arms. Most of the RFLP markers in this informative genetic map were expressed sequence tagged (EST) clones from Nipponbare callus, root and shoot libraries.

By now, 3267 markers have been used to produce the rice genetic map (http://rgp.dna.affrc.go.jp/publicdata/geneticmap2000/index.html). In addition, 332 PCR-based markers, including 161 STS markers and 171 cleaved amplified polymorphic sequence (CAPS) markers, have also been mapped on the rice chromosomes (http://rgp.dna.affrc.go.jp/publicdata/caps/index.html). The STS markers were developed using clone-specific sequences (3′ end) designed from

ESTs derived from cDNA libraries of rice (Yamamoto and Sasaki 1997). EST mapping using a physical map of YAC clones aided in assigning the chromosomal location of each marker (Wu *et al*. 2002). The CAPS markers, including 6 derived CAPS (dCAPS), on the other hand, were developed by carrying out restriction digestion of the genome that was amplified using unique primer pairs designed on the basis of 5′ and 3′ sequence data for the clones used for RFLP linkage analysis. The chromosomal location of the markers was confirmed by linkage analysis using 14 randomly selected F2 plants (Harushima *et al*. 1998; http://rgp.dna.affrc.go.jp/publicdata/caps/index.html).

All the above mentioned studies have contributed immensely to the development of over 6000 DNA markers in rice with approximately one marker every 0.25 cM, or every 75–100 kb. In addition to RFLP, RAPD and EST markers, which have been mostly used for mapping, around 500 simple sequence repeat (SSR) markers have also been genetically mapped (Akagi *et al*. 1996; Chen *et al*. 1997; Temnykh *et al*. 2000 2001). Recently, 2240 new SSR markers have been added to the list (McCouch *et al*. 2002). These were developed by PCR amplification using primer pairs designed for SSR-containing sequences released by the Monsanto rice genome sequencing effort and a few from sequences released by the IRGSP. The markers were aligned to rice chromosomes by electronic-PCR (e-PCR) that helped in identifying the BAC or PAC clones containing the marker and their association with another genetically mapped marker. The effort has resulted in providing a total of 2740 SSR markers for rice, which is approximately one SSR every 157 kb (McCouch *et al*. 2002). The use of SNP markers has also started as it is expected to provide richest source of DNA based markers since in rice one SNP per 89 bp among various genotypes or one SNP every 232 bp between two randomly selected strains has been reported (Nasu *et al*. 2002).

Physical Mapping

Physical mapping of the rice genome has received almost as much attention as genetic mapping. Various libraries consisting of large insert clones in vectors such as YACs (Burke *et al*. 1987), P1-derived artificial chromosomes (PACs; Ioannou *et al*. 1994) and BACs (Shizuya *et al*. 1992), and the availability of anchored DNA markers from genetic maps greatly facilitated the construction of physical maps.

A rice YAC library consisting of about 7000 YAC clones with an average insert size of 350 kb (Umehara *et al*. 1995) was used for the construction of a physical map of the rice genome. The YAC clones were initially mapped on the rice chromosomes by chromosome landing using the 1383-marker genetic linkage map developed by Kurata *et al*. (1994). The information was used for developing the first-generation physical maps of individual rice chromosomes (Antonio *et al*. 1996; Saji *et al*. 1996; Shimokawa *et al*. 1996; Umehara *et al*. 1996 1997; Wang *et al*. 1996; Koike *et al*. 1997; Kurata *et al*. 1997; Tanoue *et al*. 1997). On the whole, 2443 YACs formed 537 contigs and islands equivalent to 216 Mb of the rice genome, i.e. 52% of the genome. The map was updated by Saji *et al*. (2001), who utilized 1439 markers out of the 2275 markers on the genetic map developed by Harushima *et al*. (1998) to select 1892 clones from the YAC library. The YACs were used to create a physical map comprising of 297 contigs and 142 islands corresponding to ~270 Mb of the 430 Mb rice genome, yielding a coverage of about 63%.

This YAC physical map, integrated with the genetic map, has been used for positional cloning of several genes, such as the bacterial blight resistance gene, *Xa-1* (Yoshimura *et al*. 1996), the rice blast resistance gene, *Pib* (Wang *et al*. 1999) and the gibberellin-insensitive dwarf mutant gene, *d1* (Ashikari *et al*. 1999), as well as for the chromosomal assignment of 6591 ESTs to generate a rice transcript map covering 80.8% of the rice genome (Wu *et al*. 2002). Additionally, it has been utilized in the IRGSP as a backbone for the construction of a sequence-ready physical map using PACs and BACs. In spite of the usefulness of the YAC physical map, disadvantages such as instability, high chimerism and difficulty in purification of YAC DNA have limited its use. Thus, BACs and PACs have been used more to prepare physical maps of rice.

A genome-wide, BAC-based physical map of *indica* rice (cv. Teqing) was developed by Tao *et al*. (2001). They utilized 21,078 random BAC clones from three complementary libraries having average insert sizes of 130, 150 and 147 kb (Zhang *et al*. 1996), which were mapped to the rice chromosomes using DNA markers from the rice genetic maps developed by Cornell University (Causse *et al*. 1994) and the RGP (Harushima *et al*. 1998). They used the DNA sequence electrophoresis-based fingerprinting method. The map comprised of 298 contigs covering 419 Mb physical length, thus covering ~97% of the rice genome. This physical map has provided a good resource for genomics research of rice, including genome sequencing, effective map-based cloning, mapping of ESTs and target DNA marker development. It can also be used as a framework within which comparative genomics research between the two rice subspecies, viz. *indica* and *japonica*, and between rice and the other cereals can be performed. Recently, physical mapping data from *indica* and *japonica* rice was used to prepare a fine physical map of rice chromosome 4 (Zhao *et al*. 2002).

Recently, Tao *et al*. (2002) have developed one large-insert BIBAC library and three BAC libraries for *japonica* rice *cv*. Nipponbare. Each library has 23,040 clones, has an average insert size of 130 kb, 170 kb, 150 kb and 156 kb, and covers 6.7 X, 8.7 X, 7.7 X and 8.0 X rice haploid genomes, respectively. On the whole there are 92,160 clones

covering 31.1 X rice genome. The BIBAC library is the first large insert plant-transformation-competent library that will greatly aid map-based cloning, functional analysis of the rice genome sequence and molecular breeding in rice and other grass species. The libraries are also being used in the development of a whole-genome BAC-/BIBAC-based, integrated physical, genetic and sequence map of rice and in research of genome-wide comparative genomics of grass species.

Presently, emphasis is being laid to integrate cytological, genetic and physical maps. Cheng *et al*. (2001a), for instance, used chromosome-specific BACs as cytological markers for chromosome identification in rice. The fluorescent-labeled BACs were hybridized to pachytene chromosomes of rice, along with a rice centromere specific DNA probe, to construct a standardized rice karyotype that was also fully integrated with the saturated genetic linkage map of rice (Harushima *et al*. 1998). Pachytene FISH of BACs anchored by genetically mapped RFLP markers, used for generating the map, has also been applied to reveal genetic and physical relationships in specific chromosomal regions as well (Cheng *et al*. 2001b). Meiotic pachytene chromosomes were used for these mapping experiments, which exhibited a better resolving power in comparison to metaphase and interphase FISH mapping (Jiang *et al*. 1995). A BAC fingerprint-based physical map of the rice genome was also integrated with its genetic map by Chen *et al*. (2002). In comparison with the map of *indica* rice (Tao *et al*. 2001), which has limited genetic information, this map constructed for *japonica* rice cultivar Nipponbare consists of 458 contigs of 62,509 BAC clones. Out of these, 284 have been correlated to the genetic map of rice (Harushima *et al*. 1998). The map covers 90.6% of the rice genome. Such a complete physical-genetic map can facilitate the whole genome sequencing of the plant which can be of great importance for studying the organization and function of the genome.

Genome-wide sequencing

The DNA sequence of the entire genome constitutes the ultimate objective of physical mapping. It provides the most detailed description of an organism's genome and can act as a bridge between the structural and the functional phases of genomics. With the advances in sequencing strategies, including automation and the vast input of computational biology, there has been accelerated accumulation of sequence data of many organisms. Genome sequences of several higher organisms, including *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, *Drosophila melanogaster*, *Arabidopsis thaliana* and also *Homo sapiens* have been generated in separate genome projects (Mewes *et al*. 1997; The *C. elegans* Sequencing Consortium 1998; Adams *et al*. 2000; The *Arabidopsis* Genome Initiative 2000; The International Human Genome Mapping Consortium 2001; Venter *et al*. 2001). These are significant milestones in the sequence-based era of genomic research. Another step forward in the ladder of genome projects is the rice genome sequencing project.

The IRGSP, started in 1998, had the advantage of almost a decade's work on rice genomics, such as large-scale cDNA analysis, construction of fine-scale RFLP map, and physical mapping of the rice genome, to fall back upon (*see* Sasaki 1998). The IRGSP, that involves ten countries, viz. Brazil, China, France, India, Japan, Korea, Taiwan, Thailand, UK and USA, has used single variety of rice, *Oryza sativa* ssp. *japonica* cultivar Nipponbare, as the common source of DNA to avoid any allelic polymorphism (Figure 1; Sasaki and Burr 2000). The sequencing strategy of the IRGSP was a genetically-anchored physical map-based clone-by-clone shotgun approach, using large-insert genomic libraries constructed in bacterial artificial chromosomes (BACs) and P1-derived artificial chromosomes (PACs), which had already proved its utility in the case of *Arabidopsis thaliana* (The *Arabidopsis* Genome Initiative 2000) and the human genome (The International Human Genome Mapping Consortium 2001). Genomic BAC/PAC libraries of the *Oryza sativa* ssp. *japonica* cultivar Nipponbare have been used as the primary sequencing templates. Fingerprinting and physical mapping were used to make minimal tiling paths, which was followed by end sequencing, connecting and extending of contigs to obtain the genome sequence (Eckardt 2000; Sasaki and Burr 2000; Chen *et al*. 2002). This allowed genetically anchored sequence to be generated with an ultimate aim of providing whole genome sequence in public domain at an accuracy of better than 99.9%. In all, three BAC libraries (*Eco*RI, *Hin*dIII, *Mbo*I), one PAC library (*Sau*3A1) and two plasmid libraries (*Hae*III, *Sau*3A1) are being used (Baba *et al*. 2000; Chen *et al* 2002; Yang TJ *et al*. 2003). According to the procedure, the extracted DNA from individual BAC or PAC clones (ranging between 100 to 200 kb) from a sequence-ready contig is sheared by sonication or hydroshearing, and the fragments subcloned to produce shotgun libraries with average insert sizes of 2 kb or 5 kb. Random clones from the shotgun libraries were then sequenced at both ends to obtain the desired degree of coverage of about ten times of the total sequence. The resulting sequences were then assembled using computer software, like PHRED (Ewing *et al*. 1998; Ewing and Green 1998), PHRAP (www.phrap.org) and CONSED (Gordan *et al*. 1998), to reconstitute the smaller DNA fragments into the intact insert in the BAC/PAC clone.

The project was completed up to the phase 2 level (high throughput sequence with at least 10X coverage, assembled into ordered and oriented contigs, with or without gaps) in December 2002, thus providing a high-quality draft genome sequence of rice. The data are available in the public domain, deposited in databases such as GenBank,

the European Molecular Biology Laboratory (EMBL) and the DNA Databank of Japan (DDBJ), thus being accessible to researchers and scientists all over the globe (http://rgp.dna.affrc.go.jp/rgp/Dec18_NEWS.html; http://www.tigr.org/new/press_release_12-18-02.shtmal). The IRGSP has generated draft sequence of 3,446 BAC/PAC clones representing ~464 Mb of sequence including overlaps (as of 6 November 2003; http://rgp.dna.affrc.go.jp/cgi-bin/statusdb/irgsp-status.cgi). Members of IRGSP from Japan, China and USA have already published the phase 3 sequences of chromosome 1, 4 and 10, respectively (Sasaki *et al*. 2002; Feng *et al*. 2002; The Rice Chromorome 10S equencing Consortium 2003). Their main features are summarized in table 1.

While the IRGSP was pacing towards its target in a systematic way, in April 2000, a private sector company Monsanto claimed that in collaboration with Leroy Hood's group in Seattle, USA, a working draft of the rice genome has been produced (Barry 2001). The Monsanto Rice Genome Sequencing Project sequenced 3391 BAC clones of the cultivar Nipponbare with approximately 5 X coverage, yielding 399 Mb sequence. This sequence is presently organized into 52,202 contigs representing 259 Mb of assembled sequence, which is ~60.2% of the rice genome. The BACs have been ordered by anchoring them on the genetic map with 450 RFLP probes (Barry 2001). The release of the sequence draft of the rice genome by Monsanto in early 2000 was seen as a major breakthrough in genomics since this was the first crop genome to be described in such detail. To add to its effect on rice genomics and breeding, Monsanto agreed to share the data with the scientific community all over the world under an agreement. Furthermore, they also made their sequence and BAC clones available to the IRGSP to accelerate the public effort (Barry 2001; http://www.biotechinfo.net/rice_breakthrough.html).

Drafts of the rice genome sequence have also been completed and published by Beijing Genomics Institute (BGI; public sector; Yu *et al*. 2002) and Syngenta (Torrey Mesa Research Institute, San Diego, USA; private sector; Goff *et al*. 2002), who have worked on ssp. *indica* cultivar 93-11 and ssp. *japonica* cultivar Nipponbare, respectively. Both the groups followed the whole-genome shotgun sequencing approach. A brief summary of both draft sequences is given in table 2. An interesting outcome of rice genome

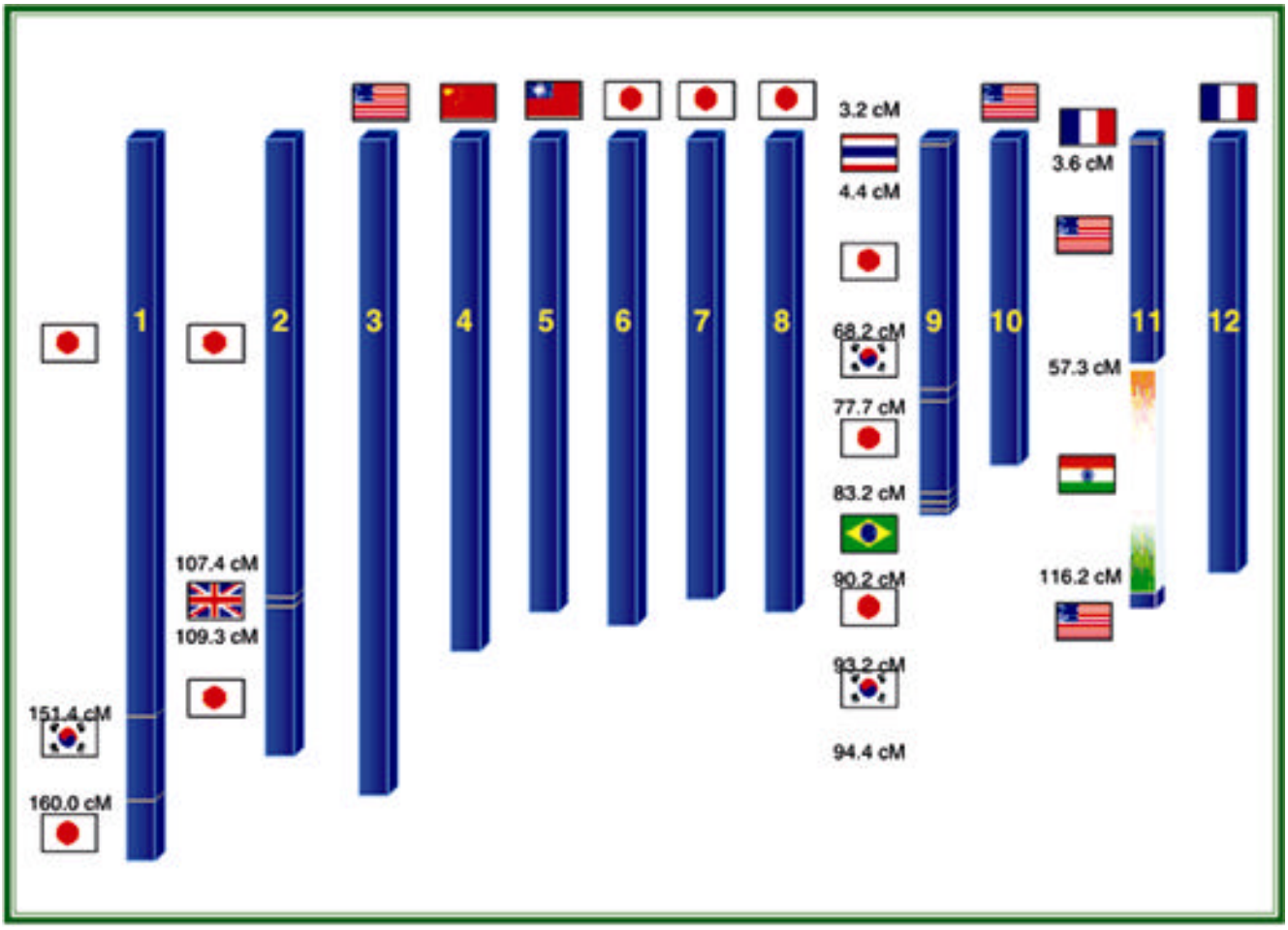

Figure 1. Sharing of rice chromosomes by member nations of IRGSP for sequencing.

Table 1. General structural features of the completely sequenced rice chromosomes.

	Chromosome 1	Chromosome 4	Chromosome 10
Non-overlapping sequences	43.2 Mb	34.6 Mb	22.4 Mb
Total length of gaps	2.5 Mb (8 gaps)	885 kb (7 gaps)	1 Mb (7 gaps)
Base composition (%GC)			
Overall	43.8	44.16	43.5
Coding	58.2	53.0	53.6
Noncoding	40.7	40.5	40.3
Predicted gene number	6,756	4,658	3,471
Gene density	6.4 kb/gene	7.4 kb/gene	6.4 kb/gene
Gene size	3.4 kb	2.77 kb	2.5 kb
Exons			
Size of exons (bp)	229	340	344
Number of exons per gene	4.8	4.4	4
Introns			
Size of introns (bp)	605	376	389
Number of introns per gene	3.8	3.4	3.0
Repetitive sequences	13%	18.2%	18%

Table 2. A comparative account of drafts of the rice genome sequence.

	Rice subspecies		
	ssp. *japonica*[1]	ssp. *indica*[2]	ssp. *japonica*[3]
Sequence generated	390 Mb	362 Mb	464 Mb*
Gene number	32,000–50,000	46,022–55,615	62,435**
Coverage	> 6 X	~ 4 X	> 10 X
Contig number	42,109	127,550	88[#]

[1]Goff *et al*. (2002); [2]Yu *et al*. (2002); [3]IRGSP release.
*including overlapping sequences; **within 366 Mb of non-overlapping sequence; [#]physical gaps notwithstanding sequencing gaps.

analysis is the discovery of unique GC content gradient not present in eudicots (Wong *et al*. 2002). Negotiations allowed the IRGSP to use the sequence data produced by Syngenta as well to complement their efforts.

All these draft sequences are a rich resource for many genomic experiments, but they have a major limitation in terms of the incomplete nature of the sequence. Further, studies using these drafts could be restricted due to ambiguities or incompleteness related to the experimental strategy, the occurrence of sequencing gaps, especially in the gene of interest, and due to the absence, under-representation or misassembly of certain regions, such as the telomeres, centromeres or regions having secondary structures. Thus, it is important to have finished genome sequence of the highest quality (Buell 2002a; Mardis *et al*. 2002). The IRGSP has thus entered the finishing phase, which includes gap-filling and quality improvement in order to meet the adopted finishing standards (http://demeter.bio.bnl.gov/Guidelines.html). The final 'finished' sequence data, i.e. refined data without gaps, is expected by the end of 2004 (http://www.tigr.org/new/press_release_12-18-02.shtml), setting the stage for future research on rice like finished sequence of human genome (Collins *et al*. 2003a, b).

Annotation and bioinformatics

The vast accumulation of raw sequence data has also elevated the importance of transforming these data into a form that is accessible to biologists, who can study and interpret it into useful biological information, a process termed annotation (Lewis *et al*. 2000). Broadly, annotation describes the genome by identifying various sites and segments along the sequence involved in the functioning of the genome. The main objective is definitely the positioning of the genes and the elements related to the functioning of these genes or their products (Rouzé *et al*. 1999). Some early rice genome fragments were annotated manually (Mayer *et al*. 2001). Today, however, genome annotation is performed using a combination of experimental and computational methods, with the latter forming a significant

part of the expanding field of bioinformatics. Computational tools or software are specially designed to accommodate the large amounts of genome-scale datasets of both prokaryotes and eukaryotes (Lewis *et al*. 2000; Fortna and Gardiner 2001; Pertea and Salzberg 2002). Automated procedures, involving minimal manual analysis have been used for annotation of the majority of available rice data (Schoof and Karlowski 2003), thus allowing simultaneous annotation of the sequences as they are released.

Many bioinformatics resources are now available to researchers around the world through the World Wide Web. A recently developed informatics resource specifically dedicated to rice is the MIPS Rice (*Oryza sativa*) database or the MOsDB (http://mips.gsf.de/proj/rice; Karlowski *et al*. 2003). This provides rice genome sequence as well as annotation information along with a complex characterization of all annotated rice genes. Another database allowing the dissection of the rice genome has been provided by The Institute for Genomic Research (TIGR). Their resource (http://www.tigr.org/tdb/e2k1/osa1/; Yuan *et al*. 2003) presents data generated by the analysis of the rice sequence data of the IRGSP. This includes the anchoring of rice BAC/PAC clones to the genetic map, annotation of the sequences of these clones for gene content, identification of motifs and domains within the predicted genes and proteins, and the identification of related sequences in other plant species. Furthermore, an automated annotation system and database for rice genome sequence, RiceGAAS (the Rice Genome Automated Annotation System) is also being extensively used to identify various structural and functional components (Sakata *et al*. 2002; http://RiceGAAS.dna.affrc.go.jp/). The system automatically collects the genome sequence data and executes the prediction of genes and long terminal repeats on the basis of multiple gene prediction programs and homology search results. It integrates results from several gene prediction software such as GENSCAN (Burge and Karlin 1997), FGENESH (Salamov and Solovyev 2000), RiceHMM (Sakata *et al*. 1999) and MZEF (Zang 1997) for mining the protein coding genes. Monocot specific version of these programmes are utilized for accurate predictions. Moreover, output from SplicePredictor programme (Kleff *et al*. 1996), which exclusively analyses presence of splice sites, is integrated, to further confirm the exon-intron boundaries. Prediction of tRNA genes is performed by software tRNAscan-SE (Lowe and Eddy 1997). The predicted genes are then searched against the nucleotide and protein databases using BLAST and the result parsed for further manual analysis. In addition to this, RiceGAAS also analyses the predicted protein for the presence of any functional domains (ProfileScan, http://hits.isb-sib.ch/cgi-bin/PFSCAN and MOTIF, http://motif.genome.ad.jp/), intracellular localization (PSORT, http://psort.ims.u-tokyo.ac.jp) and solubility (SOSui, http://sosui.proteome.bio.tuat.ac.jp/sosuimenu0.html). The RepeatMasker software (http://ftp.genome.washington.edu/cgi-bin/RepeatMasker) is used to delineate various repeat elements present in the rice sequences. Thus, RiceGAAS provides a comprehensive view of all the important genes in the sequenced clone representing a particular portion of the rice genome.

Such tools have been used to annotate and analyse the three whole genome draft sequences (two of *japonica* and one of *indica*) of rice. Gene prediction based on the whole genome draft sequences gives a highly variable number of genes in the rice genome (see table 2). This could be because the drafts are at a different level of accuracy and different gene prediction softwares have been used. Fragmented draft sequences greatly affect the accuracy of gene predictions, thus a more clear picture of the gene content of rice would be possible as IRGSP takes the draft sequence to the finishing stage. Already, a highly accurate annotated complete sequences of chromosomes 1, 4 and 10 have been made available (Sasaki *et al* 2002; Feng *et al*. 2002; The Rice Chromosome 10 Sequencing Consortium 2003). As shown in table 1, on an average, every 6–7 kb of the sequence has a gene with an average GC content higher than that in *Arabidopsis*. The distribution of genes is quite variable in all the three chromosomes. Chromosome 1 has a higher gene density on the distal regions of both the arms whereas most of the genes on chromosome 10 are on the long arm. About 53%, 44% and 40% genes predicted on chromosome 1, 4 and 10, respectively, have been classified as hypothetical because they do not show significant homology to any protein or EST in the database. Some of these could represent novel genes, as in case of chromosome 1, where 33% of the hypothetical genes show presence of known protein domains. However, the possibility of inaccurate prediction cannot be ruled out. When compared to the corresponding sequence of the *indica* rice (Yu *et al*. 2002) several discrepancies were observed. On chromosome 4, 47.7% of the genes were predicted correctly in the *indica* sequence draft, while rest were either partial (38.3%) or absent (13.85%). This is probably because *indica* draft sequence has comparatively more gaps, which would significantly affect the process of annotation (Feng *et al*. 2002).

Special resources have been created to provide information about model organisms such as the *Arabidopsis* Information Resource (TAIR; http://arabidopsis.org/; Rhee 2000; Rhee *et al*. 2003), that gives the latest information about the organism and provides further avenues for analysing the available data. Similar databases giving information about recent work on the rice genome are also available (Antonio *et al*. 2000; Yuan *et al*. 2001). These databases are a good compliment to the biological data being revealed by many rice genome projects. The INE (INtegrated rice genome Explorer; Sakata *et al*. 2000; http://www.dna.affrc.go.jp.82/giot/INE.html) site gives integrated map and sequence information of the rice genome. A web interface based on a Java applet allows rapid viewing of the database; DNA markers on the genetic map play key role in linking the YAC physical map, the EST map

and the PAC/BAC physical map to define the chromosomal locations of clones on each map. Annotated genome sequences are available via PAC/BAC clones on the physical map, along with tabulated, annotated information for predicted genes, ESTs assigned by BLASTN, protein assigned by BLASTX, and so on. The 'oryzabase' (http://www.shigen.nig.ac.jp/rice/oryzabase/top/top.jsp), developed at the National Institute of Genetics in Mishima, Japan, as a part of a huge collection of biological germplasm databases, is a comprehensive database containing information from classical rice genetics as well as genomics. A similar database working as a community resource for rice and having useful information for comparative genome mapping of the grasses is Gramene (Ware *et al*. 2002a,b; http://www.gramene.org/). This website is a significant portal for all information related to cereal genetic maps, genomic and EST sequences, genetic markers, genes, mutants, proteins, controlled vocabularies and publications.

Many centres involved in rice research have also made their resources available on the web. The rice genome-sequencing project at Arizona Genomics Institute (AGI) provides a database of the BAC end sequences as well as a fingerprint of the complete BAC library of *Oryza sativa* cultivar Nipponbare (http://www.genome.arizona.edu/). At TIGR (The Institute of Genome Research, http://www.tigr.org/tdb/rice) the genetic markers have been mapped on the BAC library thereby linking the genetic and physical maps. TIGR has also developed two repetitive sequence databases for rice both of which can be scanned using BLAST (http://www.tigr.org/tdb/e2kl/osal/blastsearch.shtml). One of these databases is developed by analysing the rice sequences available in GenBank, while the other includes repetitive sequences identified from the BAC end sequences using the MUMmer software (Delcher *et al*. 1999).

Another excitement in bioinformatics is the analysis of vast amount of genome expression data, which is being generated by the recent technologies like microarrays and SAGE (Gerstein and Jansen 2000). Certain databases, such as the Rice Expression Database (RED; http://red.dna.affrc.go.jp/ RED/; Yazaki *et al*. 2002) and the Rice Microarray Opening Site (RMOS; http://microarray.rice.dna.affrc.go.jp/; Yazaki *et al*. 2002), have been created to allow the processing and mining of expression data. These, however, are only a few examples out of the rice informatics infrastructure that is making the genomic, biological and genetic information of rice useful and relevant for studies involving other cereal species as well.

Comparative genomics

It seems certain that with the sequencing of rice and other model organisms like *Arabidopsis*, followed by the assigning of function to these sequences (drafts), there is a lot of information for applications of genomics in other species as well. This assignment is based on the fact that a remarkable degree of synteny exists between plant species as revealed by several comparative genetic mapping experiments (Schmidt 2000, 2002). Conservation of gene order and content has been detected between *Arabidopsis* and other species within the dicot family, such as the cultivated *Brassica* species (Kowalski *et al*. 1994; Lagercrantz 1998), tomato (Ku *et al*. 2000; Rossberg *et al*. 2001) and soybean (Grant *et al*. 2000). Within the monocots also, especially the cereals, extensive colinearity has been observed by comparative mapping of the genomes using genetic markers. This phenomenon of macro-colinearity was first established between seven grass species, with rice as the reference genome, and was represented in the form of a graphical consensus map that is popularly known as the 'Circle Diagram' (Moore *et al*. 1995). This map has been refined to embrace more grass species whose genomes are described using 25 rice linkage blocks (Gale and Devos 1998a; Devos and Gale; 2000).

All these studies give the general impression that all the grasses examined have similar gene order despite the large differences in DNA content or chromosome number (Devos and Gale 2000; Freeling 2001). Microcolinearity, or the conservation of gene order at the sub-megabase level, is also observed to be extensive but has frequent deviations which can be attributed to small scale rearrangements, deletions, or even local gene amplification and translocation (Bennetzen 2000; Keller and Feuillet 2000; Ware and Stein 2003). This has been examined not only between sorghum and maize, but also between rice and other crop plants as well as between rice subspecies (Song *et al*. 2002; Dubcovsky *et al*. 2001; Han and Xue 2003; Klein *et al*. 2003). The absence of microcolinearity as compared to the recombinational map level, has also been confirmed by comparison of small segments of the rice genome sequence with some cereals (Bennetzen and Ramakrishna 2002; Feuillet and Keller 2002; Bennetzen and Ma 2003). In particular, use of wheat chromosome bin mapped ESTs with rice genome sequence has predicted that order of rice genes in relation to wheat genome could emerge as a complex pattern and its utility for synteny-based analysis / application remains to be assessed (Sorrells *et al*. 2003; Qi *et al*. 2003; Singh *et al*. 2004). Nevertheless, the rice genome has come forth as a relatively stable genome compared to other cereals, which have faced most of the rearrangements during evolution.

Various investigations have also revolved around the idea of colinearity between monocot and dicot plants. However, rice genome being four times larger and containing more then twice the number of genes as that of *Arabidopsis* (The *Arabidopsis* Genome Initiative 2000; Goff *et al*. 2002; Yu *et al*. 2002), may show limited synteny with it (Bowers *et al*. 2003). Devos *et al*. (1999) studied the colinearity between two small regions of *Arabidopsis* chromosome 1 and rice and concluded that synteny between the

two has been eroded to the point that it cannot be identified by comparative mapping. Their inference was supported by the work of Mayer *et al*. (2001) who compared annotated *Arabidopsis* sequence and a segment of rice chromosome 2, and that of Liu and coworkers (2001) who compared the *Arabidopsis* genome and 126 rice BACs corresponding to ~20 Mb sequence. A similar study was performed by Salse *et al*. (2002) who extended the comparison by including 189.5 Mb of rice genome sequence and translated *Arabidopsis* annotated coding sequences. They got results depicting significantly low synteny with ~17% gene conservation in homologous segments, possibly due to polyploidization and genome rearrangements during evolution, as also suggested by earlier reports (Paterson *et al*. 1996; Salse *et al*. 2002). Large scale duplication and reshuffling has been clearly demonstrated in *Arabidopsis* (Blanc *et al*. 2000; Ku *et al*. 2000). The low level of synteny between *Arabidopsis* and rice might not be adequate for applications in map-based cloning strategies as well as for integration of functional and structural genomic data across the monocot or dicot divide (Devos *et al*. 1999), but a detailed study of the genomic data of both the plants could provide answers to questions related to the structure and evolution of genomes (Livingstone and Rieseberg 2002; Schmidt 2002; Bowers *et al*. 2003).

On the other hand, the high level of genome colinearity between plant species belonging to the same family can be exploited to carry out fine mapping and map-based cloning experiments, especially in the case of crop plants having large genomes. As in the cereals, the genetic mapping of an agronomically important locus is carried out with the large genome followed by cloning using information from the closely related model organism, i.e. rice (Foote *et al*. 1997; Kilian *et al*. 1997; Zwick *et al*. 1998; Freeling 2001).

Functional genomics

To comprehend the rapidly accumulating data as well as to understand the functioning of the cell at global level, there is a need for high throughput functional genomics as well. Hieter and Boguski (1997) describe the term functional genomics as the development and application of global or genome-wide experimental approaches to assess gene function by using the information and reagents provided by structural genomics. Several approaches have been used to explore the probable function of the genes, as well as to monitor their expression in relation to various other genes of rice.

Map-based cloning

The genetic analysis of mutants and various naturally occurring genetic variants in traits, such as disease resistance, flowering time and seed size requires plant systems that are easy to handle and are genetically simple with minimal gene redundancy such as *Arabidopsis* (Meinke *et al*. 1998; Lukowitz *et al*. 2000), tomato (Martin *et al*. 1993) and rice (Delseny *et al*. 2001). The genes involved are isolated on the basis of their map positions. The process called map-based or positional cloning involves the mapping of the mutation followed by gene isolation from large YAC/BAC libraries. Mapping narrows down the genetic interval containing a mutation by successively excluding all other parts of the genome. Positional or map-based cloning has been used as a tool to identify important genes in rice essentially by chromosome walking using the available resources and saturating the region of the gene of interest with many markers (RFLP, AFLP, EST, etc.). The *Xa21* gene conferring resistance to leaf blight disease in rice was the first important gene to be isolated by this approach (Song *et al*. 1995). Following this, several other important genes such as those conferring resistance to specific races of the rice blast agent, the fungus *Magnoporthe grisea*, have been cloned and characterized by positional cloning, *viz*. *Pib* (Wang *et al*. 1999), *Pi-ta2* (Bryan *et al*. 2000) and *Pi5(t)* (Jeon *et al*. 2003). Similarly, the bacterial blight-resistance gene *Xa1* and *Xa26 (t)* (Yoshimura *et al*. 1996, 1998; Yang *et al*. 2003), a rice spotted leaf (lesion-mimic) gene, *Spl7*, encoding a heat stress transcription factor (Yamanouchi *et al*. 2002), a semidwarfing gene, *sd-1*, encoding an enzyme involved in the gibberellin biosynthesis pathway (Monna *et al*. 2002), as well as a rice yellow mottle virus resistance gene (Albar *et al*. 2003), have also been isolated.

A lot of work has also been done for the isolation and identification of quantitative trait loci (QTL) by map-based cloning (Yano and Sasaki 1997). Major attention in this respect has been given to the heading date loci in rice, which are involved in photoperiod sensitivity. Map-based cloning has allowed the identification and consequently the analysis and characterization of quite a few of these loci (Yano *et al*. 2000; Takahashi *et al*. 2001; Kojima *et al*. 2002).

Gene Tagging by Insertional Mutagenesis

Identification of genes by insertional mutagenesis is quite advantageous due to the ease of isolating the tagged gene in comparison with functional analysis based on mutations derived from chemical or physical treatments. The process of insertional mutagenesis involves the insertion of a known segment of DNA into a gene of interest. This inserted sequence often creates a 'knockout' mutation by blocking or disrupting the expression of the gene and might result in a mutant phenotype that can be screened. In addition, the insertion sequence also tags the affected gene, which can be isolated by using hybridization probes based on the sequence of the gene tag. Once the mutated gene is known, the initial wild type gene can also be identified. Such a method has a major advantage of not requiring any prior knowledge of the gene product or its expression (Per-

eira 2000). Also, this approach provides a direct route to determine the function of a gene product *in situ* unlike other methods which are correlative and do not necessarily prove a relationship between a gene sequence and its function (Krysan *et al*. 1999). Two types of insertion sequences are commonly used for mutagenesis in case of plants: transposable elements and *Agrobacterium tumefaciens*-mediated T-DNA (transfer DNA) insertions.

T-DNA tag

The process of gene tagging using T-DNA as the insert has been used effectively to isolate genes, especially in *Arabidopsis* (Azpiroz-Leehan and Feldmann 1997; Krysan *et al*. 1999; Alonso *et al*. 2003). T-DNA insertional mutagenesis has also been used to produce 22,090 primary transgenic rice plants having approximately 25,700 taggings (Jeon *et al*. 2000). Another efficient T-DNA tagging system for *japonica* rice has been described recently (Sallaud *et al*. 2003). Over 1000 T- DNA tags in rice genome have been characterized and it revealed preferential insertion in gene rich regions (Chen *et al*. 2003)

Transposon tags

Transposons, first recognized by Barbara McClintock in maize, have become a powerful tool for gene isolation. The mutagenic potential of mobile elements and their ability to tag the mutated sequences along with their widespread distribution have been exploited for use as tools for gene isolation as these properties help in the cloning of genes (Hamer *et al*. 2001). The application of transposon tagging was initially restricted to plants, such as maize (*Zea mays*) and snapdragon (*Antirrhinum*), with active and well-characterized endogenous transposons (Walbot 1992). But, now maize transposon systems have been used for mutagenesis in heterologous transgenic plant species which otherwise lack an active endogenous transposon family (Sundaresan 1996).

The Ac element was introduced into rice and checking for hygromycin resistance identified the transposed plants, since the autonomous Ac element had been cloned between the promoter and the *hph* coding region (Izawa *et al*. 1991; Murai *et al*. 1991). Enoki *et al*. (1999) also studied the effectiveness of the transposon tagging strategy for functional analysis. They analysed about 6000 Ac element containing rice plants using PCR. A total of 559 plants of four transgenic rice families were analysed for three successive generations from R5 to R7. Out of these 18.9% plants were found to contain newly transposed Ac insertions. Transposons were also found to have a preference for protein coding genes. A strategy, using the maize Ac-Ds system, has also been effectively used for gene tagging in case of rice (Greco *et al*. 2001). Apart from this, Shimamoto *et al*. (1993) and Sugimoto *et al*. (1994) have been able to demonstrate the *trans*-activation and stable integration of the non-autonomous Ds element by direct gene transfer, i.e. from a plasmid or viral vector transfected into rice protoplasts having the Ac transposase gene. As compared to the maize Ac-Ds system, other candidate transposable elements, such as Tag1 of *Arabidopsis* and En-Spm of maize, show low activity in rice and thus are not frequently used as insertional mutagens in this plant (Liu *et al*. 1999; Hirochika 2001a).

Retrotransposons, transposable elements that transpose via an RNA intermediate and are structurally similar to integrated copies of retroviruses, have also been shown to be efficient gene tags as demonstrated by the introduction of tobacco retrotransposon *Tto1* into rice and its autonomous transposition through reverse transcription (Hirochika *et al*. 1996). Rice has also been found to have its own retrotransposons and about 32 families have been reported (Hirochika 1997). These can also be activated specifically to carry out insertional mutagenesis in rice itself during tissue culture and regeneration (Hirochika 1999). An eminent example of such a retrotransposon is *Tos17* that has been used as an effective gene tag (Hirochika 1999, 2001a). Transposition efficiency is quite high as in each regenerated plant 5–30 transposed *Tos17* have been reported (Hirochika 1999). This is advantageous over other transposable element tagging as *Tos17* is only active during tissue culture stage, therefore, mutations induced by *Tos17* insertion are fixed and inherited stably in the next generations (Hirochika 1999). Hirochika *et al*. (2001b) were able to clone the genes causing dwarf, semidwarf, viviparous, pale green, brittle and narrow leaf mutant phenotypes in rice. They have further demonstrated the possibility of using *Tos17* for screening mutants by PCR using *Tos17*-specific primers or by sequencing the *Tos17* insertion sites, thus exhibiting its use as a tool for reverse genetics.

Sato *et al*. (1999) were able to identify a mutation in the homeobox gene *OSH15* after screening 550 plants mutagenized by *Tos17*. Similarly, Agrawal *et al*. (2001) identified *Tos17* insertions in the rice zeaxanthin epoxidase gene (*OsABA1*) and in a novel *OsTATC* gene having slight homology with bacterial sec-independent translocase *TATC*, and rice phytochrome A (*phyA*) mutant lines were isolated by Takano *et al*. (2001) using a *Tos17* mutant population. A *Tos17* insertional mutant database is available (http://tos.nias.affrc.go.jp/~miyao/pub/tos17/), which contains information on 47,196 *Tos17*-induced insertion mutants. This database stores data about disruption loci and information regarding phenotypes of insertion lines, along with >40,000 photographic images. BLAST search facility is available for searching insertion lines containing gene of interest (Miyao *et al*. 2003).

Several other transposable elements have also been identified in rice on the basis of sequence analysis (Bureau *et al*. 1996; Mao *et al*. 2000; Turcotte *et al*. 2001). Recently, *in silico* repeat mining of the currently available rice genome sequence data has brought forth new transposable elements

that could serve as tools for gene tagging. These include the first active DNA transposon family in rice and the first active miniature inverted-repeat transposable element (MITE) from any organism, *miniature Ping* (*mPing*) (Jiang *et al*. 2003; Kikuchi *et al*. 2003; Nakazaki *et al*. 2003), the *transposable element of Oryza sativa* (*TEOS1*) (Chao *et al*. 2003) and *Dasheng* (Jiang *et al*. 2002).

Deletion Mutants

Leung *et al*. (2001) have generated a large collection of deletion mutants of rice IR64 by using diepoxybutane, fast neutron and gamma ray mutagenesis. Phenotyping of such mutants is underway and combination of molecular methods like TILLING (McCallum *et al*. 2000) could help pin-point the nature of mutation, thereby helping functional analysis of rice genes.

Expression profile

Another part of functional genomics is the analysis of gene expression. Having knowledge of when and where a gene product, i.e. RNA and / or protein, is expressed can give vital information about the particular gene in question. The very first step in generating a genome-wide expression profile is the preparation of EST profiles. Expressed sequence tags or ESTs are DNA sequences read from either end of cDNA molecules and since cDNAs are prepared from mRNA, these provide information about the expressed part of the genome (Rudd 2003). Thus, EST data sets have been generated on a large scale (Marra *et al*. 1998) even for rice (Yamamoto and Sasaki 1997), which has 260,782 ESTs in the NCBI (National Center for Biotechnological Information) database for ESTs (dbEST) as per release 082903 of August 29 2003. The large number of EST sequences, however, may not be a representation of the number of expressed genes because several of them are redundant. For example, total number of 252,364 sequences (221,715 ESTs + 30,649 mRNA sequences) have been clustered into only 31,080 genes. A minimally redundant set of ESTs provides a suitable substrate for a variety of high throughput techniques used for expression analyses such as microarrays. Such a collection of ESTs could be provided with quality value if ESTs represent an outcome from differential screening in relation to a particular state, e.g. drought or salt stress (Babu *et al*. 2002; Sahi *et al*. 2003). At the same time, 28000 full-length sequences of cDNA reported for rice could help annotation of genes accurately and provide resources for gene discovery and manipulation (The Rice Full-length cDNA Consortium 2003). Other techniques used in expression genomics include traps and the serial analysis of gene expression.

Traps

Classical genetic approaches to identify genes, as mentioned earlier, are generally based on the creation of mutations leading to a recognizable phenotype reflecting the gene function, such as in gene tagging. However, this is not always possible, since many genes show functional redundancy, and thus mutation in one gene or locus could be compensated for by the functioning of one or more other family members. Moreover, certain genes function at different stages of development. Mutations in such genes could cause early lethality or could be highly pleiotropic. This can thus prevent the identification of the role of the gene. Trapping techniques have been developed keeping these limitations in mind.

Entrapment strategies rely on the use of inserts, such as transposons or T-DNA, containing reporter gene constructs, whose expression is dependent on *cis*-acting regulatory sequences at the site of insertion. The inserts then allow for the identification of genes, based on their expression pattern, even though they might not display an obvious mutant phenotype (Pereira 2000; Springer 2000).

Three basic types of gene traps are constructed using reporter genes such as those encoding *b*-glucuronidase (GUS) and green fluorescent protein (GFP): enhancer trap, promoter trap, and gene trap. Ac/Ds transposable element mediated gene trapping system has been demonstrated in rice using the *GUS* reporter gene. Here Ac and Ds gene trap vectors were constructed and introduced into rice genome by *Agrobacterium*-mediated transformation. Simple and single insertion of T-DNA were analysed for the evaluation of gene-tagging efficiency. Nearly 80% Ds elements were excised from the original T-DNA sites, when Ac cDNA was expressed under a CaMV 35S promoter. About 30% of the plants carried at least one Ds which underwent secondary transposition in the later cultures. Eight per cent of transposed Ds elements expressed GUS in various tissues of rice panicles. Half of the Ds insertion sites showed simple hybridization patterns, which were used to locate the Ds element (Chin *et al*. 1999). Recently, another enhancer trap system based on GAL4/VP16-UAS elements with GUS has been developed and used to produce 31,443 independent transformants of *japonica* rice *via Agrobacterium* (Wu *et al*. 2003). Out of 2679 T_1 families field tested, 7.5% showed visible phenotype alterations and 33% of families show 3:1 segregation ratio.

The presence of the promoter-less *b*-glucuronidase (GUS) reporter gene in the binary vector used in the T-DNA tagging allowed gene trapping and thus the monitoring of the expression of the T-DNA tagged gene by histochemical assays (Jeon *et al*. 2000). Out of the total tested organs, 1.6–2.1% were GUS-positive and their GUS expression patterns were organ or tissue specific or ubiquitous in all parts of the plant. The availability of such a large population of T-DNA tagged lines of rice has given weightage to the idea of utilizing the technique for identifying insertional mutants and discovering new genes in rice. Already, such resources have been used to isolate cold-responsive genes in rice (Lee *et al*. 2003).

Another approach used to access gene function is activation tagging. This technique is based on the use of an insertion element carrying a strong enhancer (Walden *et al*. 1994). Thus, on integration into the genome, it causes activation of an adjacent gene or enhances its expression, resulting in gain-of-function mutants. Such an approach has been used extensively in *Arabidopsis* using T-DNA vectors containing multiple cauliflower mosaic virus (CaMV) enhancer sequences (Pereira 2000; Weigel *et al*. 2000) and its use has also been demonstrated recently in rice by Jeong *et al*. (2002). The latter group used a new T-DNA vector having multimerized transcriptional enhancers from the CaMV 35S promoter inserted next to the left border and the promoterless *GUS* reporter gene next to the right border. They were able to generate 13,450 fertile T-DNA insertional lines of rice using this vector. Although the enhancer element was observed to activate nearby genes, only a small number of mutants have been isolated from the tagged lines. Jeong *et al*. (2002) are investigating whether these mutants are a cause of activation tagging. On the whole, however, they have opened up the possibility of using activation tagging for forward genetic screening even in monocots.

Microarrays

Microarrays, which might also be termed as 'reverse northern-dot blots' (Wisman and Ohlrogge 2000), allow the rapid analysis of gene expression on a genome-wide scale. Hundreds to thousands of DNA fragments or oligonucleotides representing various genes spotted on a solid matrix at high density can be analysed simultaneously in single attempt by using fluorescently labelled cDNAs representing mRNA from a particular stage (Duggan *et al*. 1999).

The first plant DNA microarray was constructed for *Arabidopsis* using 48 EST clones to compare the expression patterns between leaves and roots (Schena *et al*. 1995) and highly efficient versions for the same are available. With rice under the spotlight during the genomic era, work on rice microarrays for functional genomics is also being done. RGP initiated the Rice Microarray Project directed towards the functional analysis of all rice genes using this gene expression monitoring system. The National Institute of Agrobiological Resources (NIAR) and The Institute of the Society for Techno-innovation of Agriculture, Forestry and Fisheries (STAFF) jointly conducted this project in collaboration with 64 research institutions all over Japan (Rice Microarray Opening Site or RMOS; http://microarray.rice.dna.affrc.go.jp/). They utilized ESTs generated from sequencing of cDNA clones from libraries derived from the root, shoot, leaf, panicle and also rice callus cultured in different media. Over 9000 clones searched for sequence homology by BLASTN were clustered. These clustered clones were analysed for sequence similarity against public protein and nucleic acid databases. Then BLASTN and BLASTX were performed and clones with high similarity identified putatively. One thousand two hundred sixty five cDNAs in these unique clones were randomly selected and used for the microarray system (Yazaki *et al*. 2000). More than 600 experiments in 25 physiological categories have been carried out using this array as part of the Rice Microarray Project and normalized expression data derived from these have been incorporated in the Rice Expression Database (RED; http://red.dna.affrc.go.jp/RED/; Yazaki *et al*. 2002). An example of such an experiment involves the use of the cDNA microarray to analyse the expression of the 1265 rice genes in rice plants treated with probenazole, a chemical inducer of disease resistance (Shimono *et al*. 2003). This study enabled the group to identify about ten genes involved in plant-pathogen interaction. Another microarray-based study in rice has been carried out by Kawasaki *et al*. (2001), who analysed transcript regulation in a salt-tolerant variety of rice in response to salt stress over a period of time. They used microarrays with 1728 cDNAs from libraries of roots exposed to high salt.

A rice GeneChip microarray covering almost half of the rice genome has been created by Syngenta based on the genome sequence of the rice cultivar Nipponbare (Goff *et al*. 2002; Zhu *et al*. 2003). The chip has 25-mer oligonucleotides representing 21,000 genes of rice. Zhu *et al*. (2003) used this microarray to study gene expression during different stages of rice grain filling and to identify genes involved in the synthesis and transport of carbohydrates, proteins and fatty acids. In another study, Cooper *et al*. (2003) used the same chip for investigating expression profile during stress response and seed development in rice.

Serial analysis of gene expression

Different expression profiling assay are capable of analysing limited numbers of transcript species already known and are unable to provide quantitative data on expression levels. Moreover, most of the techniques are useful in detecting the changes in expression of abundant mRNA. However, SAGE allows a rapid and detailed analysis of thousands of transcripts in single attempt. It provides quantitative information on the abundance of known transcripts and has capacity to identify novel expressed genes (Velculescu *et al*. 1995). Improved versions like Long SAGE have also become available recently (Saha *et al*. 2002). The power of SAGE for use in expression profiling has been nicely demonstrated in rice. A total of 10,122 tags from 5921 expressed genes were studied (Matsumara *et al*. 1999). SAGE analysis showed that most of the highly expressed genes belonged to the housekeeping gene category (genes encoding ribosomal protein or proteins responsible for metabolism and cell structure). Also a metallothionein gene accounted for 2.7% of total gene expression. Since SAGE is a technique that is used to analyse the abundance of tran-

scripts, overexpression or knock-out experiments can reveal the function of isolated genes. SAGE can also be used for the isolation of novel gene promoters to drive transgenes. Thus, SAGE along with DNA microarrays can prove to be an ideal tool for gene expression studies in rice.

Post-transcriptional gene silencing

The use of double-stranded RNA allows post-transcriptional silencing of target gene with high efficiency in short time and has been found useful for several organisms (Hammond *et al*. 2001). Virus-induced gene silencing (VIGS) is a variation of the same theme and virus based vectors for gene-silencing in monocots have been developed recently (Holzberg *et al*. 2002). It would be a useful system for rice functional genomics as well if such systems based on rice viruses become available.

Conclusions and perspectives

Rice genomics obtained a major boost in terms of the sequence drafts that were released by distinct projects (Delseny 2003). Notwithstanding the value of finished sequence of rice genome being carried out at present only by IRGSP (Mardis *et al*. 2002), the data released have already provided valuable information on genome structure and organization. It opens up a plethora of opportunities for research related to the life-sciences including evolutionary biology, developmental biology, biochemistry, genetics and molecular biology (Bennetzen 2002; Bowers *et al*. 2003). The latter, in particular, would be aided with the additional advantage of many newly developed genomic tools (Maheshwari *et al*. 2001), molecular markers and genomic libraries. These would be very helpful in gene cloning and functional genomics (Brent 2000), which are essential to gain an insight into the working of the genome.

Along with the already existing techniques for structural and functional genomic analysis, including gene tagging and microarrays, the use of relatively newer tools such as RNA interference (RNAi; Hannon 2002) could help in the acceleration of progress in the field. Terada *et al*. (2002) have also unveiled the use of a direct method of gene targeting through homologous recombination in rice. The procedure could prove to be useful for obtaining gene-targeted or knockout lines of rice or even other plants (Terada *et al*. 2002; Hohn and Puchta 2003). Use of rice genomic data has also extended to the *in silico* identification of regulatory elements. The approach is to identify conserved non-coding sequences based on comparisons between orthologous genes from different organisms (Guo and Moose 2003; Inada *et al*. 2003). The data revealed by rice genomics would be a global heritage to understand genetic events not only in this important crop but in related species as well due to the remarkable degree of synteny existing between them (Schmidt 2000, 2002).

Apart from the understanding of the function of genes and their expression pattern by functional genomics, pioneering studies and developments are also occurring in other areas of discovery-driven science, such as those of proteomics (Kersten *et al*. 2002; Salekdeh *et al*. 2002; Cooper *et al*. 2003; Patterson and Aebersold 2003). Since proteins are one step closer to function than genes, their study will undoubtedly bring forth more data of biological processes including metabolic pathways.

The explosion of information from rice genomics is also expected to have a major positive implication on rice breeding, especially molecular and marker assisted breeding including QTL (quantitative trait loci) analysis (Mohan *et al*. 1997; Yano and Sasaki 1997; Ashikari and Matsuoka 2002). This, however, requires the collaboration of genome researchers and plant breeders (Cyranoski 2003). The understanding of the function of genes and their expression pattern by functional genomics with the aid of transgenics (Pereira 2000) opens opportunities for human intervention at the genetic level. In rice, various useful genes, such as those for quality improvement in terms of micronutrients, vitamins etc, and for resistance against pests, diseases, herbicides and abiotic stresses, have been transferred (Klöti and Potrykus 1999; reviewed by Tyagi and Mohanty 2000). An important example in this context includes the development of 'Golden Rice', i.e. rice engineered to produce provitamin A (*b*-carotene) that could help in alleviating vitamin A deficiency, a major reason for vision impairment (Guerinot 2000; Ye *et al*. 2000). Rice genomics may be needed to find ways for higher accumulation of desired micronutrient in seeds. Such experiments could aid in decreasing malnutrition in the world by using one of the major food staples.

Recently, an International Rice Functional Genomics Consortium (IRFGC) has been formed to provide a platform for sharing materials, integrate databases, seek partnerships, implement cooperative initiatives and accelerate delivery of research results to benefit rice production (http://www.iris.irri.org/IRFGC/). Several national programs have also been developed to make use of rice genome information (Harris 2002; Tyagi and Khurana 2003; Xue *et al*. 2003). There is still a lot to be done in terms of using the knowledge of the rice genome as a boon for further research in both the basic and the applied fields (Cyranoski 2003). But, it can be foreseen that genomic era is steadily advancing towards a phase where rice would have increased productivity and all its characteristics, beneficial to both the farmer and the consumer, would have been manipulated.

Acknowledgements

The research work on rice genes and genomics in our lab is supported by the Department of Biotechnology, Government of India, New Delhi. VG, DK, VR and SV were supported by research fellowship from CSIR/UGC, Govt. of India, New Delhi.

References

Adams M. D., Celniker S. E., Holt R. A., Evans C. A., Gocayne J. D., Amanatides P. G. *et al*. 2000 The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185–2195.

Agrawal G. K., Yamazaki M., Kobayashi M., Hirochika R., Miyao A. and Hirochika H. 2001 Screening of the rice viviparous mutants generated by endogenous retrotransposon *Tos17* insertion. Tagging of a zeaxanthin epoxidase gene and a novel *OsTATC* gene. *Plant Physiol*. **125**, 1248–1257.

Akagi H., Yokozeki Y., Inagaki I. A. and Fujimura T. 1996 Microsatellite DNA markers for rice chromosomes. *Theor. Appl. Genet*. **93**, 1071–1077.

Albar L., Ndjiondjop M. N., Esshak Z., Berger A., Pinel A., Jones M. *et al*. 2003 Fine genetic mapping of a gene required for rice yellow mottle virus cell-to-cell movement. *Theor. Appl. Genet*. **107**, 371–378.

Alonso J. M., Stepanova A. N., Leisse T. J., Kim C. J., Chen H., Shinn P. *et al*. 2003 Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* **301**, 653–657.

Antonio B. A., Emoto M., Wu J., Ashikawa I., Umehara Y., Kurata N. *et al*. 1996 Physical mapping of rice chromosomes 8 and 9 with YAC clones. *DNA Res*. **3**, 393–400.

Antonio B. A., Sakata K. and Sasaki T. 2000 Rice at the forefront of plant genome informatics. *Genome Infor*. **11**, 3–11.

Apisitwanich S. 1999 Genomic *in situ* hybridization in rice. *Kasetsart J. (Nat. Sci.)* **33**, 185–190.

Arber W. 2002 Roots, strategies and prospects for functional genomics. *Curr. Sci*. **83**, 826–828.

Ashikari M. and Matsuoka M. 2002 Application of rice genomics to plant biology and breeding. *Bot. Bull. Acad. Sin*. **43**, 1–11.

Ashikari M., Wu J., Yano M., Sasaki T. and Yoshimura A. 1999 Rice gibberellin-insensitive dwarf mutant gene *Dwarf 1* encodes the alpha-subunit of GTP-binding protein. *Proc. Natl. Acad. Sci. USA* **96**, 10284–10289.

Ashikawa I., Kurata N., Nagamura Y. and Minobe Y. 1994 Cloning and mapping of telomere-associated sequences from rice. *DNA Res*. **1**, 67–76.

Azpiroz-Leehan R. and Feldmann K. A. 1997 T-DNA insertion mutagenesis in *Arabidopsis*: going back and forth. *Trends Genet.* **13**, 152–156.

Baba T., Katagiri S., Tanoue H., Tanaka R., Chiden Y., Saji S. *et al*. 2000 Construction and characterization of rice genomic libraries: PAC library of *japonica* variety, nipponbare and BAC library of *indica* variety, Kasalath. *Bulletin of the NIAR* **14**, 41–49.

Babu P. R., Sekhar A. C., Ithal N., Markandeya G. and Reddy A. R. 2002 Annotation and BAC/PAC localization of nonredundant ESTs from drought-stressed seedlings of an *indica* rice. *J. Genet*. **81**, 25–44.

Barry G. F. 2001 The use of the Monsanto draft rice genome sequence in research. *Plant Physiol*. **125**, 1164–1165.

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

Bennetzen J. L. 2000 Comparative sequence analysis of plant nuclear genomes: microcolinearity and its many exceptions. *Plant Cell* **12**, 1021–1029.

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.

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.

Blanc G., Barakat A., Guyot R., Cooke R. and Delseny M. 2000 Extensive duplication and reshuffling in the *Arabidopsis* genome. *Plant Cell* **12**, 1093–1101.

Bowers J. E., Chapman B. A., Rong J. and Paterson A. H. 2003 Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* **422**, 433–438.

Brent R. 2000 Genomic biology. *Cell* **100**, 169–183.

Bryan G. T., Wu K. S., Farrall L., Jia Y., Hershey H. P., McAdams S. A. *et al*. 2000 A single amino acid difference distinguishes resistant and susceptible alleles of the rice blast resistance gene *Pi-ta*. *Plant Cell* **12**, 2033–2045.

Buell C. R. 2002a Current status of the sequence of the rice genome and prospects for finishing the first monocot genome. *Plant Physiol*. **130**, 1585–1586.

Buell C. R. 2002b Obtaining the sequence of the rice genome and lessons learned along the way. *Trends Plant Sci*. **7**, 538–542.

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

Burge C. and Karlin S. 1997 Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol*. **268**, 78–94.

Burke D. T., Carle G. F. and Olson M. V. 1987 Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* **236**, 806–812.

Causse M. A., Fulton T. M., Cho Y. G., Ahn S. N., Chunwongse J., Wu K. *et al*. 1994 Saturated molecular map of the rice genome based on an interspecific backcross population. *Genetics* **138**, 1251–1274.

Chao Y.-T., Su C.-L., Chow T.-Y., Lee P.-F., Chung C.-I., Huang J.-J. *et al*. 2003 *TEOS1*, a novel transposable element family from *Oryza sativa*. *Bot. Bull. Acad. Sin*. **44**, 1–11.

Chen M., Presting G., Barbazuk W. B., Goicoechea J. L., Blackmon B., Fang G. *et al*. 2002 An integrated physical and genetic map of the rice genome. *Plant Cell* **14**, 537–545.

Chen S., Jin W., Wang M., Zhang F., Zhou J., Jia Q. *et al*. 2003 Distribution and characterization of over 1000 T-DNA tags in rice genome. *Plant J*. **36**, 105–113.

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

Cheng Z., Buell C. R., Wing R. A., Gu M. and Jiang J. 2001a Toward a cytological characterization of the rice genome. *Genome Res*. **11**, 2133–2141.

Cheng Z., Dong F., Langdon T., Ouyang S., Buell C. R., Gu M. *et al*. 2002 Functional rice centromeres are marked by a satellite repeat and a centromere-specific retrotransposon. *Plant Cell* **14**, 1691–1704.

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

Chin H. G., Choe M. S., Lee S.-H., Park S. H., Koo J. C., Kim N. Y. *et al*. 1999 Molecular analysis of rice plants harboring an *Ac/Ds* transposable element-mediated gene trapping system. *Plant J*. **19**, 615–623.

Collins F. S., Green E. D., Guttmacher A. E. and Guyer M. S. 2003a A vision for the future of genomics research. *Nature* **422**, 835–847.

Collins F. S., Morgan M. and Patrinos A. 2003b The human genome project: lessons from large-scale biology. *Science* **300**, 286–290.

Cooper B., Clarke J. D., Budworth P., Kreps J., Hutchison D., Park S. *et al*. 2003 A network of rice genes associated with

stress response and seed development. *Proc. Natl. Acad. Sci. USA* **100**, 4945–4950.
Cyranoski D. 2003 Rice genome: A recipe for revolution? *Nature* **422**, 796–798.
de Jong J. H., Fransz P. and Zabel P. 1999 High resolution FISH in plants – techniques and applications. *Trends Plant Sci.* **4**, 258–263.
Delcher A. L., Kasif S., Fleischmann R. D., Peterson J., White O. and Salzberg S. L. 1999 Alignment of whole genomes. *Nucl. Acids Res.* **27**, 2369–2376.
Delseny M. 2003 Towards an accurate sequence of the rice genome. *Curr. Opin. Plant Biol.* **6**, 101–105.
Delseny M., Salses J., Cooke R., Sallaud C., Regad F., Lagoda P. *et al.* 2001 Rice genomics: present and future. *Plant Physiol. Biochem.* **39**, 323–334.
Devos K. M., Beales J., Nagamura Y. and Sasaki T. 1999 *Arabidopsis*-rice: will colinearity allow gene prediction across the eudicot-monocot divide? *Genome Res.* **9**, 825–829.
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. J., Busso C. S., Yan L., Shiloff B. A. *et al.* 2001 Comparative sequence analysis of colinear barley and rice bacterial artificial chromosomes. *Plant Physiol.* **125**, 1342–1353.
Duggan D. J., Bittner M., Chen Y., Meltzer P. and Trent J. M. 1999 Expression profiling using cDNA microarrays. *Nat. Genet.* **21**, 10–14.
Eckardt N. A. 2000 Sequencing the rice genome. *Plant Cell* **12**, 2011–2017.
Enoki H., Izawa T., Kawahara M., Komatsu M., Koh S., Kyozuka J. *et al.* 1999 *Ac* as a tool for the functional genomics of rice. *Plant J.* **19**, 605–613.
Ewing B. and Green P. 1998 Base-calling of automated sequencer traces using PHRED. II. Error probabilities. *Genome Res.* **8**, 186–194.
Ewing B., Hillier L., Wendl M. C. and Green P. 1998 Base-calling of automated sequencer traces using PHRED. I. Accuracy assessment. *Genome Res.* **8**, 175–185.
Feng Q., Zhang Y., Hao P., Wang S., Fu G., Huang Y. *et al.* 2002 Sequence and analysis of rice chromosome 4. *Nature* **420**, 316–320.
Feuillet C. and Keller B. 2002 Comparative genomics in the grass family: molecular characterization of grass genome structure and evolution. *Ann. Bot.* **89**, 3–10.
Fischer K. S., Barton J., Khush G. S., Leung H. and Cantrell R. 2000 Collaborations in rice. *Science* **290**, 279–280.
Foote T., Roberts M., Kurata N., Sasaki T. and Moore G. 1997 Detailed comparative mapping of cereal chromosome regions corresponding to the *Ph1* locus in wheat. *Genetics* **147**, 801–807.
Fortna A. and Gardiner K. 2001 Genomic sequence analysis tools: a user's guide. *Trends Genet.* **17**, 158–164.
Freeling M. 2001 Grasses as a single genetic system: reassessment 2001. *Plant Physiol.* **125**, 1191–1197.
Fukui K. and Ohmido N. 2000 Rice genome research: an alternative approach based on molecular cytology. In *Genomes* (ed. J. P. Gustafson), pp. 109–121. Kluwer Academic/Plenum Publishers, New York.
Fukui K., Shishido R. and Kinoshita T. 1997 Identification of rice D-genome chromosomes by genomic *in situ* hybridization. *Theor. Appl. Genet.* **95**, 1239–1245.
Fukuoka S., Inoue T., Miyao A., Monna L., Zhong H. S., Sasaki T. *et al.* 1994 Mapping of sequence-tagged sites in rice by single strand conformation polymorphism. *DNA Res.* **1**, 271–277.
Gale M. D. and Devos K. M. 1998a Comparative genetics in the grasses. *Proc. Natl. Acad. Sci. USA* **95**, 1971–1974.
Gale M. D. and Devos K. M. 1998b Plant comparative genetics after 10 years. *Science* **282**, 656–659.
Gerstein M. and Jansen R. 2000 The current excitement in bioinformatics-analysis of whole-genome expression data: how does it relate to protein structure and function. *Curr. Opin. Struct. Biol.* **10**, 574–584.
Goff S. A. 1999 Rice as a model for cereal genomics. *Curr. Opin. Plant Biol.* **2**, 86–89.
Goff S. A., Ricke D., Lan T. H., Presting G., Wang R., Dunn M. *et al.* 2002 A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* **296**, 92–100.
Gordon D., Abajian C. and Green P. 1998 CONSED: a graphical tool for sequence finishing. *Genome Res.* **8**, 195–202.
Grant D., Cregan P. and Shoemaker R. C. 2000 Genome organization in dicots: genome duplication in *Arabidopsis* and synteny between soybean and *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **97**, 4168–4173.
Greco R., Ouwerkerk P. B., Taal A. J., Favalli C., Beguiristain T., Puigdomenech P. *et al.* 2001 Early and multiple *Ac* transpositions in rice suitable for efficient insertional mutagenesis. *Plant Mol. Biol.* **46**, 215–227.
Guerinot M. L. 2000 The green revolution strikes gold. *Science* **287**, 241–242.
Guo H. and Moose S. P. 2003 Conserved noncoding sequences among cultivated cereal genomes identify candidate regulatory sequence elements and patterns of promoter evolution. *Plant Cell* **15**, 1143–1158.
Hamer L., Dezwaan T. M., Montenegro-Chamorro M. V., Frank S. A. and Hamer J. E. 2001 Recent advances in large-scale transposon mutagenesis. *Curr. Opin. Chem. Biol.* **5**, 67–73.
Hammond S. M., Caudy A. A. and Hannon G. J. 2001 Post-transcriptional gene silencing by double-stranded RNA. *Nat. Rev. Genet.* **2**, 110–119.
Han B. and Xue Y. 2003 Genome-wide intraspecific DNA-sequence variations in rice. *Curr. Opin. Plant. Biol.* **6**, 134–138.
Hannon G. J. 2002 RNA interference. *Nature* **418**, 244–251.
Harris S. B. 2002 Virtual rice: Japan sets up the rice simulator project to create an *in silico* rice plant. *EMBO Rep.* **3**, 511–513.
Harushima Y., Yano M., Shomura A., Sato M., Shimano T., Kuboki Y. *et al.* 1998 A high-density rice genetic linkage map with 2275 markers using a single F2 population. *Genetics* **148**, 479–494.
Heng H. H., Spyropoulos B. and Moens P. B. 1997 FISH technology in chromosome and genome research. *BioEssays* **19**, 75–84.
Hieter P. and Boguski M. 1997 Functional genomics: it's all how you read it. *Science* **278**, 601–602.
Hirochika H. 1997 Retrotransposons of rice: their regulation and use for genome analysis. *Plant Mol. Biol.* **35**, 231–240.
Hirochika H. 1999 Retrotransposon of rice as a tool for forward and reverse genetics. In *Molecular Biology of Rice* (ed. K. Shimamoto), pp. 43–58. Springer-Verlag, Tokyo.
Hirochika H. 2001a Contribution of the *Tos17* retrotransposon to rice functional genomics. *Curr. Opin. Plant Biol.* **4**, 118–122.
Hirochika H., Miyao A., Yamazaki M., Takeda S., Abe K., Hirochika R. *et al.* 2001b Retrotransposons of rice as a tool for the functional analysis of genes. In *Rice Genetics IV. Proceedings of the Fourth International Rice Genetics Symposium, 22–27 October 2000* (ed. G. S. Khush, D. S. Brar and B. Hardy), pp. 279–292. Science Publishers, Inc, New Delhi (India) and International Rice Research Institute, Los Banos (Philippines).
Hirochika H., Otsuki H., Yoshikawa M., Otsuki Y., Sugimoto K. and Takeda S. 1996 Autonomous transposition of the tobacco retrotransposon *Tto1* in rice. *Plant Cell* **8**, 725–734.

Hohn B. and Puchta H. 2003 Some like it sticky: targeting of the rice gene *Waxy*. *Trends Plant Sci*. **8**, 51–53.

Holzberg S., Brosio P., Gross C. and Pogue G. P. 2002 Barley stripe mosaic virus-induced gene silencing in a monocot plant. *Plant J*. **30**, 315–327.

Inada D. C., Bashir A., Lee C., Thomas B. C., Ko C., Goff S. A. *et al*. 2003 Conserved noncoding sequences in the grasses. *Genome Res*. **13**, 2030–2041.

Ioannou P. A., Amemiya C. T., Garnes J., Kroisel P. M., Shizuya H., Chen C. *et al*. 1994 A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. *Nat. Genet.* **6**, 84–89.

Inoue T., Zhong H. S., Miyao A., Ashikawa I., Monna L., Fukuoka S. *et al*. 1994 Sequence-tagged sites (STSs) as standard landmarkers in the rice genome. *Theor. Appl. Genet*. **89**, 728–734.

Iwata N. 1986 The relationship between cytologically identified chromosomes and linkage groups in rice. In *Rice Genetics. Proceedings of the International Rice Genetics Symposium, 27–31 May 1985*, pp. 229–238. International Rice Research Institute Los Banos (Philippines).

Izawa T., Miyazaki C., Yamamoto M., Terada R., Iida S. and Shimamoto K. 1991 Introduction and transposition of the maize transposable element *Ac* in rice (*Oryza sativa* L.). *Mol. Gen. Genet*. **227**, 391–396.

Izawa T. and Shimamoto K. 1996 Becoming a model plant: the importance of rice to plant science. *Trends Plant Sci*. **1**, 95–99.

Jeon J. S., Chen D., Yi G. H., Wang G. L. and Ronald P. C. 2003 Genetic and physical mapping of *Pi5(t)*, a locus associated with broad-spectrum resistance to rice blast. *Mol. Genet. Genomics* **269**, 280–289.

Jeon J.-S., Lee S., Jung K.-H., Jun S.-H., Jeong D.-H., Lee J. *et al*. 2000 T-DNA insertional mutagenesis for functional genomics in rice. *Plant J*. **22**, 561–570.

Jeong D.-H., An S., Kang H.-G., Moon S., Han J.-J., Park S. *et al*. 2002 T-DNA insertional mutagenesis for activation tagging in rice. *Plant Physiol*. **130**, 1636–1644.

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

Jiang N., Bao Z., Temnykh S., Cheng Z., Jiang J., Wing R. A. *et al*. 2002 *Dasheng*: a recently amplified nonautonomous long terminal repeat element that is a major component of pericentromeric regions in rice. *Genetics* **161**, 1293–1305.

Jiang N., Bao Z., Zhang X., Hirochika H., Eddy S. R., McCouch S. R. *et al*. 2003 An active DNA transposon family in rice. *Nature* **421**, 163–167.

Karlowski W. M., Schoof H., Janakiraman V., Stuempflen V. and Mayer K. F. X. 2003 MOsDB: an integrated information resource for rice genomics. *Nucl. Acids Res*. **31**, 190–192.

Kawasaki S., Borchert C., Deyholos M., Wang H., Brazille S., Kawai K. *et al*. 2001 Gene expression profiles during the initial phase of salt stress in rice. *Plant Cell* **13**, 889–905.

Keller B. and Feuillet C. 2000 Colinearity and gene density in grass genomes. *Trends Plant Sci*. **5**, 246–251.

Kersten B., Burkle L., Kuhn E. J., Giavalisco P., Konthur Z., Lueking A. *et al*. 2002 Large-scale plant proteomics. *Plant Mol. Biol*. **48**, 133–141.

Khush G. S. 1997 Origin, dispersal, cultivation and variation of rice. *Plant Mol. Biol*. **35**, 25–34.

Khush G. S. and Brar D. S. 2001 Rice genetics from Mendel to functional genomics. In *Rice Genetics IV. Proceedings of the Fourth International Rice Genetics Symposium, 22–27 October 2000* (ed. G. S. Khush, D. S. Brar and B. Hardy), pp. 3–25. Science Publishers, Inc., New Delhi (India) and International Rice Research Institute, Los Banos (Philippines).

Khush G. S. and Singh R. J. 1985 Relationships between linkage groups and cytologically identifiable chromosomes of rice. In *Rice Genetics. Proceedings of the International Rice Genetics Symposium, 27–31 May 1985*, pp. 239–248. Island Publishing House, Inc, Manila, Philippines.

Khush G. S., Singh K., Ishii T., Parco A., Huang N., Brar D. S. *et al*. 1996 Centromere mapping and orientation of the cytological, classical and molecular linkage maps of rice. In *Rice Genetics III. Proceedings of the Third International Rice Genetics Symposium, 16–20 October 1995* (ed. G. S. Khush), pp. 57–75. International Rice Research Institute, Los Banos (Philippines).

Kikuchi K., Terauchi K., Wada M. and Hirano H.-Y. 2003 The plant MITE *mPing* is mobilized in anther culture. *Nature* **421**, 167–170.

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.

Kleffe J., Hermann K., Vahrson W., Wittig B. and Brendel V. 1996 Logitlinear models for the prediction of splice sites in plant pre-mRNA sequences. *Nucl. Acids Res*. **24**, 4709–4718.

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.

Klöti A. and Potrykus I. 1999 Rice improvement by genetic transformation. In *Molecular Biology of Rice* (ed. K. Shimamoto), pp. 283–301. Springer, Tokyo.

Koike K., Yoshino K., Sue N., Umehara Y., Ashikawa I., Kurata N. *et al*. 1997 Physical mapping of rice chromosomes 4 and 7 using YAC clones. *DNA Res*. **4**, 27–33.

Kojima S., Takahashi Y., Kobayashi Y., Monna L., Sasaki T., Araki T. *et al*. 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.

Kowalski S. P., Lan T.-H., Feldmann K. A. and Paterson A. H. 1994 Comparative mapping of *Arabidopsis thaliana* and *Brassica oleracea* chromosomes reveals islands of conserved organization. *Genetics* **138**, 499–510.

Krysan P. J., Young J. C. and Sussman M. R. 1999 T-DNA as an insertional mutagen in *Arabidopsis*. *Plant Cell* **11**, 2283–2290.

Ku H. M., Vision T., Liu J. and Tanksley S. D. 2000 Comparing sequenced segments of the tomato and *Arabidopsis* genomes: large-scale duplication followed by selective gene loss creates a network of synteny. *Proc. Natl. Acad. Sci. USA* **97**, 9121–9126.

Kurata N., Nagamura Y., Yamamoto K., Harushima Y., Sue N., Wu J. *et al*. 1994 A 300 kilobase interval genetic map of rice including 883 expressed sequences. *Nat. Genet*. **8**, 365–372.

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

Lagercrantz U. 1998 Comparative mapping between *Arabidopsis thaliana* and *Brassica nigra* indicates that *Brassica* genomes have evolved through extensive genome replication accompanied by chromosome fusions and frequent rearrangements. *Genetics* **150**, 1217–1228.

Lee S., Kim J. -Y. Kim S. -H., Kim S. -J., Lee K., Han S.-K. *et al*. 2004 Trapping and characterization of cold-responsive genes from T-DNA tagging lines in rice. *Plant Sci*. **166**, 69–79.

Leung H., Wu C., Baraoidan M., Bordeos A., Ramos M., Madamba, S. *et al*. 2001 Deletion mutants for functional genomics:

progress in phenotyping, sequence assignment and database development. In *Rice Genetics IV. Proceedings of the Fourth International Rice genetics Symposium, 22–27 October 2000* (ed. G. S. Khush, D. S. Brar and B. Hardy), pp. 239–251. Science Publishers, Inc. New Delhi (India), International Rice Research Institute Los Baños (Philippines).

Lewis S., Ashburner M. and Reese M. G. 2000 Annotating eukaryote genomes. *Curr. Opin. Struct. Biol.* **10**, 349–354.

Li C. B., Zhang D. M., Ge S. and Hong D. Y. 2000 Identification of genomic constitution of three tetraploid *Oryza* species through two-probe genomic *in situ* hybridization. *International Rice Research Notes* **25**, 19–22.

Liu D., Zhang S., Fauquet C. and Crawford N. M. 1999 The *Arabidopsis* transposon *Tag1* is active in rice, undergoing germinal transposition and restricted, late somatic excision. *Mol. Gen. Genet.* **262**, 413–420.

Liu H., Sachidanandam R. and Stein L. 2001 Comparative genomics between rice and *Arabidopsis* shows scant collinearity in gene order. *Genome Res.* **11**, 2020–2026.

Livingstone K. and Rieseberg L. H. 2002 Rice genomes: a grainy view of future evolutionary research. *Curr. Biol.* **12**, R470–R471.

Lowe T. M. and Eddy S. R. 1997 tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucl. Acids Res.* **25**, 955–964.

Lukowitz W., Gillmor C. S. and Scheible W. R. 2000 Positional cloning in *Arabidopsis*. Why it feels good to have a genome initiative working for you. *Plant Physiol.* **123**, 795–805.

Maclean J. L., Dawe D. C., Hardy B. and Hettel G. P. 2002 Rice almanac. Los Baños (Phillippines): International Rice Research Institute, Bouaké (Côte d'Ivoire): West Africa Rice Development Association, Cali (Colombia): International Center for Tropical Agriculture, Rome (Italy): Food and Agriculture Organization, pp. 253.

Maheshwari S. C., Maheshwari N. and Sopory S. K. 2001 Genomics, DNA chips and a revolution in plant biology. *Curr. Sci.* **80**, 252–261.

Mao L., Wood T. C., Yu Y., Budiman M. A., Tomkins J., Woo S. *et al.* 2000 Rice transposable elements: a survey of 73,000 sequence-tagged-connectors. *Genome Res.* **10**, 982–990.

Mardis E., McPherson J., Martienssen R., Wilson R. K. and McCombie W. R. 2002 What is finished and why does it matter. *Genome Res.* **12**, 669–671.

Marra M. A., Hillier L. and Waterston R. H. 1998 Expressed sequence tags-ESTablishing bridges between genomes. *Trends Genet.* **14**, 4–7.

Martin G. B., Brommonschenkel S. H., Chunwongse J., Frary A., Ganal M. W., Spivey R. *et al.* 1993 Map-based cloning of a protein kinase gene conferring disease resistance in tomato. *Science* **262**, 1432–1436.

Matsumura H., Nirasawa S. and Terauchi R. 1999 Transcript profiling in rice (*Oryza sativa* L.) seedlings using serial analysis of gene expression (SAGE). *Plant J.* **20**, 719–726.

Mayer K., Murphy G., Tarchini R., Wambutt R., Volckaert G., Pohl T. *et al.* 2001 Conservation of microstructure between a sequenced region of the genome of rice and multiple segments of the genome of *Arabidopsis thaliana*. *Genome Res.* **11**, 1167–1174.

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

McCouch S. R., Kochert G., Yu Z. H., Wang Z. Y., Khush G. S., Coffman W. R. *et al.* 1988 Molecular mapping of rice chromosomes. *Theor. Appl. Genet.* **76,** 815–829.

McCouch S. R., Teytelman L., Xu Y., Lobos K. B., Clare K., Walton M. *et al.* 2002 Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.). *DNA Res.* **9**, 199–207.

Meinke D. W., Cherry J. M., Dean C., Rounsley S. D. and Koornneef M. 1998 *Arabidopsis thaliana:* a model plant for genome analysis. *Science* **282**, 662–682.

Mewes H. W., Albermann K., Bahr M., Frishman D., Gleissner A., Hani J. *et al.* 1997 Overview of the yeast genome. *Nature* **387**, 7–8.

Miyao A., Tanaka K., Murata K., Sawaki H., Takeda S., Abe K. *et al.* 2003 Target site specificity of the *Tos17* retrotransposon shows a preference for insertion within genes and against insertion in retrotransposon-rich regions of the genome. *Plant Cell* **15**, 1771–1780.

Mohan M., Nair S., Bhagwat A., Krishna T. G., Yano M. and Bhatia C. R. *et al.* 1997 Genome mapping, molecular markers and marker-assisted selection in crop plants. *Mol. Breed.* **3**, 87–103.

Monna L., Kitazawa N., Yoshino R., Suzuki J., Masuda H., Maehara Y. *et al.* 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.

Monna L., Miyao A., Inoue T., Fukuoka S., Yamazaki M., Zhong H. S. *et al.* 1994 Determination of RAPD markers in rice and their conversion into sequence tagged sites (STSs) and STS-specific primers. *DNA Res.* **1**, 139–148.

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

Murai N., Li Z. J., Kawagoe Y. and Hayashimoto A. 1991 Transposition of the maize activator element in transgenic rice plants. *Nucl. Acids Res.* **19**, 617–622.

Nagamura Y., Antonio B. A. and Sasaki T. 1997 Rice molecular genetic map using RFLPs and its applications. *Plant Mol. Biol.* **35**, 79–87.

Nakagahra M., Okuno K. and Vaughan D. 1997 Rice genetic resources: history, conservation, investigative characterization and use in Japan. *Plant Mol. Biol.* **35**, 69–77.

Nakazaki T., Okumoto Y., Horibata A., Yamahira S., Teraishi M., Nishida H. *et al.* 2003 Mobilization of a transposon in the rice genome. *Nature* **421**, 170–172.

Nasu S., Suzuki J., Ohta R., Hasegawa K., Yui R., Kitazawa N. *et al.* 2002 Search for and analysis of single nucleotide polymorphisms (SNPs) in rice (*Oryza sativa*, *Oryza rufipogon*) and establishment of SNP markers. *DNA Res.* **9**, 163–171.

Ohmido N., Akiyama Y. and Fukui K. 1998 Physical mapping of unique nucleotide sequences on identified rice chromosomes. *Plant Mol. Biol.* **38**, 1043–1052.

Ohmido N., Kijima K., Ashikawa I., De Jong J. H. and Fukui K. 2001 Visualization of the terminal structure of rice chromosomes 6 and 12 with multicolor FISH to chromosomes and extended DNA fibers. *Plant Mol. Biol.* **47**, 413–421.

Paterson A. H., Lan T. H., Reischmann K. P., Chang C., Lin Y. R., Liu S. C. *et al.* 1996 Toward a unified genetic map of higher plants, transcending the monocot-dicot divergence. *Nat. Genet.* (suppl.) **14**, 380–382.

Patterson S. D. and Aebersold R. H. 2003 Proteomics: the first decade and beyond. *Nat. Genet.* **33**, 311–323.

Pereira A. 2000 A transgenic perspective on plant functional genomics. *Transgenic Res.* **9**, 245–260.

Pertea M. and Salzberg S. L. 2002 Computational gene finding in plants. *Plant Mol. Biol.* **48**, 39–48.

Qi L., Echalier B., Friebe B. and Gill B. S. 2003 Molecular characterization of a set of wheat deletion stocks for use in chromosome bin mapping of ESTs. *Funct. Integr. Genomics* **3**, 39–55.

Rhee S. Y. 2000 Bioinformatic resources, challenges and oppor-

tunities using *Arabidopsis* as a model organism in a post-genomic era. *Plant Physiol*. **124**, 1460–1464.

Rhee S. Y., Beavis W., Berardini T. Z., Chen G., Dixon D., Doyle A. *et al*. 2003 The *Arabidopsis* Information Resource (TAIR): a model organism database providing a centralized, curated gateway to *Arabidopsis* biology, research materials and community. *Nucl. Acids Res*. **31**, 224–228.

Rossberg M., Theres K., Acarkan A., Herrero R., Schmitt T., Schumacher K. *et al*. 2001 Comparative sequence analysis reveals extensive microcolinearity in the lateral suppressor regions of the tomato, *Arabidopsis* and *Capsella* genomes. *Plant Cell* **13**, 979–988.

Rouze P., Pavy N. and Rombauts S. 1999 Genome annotation: which tools do we have for it? *Curr. Opin. Plant Biol*. **2**, 90–95.

Rudd S. 2003 Expressed sequence tags: alternative or complement to whole genome sequences? *Trends Plant Sci*. **8**, 321–329.

Saha S., Sparks A. B., Rago C., Akmaev V., Wang C. J., Vogelstein B. *et al*. 2002 Using the transcriptome to annotate the genome. *Nat. Biotechnol*. **20**, 508–512.

Sahi C., Agarwal M., Reddy M. K., Sopory S. K. and Grover A. 2003 Isolation and expression analysis of salt stress-associated ESTs from contrasting rice cultivars using a PCR-based subtraction method. *Theor. Appl. Genet*. **106**, 620–628.

Saji S., Umehara Y., Antonio B. A., Yamane H., Tanoue H., Baba T. *et al*. 2001 A physical map with yeast artificial chromosome (YAC) clones covering 63% of the 12 rice chromosomes. *Genome* **44**, 32–37.

Saji S., Umehara Y., Kurata N., Ashikawa I. and Sasaki T. 1996 Construction of YAC contigs on rice chromosome 5. *DNA Res*. **3**, 297–302.

Sakata K., Antonio B. A., Mukai Y., Nagasaki H., Sakai Y., Makino K. *et al*. 2000 INE: a rice genome database with an integrated map view. *Nucl. Acids Res*. **28**, 97–101.

Sakata K., Nagamura Y., Numa H., Antonio B. A., Nagasaki H., Idonuma A. *et al*. 2002 RiceGAAS: an automated annotation system and database for rice genome sequence. *Nucl. Acids Res*. **30**, 98–102.

Sakata K., Nagasaki H., Idonuma A., Waki K., Kise M. and Sasaki T. 1999 A computer program for prediction of gene domain on rice genome sequence. In *The 2nd Georgia Tech International Conference on Bioinformatics, Abstracts*, pp. 78.

Sallaud C., Meynard D., Van Boxtel J., Gay C., Bes M., Brizard J. P. *et al*. 2003 Highly efficient production and characterization of T-DNA plants for rice (*Oryza sativa* L.) functional genomics. *Theor. Appl. Genet*. **106**, 1396–1408.

Salamov A. A. and Solovyev V. V. 2000 *Ab initio* gene finding in Drosophila genomic DNA. *Genome Res*. **10**, 516–522.

Salekdeh G. H., Siopongco J., Wode L. J., Gharlyazie B. and Bennet J. 2002 A proteomic approach to analysing drought- and salt-responsiveness in rice. *Field Crops Res*. **76**, 199–219.

Salse J., Piegu B., Cooke R. and Delseny M. 2002 Synteny between *Arabidopsis thaliana* and rice at the genome level: a tool to identify conservation in the ongoing rice genome sequencing project. *Nucl. Acids Res*. **30**, 2316–2328.

Sasaki T. 1998 The rice genome project in japan. *Proc. Natl. Acad. Sci. USA* **95**, 2027–2028.

Sasaki T. 2002 Rice genomics to understand rice plant as an assembly of genetic codes. *Curr. Sci*. **83**, 834–839.

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

Sasaki T., Matsumoto T., Yamamoto K., Sakata K., Baba T., Katayose Y. *et al*. 2002 The genome sequence and structure of rice chromosome 1. *Nature* **420**, 312–316.

Sato Y., Sentoku N., Miura Y., Hirochika H., Kitano H. and Matsuoka M. 1999 Loss-of-function mutations in the rice homeobox gene *OSH15* affect the architecture of internodes resulting in dwarf plants. *EMBO J*. **18**, 992–1002.

Schena M., Shalon D., Davis R. W. and Brown P. O. 1995 Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467–470.

Schmidt R. 2000 Synteny: recent advances and future prospects. *Curr. Opin. Plant Biol*. **3**, 97–102.

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

Schoof H. and Karlowski W. M. 2003 Comparison of rice and *Arabidopsis* annotation. *Curr. Opin. Plant Biol*. **6**, 106–112.

Shimamoto K., Miyazaki C., Hashimoto H., Izawa T., Itoh K., Terada R. *et al*. 1993 Trans-activation and stable integration of the maize transposable element *Ds* cotransfected with the *Ac* transposase gene in transgenic rice plants. *Mol. Gen. Genet*. **239**, 354–360.

Shimokawa T., Kurata N., Wu J., Umehara Y., Ashikawa I. and Sasaki T. 1996 Assignment of YAC clones spanning rice chromosomes 10 and 12. *DNA Res*. **3**, 401–406.

Shimono M., Yazaki J., Nakamura K., Kishimoto N., Kikuchi S., Iwano M. *et al*. 2003 Disease control: cDNA microarray analysis of gene expression in rice plants treated with probenazole, a chemical inducer of disease resistance. *J. Gen. Plant Pathol*. **69**, 76–82.

Shishido R., Apisitwanich S., Ohmido N., Okinaka Y., Mori K. and Fukui K. 1998 Detection of specific chromosome reduction in rice somatic hybrids with the A, B and C genomes by multi-colour genomic *in situ* hybridization. *Theor. Appl. Genet*. **97**, 1013–1018.

Shizuya H., Birren B., Kim U. J., Mancino V., Slepak T., Tachiiri Y. *et al*. 1992 Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an *F-factor*-based vector. *Proc. Natl. Acad. Sci. USA* **89**, 8794–8797.

Singh K., Ishii T., Parco A., Huang N., Brar D. S. and Khush G. S. 1996 Centromere mapping and orientation of the molecular linkage map of rice (*Oryza sativa* L.). *Proc. Natl. Acad. Sci. USA* **93**, 6163–6168.

Singh N. K., Raghuvanshi S., Srivastava S. K., Gaur A., Pal A. K., Dalal V. *et al*. 2004 Sequence analysis of the long arm of rice chromosome 11 for rice-wheat synteny. *Funct. Integr. Genomics* (in press).

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

Song W. Y., Wang G. L., Chen L. L., Kim H. S., Pi L. Y., Holsten T. *et al*. 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., Kantety R., Munkvold J. D. *et al*. 2003 Comparative DNA sequence analysis of wheat and rice genomes. *Genome Res*. **13**, 1818–1827.

Springer P. S. 2000 Gene traps: tools for plant development and genomics. *Plant Cell* **12**, 1007–1020.

Sugimoto K., Otsuki Y., Saji S. and Hirochika H. 1994 Transposition of the maize *Ds* element from a viral vector to the rice genome. *Plant J*. **5**, 863–871.

Sundaresan V. 1996 Horizontal spread of transposon mutagenesis: new uses for old elements. *Trends Plant Sci*. **1**, 184–190.

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.

Takano M., Kanegae H., Shinomura T., Miyao A., Hirochika H. and Furuya M. 2001 Isolation and characterization of rice phytochrome A mutants. *Plant Cell* **13**, 521–534.

Tanoue H., Shimokawa T., Wu J., Sue N., Umehara Y., Ashikawa I. *et al*. 1997 Ordered YAC clone contigs assigned to rice chromosomes 3 and 11. *DNA Res*. **4**, 133–140.

Tao Q., Chang Y.-L., Wang J., Chen H., Islam-Faridi M. N., Scheuring C. *et al*. 2001 Bacterial artificial chromosome-based physical map of the rice genome constructed by restriction fingerprint analysis. *Genetics* **158**, 1711–1724.

Tao Q., Wang A. and Zhang H. B. 2002 One large-insert plant-transformation-competent BIBAC library and three BAC libraries of *japonica* rice for genome research in rice and other grasses. *Theor. Appl. Genet*. **105**, 1058–1066.

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

Temnykh S., Park W. D., Ayres N., Cartinhour S., Hauck N. and Lipovich L. *et al*. 2000 Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa* L.). *Theor. Appl. Genet*. **100**, 697–712.

Terada R., Urawa H., Inagaki Y., Tsugane K. and Iida S. 2002 Efficient gene targeting by homologous recombination in rice. *Nat. Biotechnol*. **20**, 1030–1034.

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

The *C. Elegans* Sequencing Consortium 1998 Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* **282**, 2012–2018.

The International Human Genome Mapping Consortium 2001 Initial sequencing and analysis of human genome. *Nature* **409**, 860–921.

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.

Turcotte K., Srinivasan S. and Bureau T. 2001 Survey of transposable elements from rice genomic sequences. *Plant J*. **25**, 169–179.

Tyagi A. K. and Khurana J. P. 2003 Plant molecular biology and biotechnology research in the post-recombinant DNA era. *Adv. Biochem. Eng. Biotechnol*. **84**, 91–121.

Tyagi A. K., Khurana J. P., Khurana P., Mohanty A. and Bharti A. K. 2003 Genome-wide molecular approaches in plants: From structure to function. In *Proceedings of Diamond Jubilee Symposium of Indian Society for Genetics and Plant Breeding. ISGPB, New Delhi*, pp. 235–250.

Tyagi A. K. and Mohanty A. 2000 Rice transformation for crop improvement and functional genomics. *Plant Sci*. **158**, 1–18.

Umehara Y., Inagaki A., Tanoue H., Yasukochi Y., Nagamura Y., Saji S. *et al*. 1995 Construction and characterization of a rice YAC library for physical mapping. *Mol. Breed*. **1**, 79–89.

Umehara Y., Kurata N., Ashikawa I. and Sasaki T. 1997 Yeast artificial chromosome clones of rice chromosome 2 ordered using DNA markers. *DNA Res*. **4**, 127–131.

Umehara Y., Tanoue H., Kurata N., Ashikawa I., Minobe Y. and Sasaki T. 1996 An ordered yeast artificial chromosome library covering over half of rice chromosome 6. *Genome Res*. **6**, 935–942.

Vaughan D. A., Morishima H. and Kadowaki K. 2003 Diversity in the *Oryza* genus. *Curr. Opin. Plant Biol*. **6**, 139–146.

Velculescu V. E., Zhang L., Vogelstein B. and Kinzler K. W. 1995 Serial analysis of gene expression. *Science* **270**, 484–487.

Venter J. C., Adams M. D., Myers E. W., Li P. W., Mural R. J., Sutton G. G. *et al*. 2001 The sequence of the human genome. *Science* **291**, 1304–1351.

Walbot V. 1992 Strategies for mutagenesis and gene cloning using transposon tagging and T-DNA insertional mutagenesis. *Annu. Rev. Plant Physiol. Plant Mol. Biol*. **43,** 49–82.

Walden R., Fritze K., Hayashi H., Miklashevichs E., Harling H. and Schell J. 1994 Activation tagging: a means of isolating genes implicated as playing a role in plant growth and development. *Plant Mol. Biol*. **26**, 1521–1528.

Wang Z. -X., Idonuma A., Umehara Y., Van Houten W., Ashikawa I., Minobe Y. *et al*. 1996 Physical mapping of rice chromosome 1 with yeast artificial chromosomes (YACs). *DNA Res*. **3**, 291–296.

Wang Z. X., Yano M., Yamanouchi U., Iwamoto M., Monna L., Hayasaka H. *et al*. 1999 The *Pib* gene for rice blast resistance belongs to the nucleotide binding and leucine-rich repeat class of plant disease resistance genes. *Plant J*. **19**, 55–64.

Ware D. and Stein L. 2003 Comparison of genes among cereals. *Curr. Opin. Plant Biol*. **6**, 121–127.

Ware D. H., Jaiswal P., Ni J., Yap I. V., Pan X., Clark K. Y. *et al*. 2002a Gramene, a tool for grass genomics. *Plant Physiol*. **130**, 1606–1613.

Ware D., Jaiswal P., Ni J., Pan X., Chang K., Clark K. *et al*. 2002b Gramene: a resource for comparative grass genomics. *Nucl. Acids Res*. **30**, 103–105.

Weigel D., Ahn J. H., Blazquez M. A., Borevitz J. O., Christensen S. K., Fankhauser C. *et al*. 2000 Activation tagging in *Arabidopsis*. *Plant Physiol*. **122**, 1003–1013.

Wisman E. and Ohlrogge J. 2000 *Arabidopsis* microarray service facilities. *Plant Physiol*. **124**, 1468–1471.

Wong G. K. -S., Wang J., Tao L., Tan J., Zhang J., Passey D. A. *et al*. 2002 Compositional gradients in Gramineae genes. *Genome Res*. **12**, 851–856.

Wu C., Li X., Yuan W., Chen G., Kilian A., Li J. *et al*. 2003 Development of enhancer trap lines for functional analysis of the rice genome. *Plant J*. **35**, 418–427.

Wu J., Maehara T., Shimokawa T., Yamamoto S., Harada C., Takazaki Y. *et al*. 2002 A comprehensive rice transcript map containing 6591 expressed sequence tag sites. *Plant Cell* **14**, 525–535.

Wu J., Matsui E., Yamamoto K., Nagamura Y., Kurata N., Takuji S. *et al*. 1995 Genomic organization of 57 ribosomal protein genes in rice (*Oryza sativa* L.) through RFLP mapping. *Genome* **38**, 1189–1200.

Xue Y., Li J. and Xu Z. 2003 Recent highlights of the China rice functional genomics program. *Trends Genet*. **19**, 390–394.

Yamamoto K. and Sasaki T. 1997 Large-scale EST sequencing in rice. *Plant Mol. Biol*. **35**, 135–144.

Yamanouchi U., Yano M., Lin H., Ashikari M. and Yamada K. 2002 A rice spotted leaf gene, *Spl7*, encodes a heat stress transcription factor protein. *Proc. Natl. Acad. Sci. USA* **99**, 7530–7535.

Yang Z., Sun X., Wang S. and Zhang Q. 2003 Genetic and physical mapping of a new gene for bacterial blight resistance in rice. *Theor. Appl. Genet*. **106**, 1467–1472.

Yang T. -J., Yu Y., Nah G., Atkins M., Lee S., Frisch D. A. *et al*. 2003 Construction and utility of 10-kb libraries for efficient clone-gap closure for rice genome sequencing. *Theor. Appl. Genet*. **107**, 652–660.

Yano M. and Sasaki T. 1997 Genetic and molecular dissection of quantitative traits in rice. *Plant Mol. Biol*. **35**, 145–153.

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

Yazaki J., Kishimoto N., Ishikawa M. and Kikuchi S. 2002 Rice expression database: the gateway to rice functional genomics. *Trends Plant Sci*. **7**, 563–564.

Yazaki J., Kishimoto N., Nakamura K., Fujii F., Shimbo K., Otsuka Y. *et al*. 2000 Embarking on rice functional genomics via cDNA microarray: use of 3′ UTR probes for specific gene expression analysis. *DNA Res*. **7**, 367–370.

Ye X., Al-Babili S., Kloti A., Zhang J., Lucca P., Beyer P. *et al*. 2000 Engineering the provitamin A (beta-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. *Science* **287**, 303–305.

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

Yoshimura S., Umehara Y., Kurata N., Nagamura Y., Sasaki T., Minobe Y. *et al*. 1996 Identification of a YAC clone carrying the *Xa-1* allele, a bacterial blight resistance gene in rice. *Theor*. *Appl*. *Genet*. **93**, 117–122.

Yoshimura S., Yamanouchi U., Katayose Y., Toki S., Wang Z. -X., Kono I. *et al*. 1998 Expression of *Xa1*, a bacterial blight-resistance gene in rice, is induced by bacterial inoculation. *Proc*. *Natl*. *Acad*. *Sci*. *USA* **95**, 1663–1668.

Yu J., Hu S., Wang J., Wong G. K., Li S., Liu B. *et al*. 2002 A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* **296**, 79–92.

Yuan Q., Ouyang S., Liu J., Suh B., Cheung F., Sultana R. *et al*. 2003 The TIGR rice genome annotation resource: annotating the rice genome and creating resources for plant biologists. *Nucl*. *Acids Res*. **31**, 229–233.

Yuan Q., Quackenbush J., Sultana R., Pertea M., Salzberg S. L. and Buell C. R. 2001 Rice bioinformatics. Analysis of rice sequence data and leveraging the data to other plant species. *Plant Physiol*. **125**, 1166–1174.

Zang M. 1997 Identification of protein coding region in the human genome based on quadratic discriminant analysis. *Proc*. *Natl*. *Acad*. *Sci*. *USA* **94**, 565–568.

Zhang H.-B., Choi S., Woo S.-S., Li Z. and Wing R. A. 1996 Construction and characterization of two rice bacterial artificial chromosome libraries from the parents of a permanent recombinant inbred mapping population. *Mol*. *Breed*. **2**, 11–24.

Zhao Q., Zhang Y., Cheng Z., Chen M., Wang S., Feng Q. *et al*. 2002 A fine physical map of the rice chromosome 4. *Genome Res*. **12**, 817–823.

Zhu T., Budworth P., Chen W., Provart N., Chang H.-S., Guimil S. *et al*. 2003 Transcriptional control of nutrient partitioning during rice grain filling. *Plant Biotech*. *J*. **1**, 59–70.

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