Mol Breeding (2010) 26:257–273
DOI 10.1007/s11032-009-9367-7

Rice genomics moves ahead

Saurabh Raghuvanshi · Meenu Kapoor · Shashi Tyagi · Sanjay Kapoor · Paramjit Khurana · Jitendra Khurana · Akhilesh Tyagi

Received: 22 August 2009 / Accepted: 2 December 2009 / Published online: 18 December 2009

Abstract Rice is one of the pillars of world-wide food security. Improvement in its yield is necessary to mitigate hunger of millions of people who depend on rice as a staple. Decoding rice genome sequence is expected to complement efforts being made to improve rice and its yield. The information about more than 32,000 genes, regulatory elements, repeat DNA, and DNA markers opens-up new horizons for molecular analysis and genetic enhancement not only for rice but also for other cereal crops. In the post-genomic era, significant progress has been made on defining transcriptome and epigenome as well as gene discovery by way of forward and reverse genetic approaches. Efforts are on to fill the gap between the genome and the phenotype. This may lead to regular practice of genomics-assisted breeding of rice.

S. Raghuvanshi · S. Kapoor · P. Khurana · J. Khurana
Interdisciplinary Centre for Plant Genomics and Department of Plant Molecular Biology, University of Delhi, South Campus, New Delhi 110021, India

M. Kapoor
University School of Biotechnology, GGS Indraprastha University, Delhi 110006, India

S. Tyagi
Botany Department, Gargi College, New Delhi 110049, India

A. Tyagi (✉)
National Institute of Plant Genome Research, Aruna Asaf Ali Marg, New Delhi 110067, India
e-mail: akhilesh@genomeindia.org

Keywords Gene function · Genome analysis · Epigenomics · Transcriptomics · Genomics-assisted breeding · Rice

Introduction

Despite significant ongoing efforts to breed improved rice, the world-wide yield of rice has shown signs of stagnation after registering an increase of almost two-fold during the 1960s and the 1990s (www.fao.org, www.irri.org). This is also in part due to dwindling land resources and climate change. It is imperative to increase rice yield commensurate with population growth to fulfill the demand since expectations for such improvement are high for rice in comparison to other major cereals like maize and wheat. Further, qualitative improvement of rice will help alleviate under-nutrition of people of the world.

Rice is the first food crop whose genome has been completely sequenced, more than once and for both *indica* and *japonica* subspecies, reflecting its importance as major source of food world-wide (Goff et al. 2002; Yu et al. 2002, 2005; International Rice Genome Sequencing Project, IRGSP 2005). The map-based sequence of japonica rice revealed information about 370 Mb out of an estimated 389 Mb genome, which is used to provide 'gold standard' of 12 pseudomolecules representing rice chromosomes (Matsumoto et al. 2008). The smaller genome size, 6

and 40 times smaller than maize and wheat, respectively, synteny, and demonstrated potential for genetic manipulation as well as diversity make rice a model system to investigate its genome and for crop improvement. Various aspects of rice genome sequencing and subsequent developments in terms of forward and reverse genetics, proteomics, phylogenomics, and molecular breeding have been reviewed earlier (Vij et al. 2006; Collard et al. 2008; Han and Zhang 2008; Jung et al. 2008; Matsumoto et al. 2008). Here, we present a brief overview of recent developments in rice genomics particularly on genome annotation, transcriptomics, epigenomics and gene discovery.

The rice genome and its annotation

Both *indica* and *japonica* rice genomes have been sequenced. While *indica* genome was sequenced by whole genome shot-gun approach (Yu et al. 2002, 2005), *japonica* rice genome was sequenced by both whole genome shot-gun (Goff et al. 2002) and map-based clone-by-clone (IRGSP 2005) approaches. The map-based clone-by-clone approach involved construction of a high-density linkage map, YAC-based physical map, transcript map and BAC/PAC physical map. The sequence-ready physical map comprised of both PAC and BAC libraries (Chen et al. 2002; Wu et al. 2002). The finished quality sequence (370 Mb) of more than 3,000 BAC/PAC clones represented 95% of the whole genome and covered virtually the entire euchromatic region (IRGSP 2005). Although, significant progress has been made in sequencing centromeric and telomeric regions in rice unraveling their complex architecture, efforts to complete their sequence are still on (Matsumoto et al. 2008). Almost 35% of the genome represents repeat elements. The finished sequence of the genome had a total of 37,544 non-transposable-element-related protein-coding sequences. Interestingly, evidence for widespread and recurrent gene transfer from the organelles to the nuclear genome was observed. Analysis of duplications in the genome revealed three main classes of duplications—whole genome, tandem and background duplications (Paterson et al. 2004; Vij et al. 2006). It was observed that almost 60% of the genome is duplicated and duplications are present in all chromosomes. However, chromosomes 11 and 12 share a recent duplication block. The wealth of SSRs (>18,000) reported has accelerated research on marker-assisted breeding and positional cloning for genes of agronomic importance.

Initial studies on rice genome annotation, largely based on the in silico predictions, over-estimated the number (40,000–50,000) of protein coding genes. Surprisingly, a considerable proportion (~50%) of the predicted genes did not have any homolog in *Arabidopsis*. These genes had an unusually high GC content, smaller size and failed to map to any known ESTs. Subsequent analysis (Bennetzen et al. 2004; Jabbri et al. 2004) suggested that most of these genes were either wrong annotations or transposon related, clearly indicating a review of the annotation strategy. As evidenced by the annotation of *Drosophila* (Misra et al. 2002), human (Imanishi et al. 2004) and *Arabidopsis* genomes (Haas et al. 2002; Wortman et al. 2003), automated annotations need to be manually curated to remove obvious discrepancies and refine the predictions. ESTs and full-length cDNA sequences are invaluable, as they can be used to validate the predicted coding loci. Therefore, whole genome automated annotations need to be refined by manual curation on the basis of known ESTs, fl-cDNAs, MPSS data, known proteins etc. from the same as well as related organisms. Based on a similar basic idea, four major annotation portals of the rice genome are available at Rice Annotation Project-Database (RAP-DB; Rice Annotation Project 2007; rapdb.dna.affrc.go.jp), Osa1, MSU (Ouyang et al. 2007; rice.plantbiology.msu.edu), Beijing Genomics Institute-Rice Information System (BGI-RIS; Zhao et al. 2004; rice.genomics.org.cn/rice) and NCBI-genomes (www.ncbi.nlm.nih.gov/sites/entrez?db=genome). The RAP-DB differentiates the gene models supported by fl-cDNAs from those having only evidence of expression (i.e., no support of fl-cDNA but ESTs/MPSS provide the proof of expression) as well as ab initio prediction without any evidence of expression. RAP-DB can be accessed through GBrowse (Stein et al. 2002), which provides a chromosome-oriented access to the annotations and G-Integra (Imanishi et al. 2004) for global view that integrates information from other plant genomes as well. The database also provides genome-wide comparison of the *japonica* and *indica* rice along with sorghum. The Osa1 database (MSU) also refines the automated prediction of rice genes with the help of

transcript assemblies (Haas et al. 2002). In addition, it provides a 'community annotation facility' wherein research groups can annotate the gene family of their interest and submit it to the database. RAP-DB provides a 'gene id' converter based on overlapping exons in the two databases (Osa1 and RAP-DB) that helps to fetch similar gene models. Osa1 also provides information, based on ESTs, cDNAs, MPSS and SAGE data, about the spatial and temporal expression profile of the predicted gene models. The BGI-RIS has the core dataset based on the *indica* (cultivar 93-11) rice genome. Apart from these, several other databases have been established to provide valuable information to enrich the annotations further. 'Rice *indica* cDNA Database' (RICD; Lu et al. 2008b) has a collection of 20,000 putative full-length cDNAs and >40,000 ESTs isolated from various cDNA libraries of *indica* rice varieties Guangluai 4 and Minghui 63. The cDNAs/ESTs have been mapped to the genome and putative function assigned on the basis of sequence similarity. Similarly, 'Knowledge-based Oryza Molecular biological Encyclopedia' (KOME, Kikuchi et al. 2003) compiles data on >28,000 full length cDNA of *japonica* rice (cv. Nipponbare). Based on these cDNAs 13,046 putative promoter regions in rice have been identified at the Eukaryotic Promoter Database (EPD; www.epd.isb-sib.ch). Subsequent genome-wide computational analysis revealed that only ~19% of rice gene promoters have a 'TATA' box (Civán and Svec 2009). 'OryGenesDB' (Droc et al. 2009) is a resource for reverse genetic studies in rice and contains 1,71,000 flanking sequence tags (FSTs) of rice insertion lines (Tos17, T-DNA and Ac/Ds) available from 10 major sources. It also offers a web-based utility 'Orylink' for an organized search among the three databases, viz. OryGenesDB, Oryza Tag Line and GreenPhylDB. 'Oryza Tag Line' (Larmande et al. 2008) is a collection of phenotypic characteristics of about 30,000 enhancer-trap lines of *Oryza sativa* cv. Nipponbare. On the other hand, GreenPhlyDB is based on the concept of 'phylogenomics' (Eisen and Fraser 2003), i.e., a throughtput analysis combining genomic and phylogenetic data. The database compiles the comparative functional genomics data of rice and *Arabidopsis* and assigns the proteins of both model plants to different orthologous groups. As a result, 6,421 gene families, perhaps the largest collection of plant gene families, have been curated manually. Similarly, 'SALAD database' (Mihara et al. 2009; salad.dna.affrc.go.jp/salad) provides a portal to analyze and compare proteomes of rice, *Arabidopsis, Sorghum*, *Vitis* as well as *Selaginella*, *Physcometrella*, *Chlamydomonas* and *Saccharomyces*. Another phylogenomic database is the 'Rice kinase database' where >1,400 protein kinases have been identified in the rice proteome (Dardick et al. 2007). 'ARACHIPELAGO' is a compilation of information for over 2,500 rice genes known to be involved in response to abiotic stress. Another important resource to enrich genome annotation is 'RiceGeneThresher', a web based utility to identify genes underlying known QTLs in rice (Thongjuea et al. 2009).

Looking deeper in the annotations, the latest release of RAP-DB (Rice Annotation Project 2008) catalogues 30,192 protein coding gene models with evidence of expression, ~27% of which are similar to known proteins (including rice) and 45% could only be annotated by the presence of a protein domain. It may be noted that putative function was only mapped if the database hit was linked to a relevant published study. Besides, 22,022 gene models have been identified by ab initio predictions which do not have any evidence of expression. Availability of genome sequence has spearheaded many individual ventures that have enriched the annotation. Detailed annotations of over 30 gene families have already been deposited to the community annotation facility of Osa1 database. Similarly many gene families like glutamate dehydrogenase gene family, BURP-domain containing gene family, HAk potassium transporter gene family, HSP20 gene family, Argounautes, receptor like cytoplasmic kinase gene family, A20/AN1 zinc-finger domain-containing proteins, and basic leucine zipper (bZIP) transcription factor family (Kapoor et al. 2008; Nijhawan et al. 2008; Vij et al. 2008; Vij and Tyagi 2008; Ding et al. 2009; Ouyang et al. 2009; Yang et al. 2009), to name a few, have been studied in detail.

Apart from identification of the protein coding genes, studies have also been done to identify non-coding small RNA loci in rice. MPSS data for small RNAs (2,953,855 tags) from untreated flower, seedling and stem tissues (Nobuta et al. 2007) has been mapped to the genome and is available at RAP-DB. Over 350 miRNAs have been identified and are available at miRBase (Griffiths-Jones et al. 2008) and RAP-DB. Similarly, other important components of

the genome are the repetitive DNA elements. The 'Oryza repeat database' (a component of Plant Repeat Database) at Osa1 has a compilation of known transposable elements and centromere/telomere associated repeats of rice. The database has 24,966 repeat elements covering >11.0 Mb of the genome. The most abundant are the transposable elements followed by centromere-related, telomere-related, rDNAs and unclassified sequences. Recently, a genome-wide analysis has identified a new active retro-transposon 'Lullaby' from rice calli (Picault et al. 2009).

Thus, there has been a substantial increase in knowledge since the initial release of the rice genome sequence. Various portals have been developed which attempt to integrate the knowledge with a genome-centric view. It is important to complement them with a slight change in perspective and collating information in a protein-centric manner.

Comparative genomics

Comparative studies in the grasses laid the foundation for comparative genomics. Comprehensive data sets are in place for the major crop plants like rice, wheat, maize, barley, sorghum, and oats which provide evidence for the presence of genic colinearity between genomes. This phenomenon of macro-colinearity was first established in seven grass genomes, rice serving as the central reference genome, and is often referred to as the 'Crop Circle' (Devos 2005). This work when extended to DNA sequence level (micro-colinearity), aiding studies of the genic and non-genic regions, has frequent deviations attributed to small scale rearrangements, deletions, or local gene amplifications (Bennetzen 2000; Keller and Feuillet 2000). The rice genome sequence has led to seeding of information not only for positional cloning in other crop plants, but also provides the ability to gain insight into gene family organization. Exceptions to micro-colinearity have provided information into mechanisms involved in evolution of grass genomes.

The elucidation of syntenic relationship of rice with other cereal genomes was considered as one of the major benefits of sequencing the rice genome. Several important genes have been identified in other cereals based on their synteny with rice. These include *Ror2*, a gene conferring resistance to powdery mildew in barley, malting quality QTL in barley, liguleless in *sorghum* and major heading date QTL in ryegrass (Armstead et al. 2004; Han et al. 1998; Zwick et al. 1998). Exceptions to colinearity were observed in *Rpg1*, *Rph7* and *PhD-H1* genes in barley (Dunford et al. 2002; Han et al. 1999; Leister et al. 1998; Brunner et al. 2003). The finished sequences of rice chromosomes 11 and 12 allowed detailed studies of rice–wheat synteny at the gene level and indicated that although synteny is conserved at the gross genome level, microcolinearity may have been disturbed during evolution of the cereals (Singh et al. 2004; The Rice Chromosome 11 and 12 Sequencing Consortia 2005). Although, majority of rice gene models from chromosome 11 mapped to group 4 chromosomes of wheat, indicating a common origin, many of these which mapped to the short arm of wheat chromosome 4A also mapped to the short arms of chromosomes 4B and 4D indicating significant rearrangements. A similar situation was seen in case of chromosome 12 where most of the gene models mapped to wheat group 5 chromosomes (The Rice Chromosome 11 and 12 Sequencing Consortia 2005). A comparative distribution of rice chromosome 11–12 gene homologs to the wheat homoeologous groups indicates different origins of the two chromosomes but does not support the earlier observation of the evolution of 11 and 12 chromosomes via polyploidization (Paterson et al. 2004). The utility of single-copy genes for defining syntenic and colinear regions between rice and wheat has also been emphasized (Singh et al. 2007). Based on microarray experiments, a genome-wide expression map of different tissues identified a vast majority of paralogous genes pairs with neo- and sub-functionalization over a period of time during the course of evolution (Throude et al. 2009).

High resolution of physical maps of rice chromosomes across the 11 wild genomes provides a suitable FPC and web-based platform to access and understand the Oryza genome (www.omap.org). Within the rice subspecies, a 172-kb genomic DNA region associated with the yld1.1 QTL was found to be highly conserved between *O. sativa* ssp. *japonica* and *O. sativa* ssp. *indica* and the common ancestor *O. rufipogon* plays an important role in conservation of synteny in terms of the content, homology, structure, orientation, and physical distance of the predicted 14

genes within this region (Song et al. 2008). Since *Oryza* species have 10 different genome types, including 6 diploid genome types (AA, BB, CC, EE, FF and GG) and 4 allotetraploid genome types (BBCC, CCDD, HHKK and HHJJ), it makes for interesting study dealing with evolutionary and phylogenetic relevance of important genes like the *MONOCULM1*. Thus, 14 different orthologous regions of the *MOC1* locus were recently analyzed by Lu et al. (2009). It was found that transposons were only conserved between genomes of the same type (i.e., AA or BB) and the allotetraploids were observed to be the result of more recent polyploidization involving pseudogenization of duplicated genes caused by large deletions and small frame-shift insertions/deletions, or nonsense mutations.

Despite the differences in genome size amongst the cereals, it is indeed remarkable that the size of gene-rich regions is similar (Feuillet and Keller 1999). Synteny is generally the highest and retroelements lowest in the distal regions of the chromosomes prone to high rates of recombination. Conservation of gene order has also been investigated between sorghum and maize, rice and rye, rice subspecies and with other crop plants (Bennetzen and Ramakrishana 2002; Feuillet and Keller 2002; Bennetzen and Ma 2003; Hackauf et al. 2008). Analysis of the rice genome indicates that more than 90% wheat, barley and maize proteins could be identified in rice (Goff et al. 2002). Analysis of 2,629 maize markers with rice sequences revealed 656 putative orthologs (Salse et al. 2004). Gene rearrangements are common and even amongst closely related species like barley–wheat or maize–sorghum, at least 20% differences are observed. Despite these non-cohesive reports, the rice genome has proved to be a stable genome over a period of time and absorbing extensive rearrangements during the course of evolution.

The recently completed *Sorghum* genome indicates that, as compared to rice, sorghum has ~75% larger heterochromatin DNA, inferring that euchromatin is 252 and 309 Mb, respectively, in sorghum and rice (Paterson et al. 2009). The net increase in size of sorghum is largely due to LTR retrotransposons, and sorghum resembles rice in having a higher ratio of *gypsy*-like to *copia*-like elements (3.7–1 and 4.9–1) than maize (1.6–1). The major deviations are in the C4 biosynthetic pathway, the NBS-LRR proteins and the cell wall biogenesis pathways. Some characteristic drought-related adaptations, which differ with rice, are represented by the miRNA 169 g for which five homologs are present in sorghum.

Nonetheless, with the recent completion of the high-quality genetically and physically anchored sorghum genome and the imminent completion of the whole genome shot-gun sequence for *Brachypodium distachyon*, coupled with newer tools for functional analysis and massive information generated, make it exciting times for studies on comparative genomics of the grasses (Bossolini et al. 2007). These efforts are greatly aided by development of tools and resources for use in comparative genomics efforts. Chief amongst these are the Gramene (www.gramene.org) database and its ever increasing size as well as use and GRASSIUS (www.grassius.org) for comparative regulatory genomics. The comparative genomics of cereal genomes has led to an attempt to reconstruct 'ancestral cereal genome' defining 'inner circle' recently (Bolot et al. 2009).

Functional genomics of rice

The aim of functional genomic programs is to define molecular function of individual genes, identify both upstream and downstream interacting partners and eventually build regulatory and biochemical networks to understand functioning of a system, be it a cell, a tissue or an organism, in a holistic manner. For any organism, once the genic content is defined by the genome sequencing and annotation programs, the transcriptional units (TUs) need to be validated and the gene products, i.e., transcripts, proteins and metabolites, need to be segregated in temporal, developmental and/or tissue/cell type based co-expression groups, which define the span of individual networks. The members of individual co-expression groups can further be categorized into regulatory and/or biochemical pathways by delineating their function by using various forward and reverse genetics resources and tools. The development of these genetic resources and their easy accessibility to research community, therefore, is of immense importance for the success of any functional genomics initiative (Rensink and Buell 2005; Jung et al. 2008). The forward genetics resources include physically (fast-neutron, gamma-rays, and ion beam irradiation), or

chemically (ethyl methanesulfonate, methyl nitrosourea, or diepoxybutane) generated mutants, insertion mutants (T-DNA or transposon), gene entrapment and activation tag lines. It further requires development of high throughput mutation screening panels and preliminary characterization of the site of mutation by identifying flanking sequence tags (FSTs) in case of insertion mutations to aid reverse genetics-based gene function validation (Krishnan et al. 2009).

Querying the transcriptome

Like in case of yeast, fly, worm, human and *Arabidopsis* genome projects, the quest to understand rice transcriptomes also started with sequencing of Expressed Sequenced Tags (ESTs). During initial phase of the rice genome project, ~29 k ESTs were sequenced that helped in the identification of ~10 k unique cDNAs from various tissues and callus-specific libraries (Uchimiya et al. 1992; Sasaki et al. 1994; Liu et al. 1995; Yamamoto and Sasaki 1997). For unknown genomes, EST information provided the fastest alternative to gain insights into gene structure, expression and function of genes along with genome organization (Fukuoka et al. 1994; Monna et al. 1994; Yamamoto and Sasaki 1997). Since initial studies were based on random selection of clones for sequencing from cDNA libraries, the proportion of redundant clones increased logarithmically with the progress of the program and soon crossed the 50% mark. This situation demanded a strategic course correction in order to enrich the cDNA libraries for rarely expressed transcripts. During this time, rice genome sequencing initiatives were also gaining momentum and so was the need to define the rice transcriptional units to be able to make sense of the genome sequence information. But for defining the transcriptional units EST data was not enough. It required sequence information of the entire cDNA to be able to define the ORFs, the intron–exon boundaries and alternatively spliced transcripts. An ambitious full-length cDNA (FL-cDNA) project was then initiated as part of Japan's Rice Genome Research Program (RGP) and by the time the shot-gun method based draft sequences of *indica* (Yu et al. 2002) and *japonica* (Goff et al. 2002) rice were published, Kikuchi and coworkers were ready with the sequences of 28,469 FL-cDNAs (KOME: http://cdna01.dna.affrc.go.jp/cDNA/) that could be mapped onto these sequences (Kikuchi et al. 2003). 18,933 TUs could be localized on the *japonica* sequence published by Syngenta (Goff et al. 2002) and 5,045 of these sequences were multi-exon TUs, suggesting that they might have originated from alternative splicing, initiation or termination. At this stage, based on rice-specific EST and Fl-cDNA data and known TU from other organism a number of algorithms were developed for gene prediction in rice (described in detail in the previous section on gene annotations). Eventually, map-based sequence of *japonica* rice, Nipponbare, also became available along with improvement in annotation algorithms leading to the refinement of both *indica* and *japonica* draft sequences (Yuan et al. 2005; IRGSP 2005; Ouyang et al. 2007). Further additions to FL-cDNA database resulted in increase in the number of mapped TUs to 32,775 (TIGR4), 32,730 (IRGSP4) and 30,162 (93-11, BGI) (Satoh et al. 2007).

High throughput technologies for transcriptomic research

As the data related to structural genomics was accumulating, it became imperative that studies related to understanding function of the genes also gained pace. High throughput transcriptome analysis technologies, namely, serial analysis of gene expression (SAGE), massively parallel signature sequencing (MPSS) and microarrays, were still in their nascent stage (Schena et al. 1995; Velculescu et al. 1995; Brenner et al. 2000; Reinartz et al. 2002) The SAGE and MPSS (sequence-based technologies) had the advantage of being open systems, which were capable of detecting both known and unknown transcripts, alternative spliced forms, as well as, antisense transcripts. The hybridization-based microarray technology, even though, was effective only in case of known transcripts, its lower direct costs (5–10 time lower than SAGE and MPSS), higher throughput and higher specificity made it the favorite for gene expression analysis research (Wang 2007). The initial rice microarrays consisted of a small number [1,265 (Yazaki et al. 2000); 1,728 (Kawasaki et al. 2001); 8,987 (Yazaki et al. 2003)] of specifically amplified cDNAs. The unsuitability of these arrays for high-

throughput transcriptomic analysis soon became evident due to lack of reproducibility in manufacturing process and cross-hybridizations resulting from lack of precise control over target sequence selection (Kikuchi et al. 2007). The second phase of arrays were based on 25–60 mer oligonucleotides from unique regions of the transcripts that were either chemically synthesized on glass slides (Agilent 22K Rice Gene Expression Arrays), synthesized on silicon wafer by photolithography technology (Affymetrix 57K GeneChip Rice) or printed on glass slides using inkjet technology (Beijing Genomics Institute, 61K; NSF Rice Oligonucleotide Array Project, 20K and later 45K). Although, these arrays were primarily designed for assaying transcript abundance, other uses of these chips, including genome-wide polymorphism (Hazen and Kay 2003; Cui et al. 2005; Edwards et al. 2008), and copy number estimations (Skvortsov et al. 2007) have also been suggested. Recent advances in microarray technologies have made it possible to accommodate even larger number of probes per unit area, thereby making it possible to have the entire genome placed on manageable number of microarrays. Since in these slides (called tiling arrays) the entire genomic information is placed on microarrays in a sequential, unbiased manner, they can be used to query the transcriptome in terms of number of TUs, exon usage diversity, and antisense transcription in a more comprehensive manner (Li et al. 2005, 2006; Stolc et al. 2005). Li et al. (2007) used 37 of these chips covering the entire rice genome to find 25,352 and 27,744 transcriptionally active regions (TARs) from non-exonic regions indicating the presence of uncharacterized splice variants or regions of incompletely annotated genes, antisense transcripts, duplicated gene fragments, or potential non-coding RNAs.

The sequence-based analysis of transcriptomes started with SAGE, which later matured into RobustlongSAGE (Gowda et al. 2004) and then SuperSAGE (Matsumura et al. 2006), where the abundance of short sequence TAGs from the 3′ end of individual transcripts was taken as measure of their relative abundance. Although SAGE was developed on a sound logic that offered unbiased analysis of transcriptomes, it did not become popular with the researchers because of higher operational costs and limited number of transcriptional units assayed. The ideology behind SAGE, however, persisted in the form of MPSS and it proved that this approach was as important for transcriptomics as microarrays. But, the MPSS too remained out-of-bound for many researchers for being costly, and therefore, not many time-points could be analyzed using this technology. Nevertheless, MPSS libraries for 32 human samples, 20 rice samples, and 17 *Arabidopsis* samples derived from various tissues and/or physiological/developmental states were generated, making it an indispensable resource for transcriptomic research (Jongeneel et al. 2005; Nakano et al. 2006; Nobuta et al. 2007). With the advent of 'next-generation' sequencing technologies, genome Sequencer FLX from 454 Life Sciences/Roche, Illumina Genome Analyzer and Applied Biosystem's SOLiD, the sequence-based transcriptome analysis concept seems to have come of age (Lister et al. 2009). All these technologies are capable of generating giga-bases of relatively shorter reads (50–400) in a single run, thereby, not only enhancing the resolution of existing concepts like, transcriptome (mRNA/small RNA) profiling, alternative splicing, DNA methylation, genome re-sequencing, etc. but are also revealing newer ways to unravel genomic treasures, e.g., RNA-sequencing (Ozsolak et al. 2009; Wang et al. 2009b) and deep cap analysis gene expression (CAGE) for genome-wide identification of promoters and quantification of their expression (de Hoon and Hayashizaki 2008).

Understanding function of genes the high-throughput way

So far, there is no technology available for high throughput validation of gene function, but what all the present day transcriptomics technologies have been able to achieve is to have created bins for segregating genes based on co-expression patterns. It is hoped that, commensurate to our knowledge of individual gene functions, it will be possible to segregate these bins further into smaller bins of biochemical/regulatory "direct linkage groups" and use them for building the network of life.

Hormone responsive genes, including those for gibberellins (GA), abscisic acid (ABA) and brassinosteroids (BRs), were the first targets of transcriptome profiling studies in rice. Besides identifying specifically and commonly affected genes these studies also

helped in validating the respective promoter regions involved in hormone response (Yazaki et al. 2003; Yang et al. 2004). A large number of global expression profiling studies have targeted abiotic stress tolerance genes that show differential expression in response to salt, drought or cold stress, highlighting the effect of these factors on rice productivity and global economy (Rabbani et al. 2003). The genes involved in salt stress have been identified by comparing the transcriptome profiles of salt sensitive (IR29) and salt tolerant (Pokkali) cultivars (Kawasaki et al. 2001) and in vegetative and reproductive tissues (Zhou et al. 2007) under stressed and unstressed conditions. Since cold temperatures negatively affect male fertility in rice, microarray-based expression profiling was used to identify ~160 differentially accumulated transcripts in the anthers of cold-stressed plants (Yamaguchi et al. 2004). To assay the impact of high temperatures on grain filling metabolism when transcriptomes of unstressed and heat-stressed developing seed were compared, pronounced effects were observed on the expression of genes involved in starch biosynthesis (Yamakawa et al. 2007). Besides abiotic stresses, rice blast and sheath blight are major factors affecting productivity of rice. To understand the molecular basis of host-pathogen (*Magnaporthe grisea/Rhizoctonia solani*) relationships, studies have been carried out using microarrays and RL-SAGE (Shim et al. 2004; Kim et al. 2005; Soderlund et al. 2006; Venu et al. 2007). Using these studies, Shimono et al. (2007) were able to associate *WRKY45* gene to benzothiadiazole (BTH) activated protection of plants from blast by activating the salicylic acid (SA) signaling pathway.

Global expression profiles have also been generated at various stages/tissue of vegetative and reproductive development to identify genes involved in control of developmental phases and manifestation of developmental stage-specific tissue and organs. Rice FL-cDNA project pioneered in generating and cataloging EST/cDNA based tissue/organ-specific transcriptomes (Kikuchi et al. 2003). Subsequently, various stages of panicle and seed development were queried by microarray-, SAGE- and MPSS- (Nakano et al. 2006) based methods at the level of stage of development (Furutani et al. 2006), organ (Endo et al. 2004; Wang et al. 2005; Li et al. 2007), or even single cell types (Hirano et al. 2008; Suwabe et al. 2008; Jiao et al. 2009) isolated by laser dissection microscopy. These analyses indicated towards involvement of certain gene families of transcription factors and those coding for signal transduction components, e.g., AUX-IAA, bZIP, C_2H_2 zinc finger, CDPKs, F-box, homeobox, HSP20 MADS-box, Argonautes etc. which were analyzed in detail to identify developmental stage-/tissue-specific key members (Agarwal et al. 2007; Arora et al. 2007; Jain et al. 2007; Ray et al. 2007; Kapoor et al. 2008; Nijhawan et al. 2008; Ouyang et al. 2009).

Genome-wide expression profiles generated from F1 hybrids and their inbred parents have been exploited to help solve the century old puzzle of heterosis (Swanson-Wagner et al. 2006). It is believed that categorization of differential expression profiles into additive and non-additive modes followed by their association to vigor and productivity-related biochemical pathways might hold the key to the understanding of molecular basis of heterosis. Such expression analyses have been carried out on seedlings, roots, leaves, panicles and embryos of maize and rice to identify the set of differentially expressed genes and assess their linkage to yield related quantitative trait loci (Meyer et al. 2007; Hoecker et al. 2008; Wei et al. 2009).

Various approaches used for gene function analysis also involve use of transgenic rice system either to generate tagged mutants or to validate function by complementation or by raising gene overexpression/suppression lines (Kathuria et al. 2007). This has generated a wealth of information about genes with possible functions in stress tolerance, quality control, yield and plant development. In addition, a large number of regulatory elements have been evaluated for their activity in transgenic rice which could be utilized for stage-/state-specific expression of desirable genes. Complementing these efforts, gene tagging and protein level interactions have also identified a few agronomically useful genes (Jung et al. 2008). Such efforts need to be intensified to identify function of a large number of genes in conjunction with map-based cloning.

Epigenomics

The expressed or suppressed state of any gene is further governed by covalent modifications such as, methylation, acetylation, ubiquitination and phosphorylation

of DNA and the underlying histone proteins that are mediated by regulatory proteins or small non-coding RNAs (Strahl and Allis 2000; Jenuwein and Allis 2001; Fischle et al. 2003). Combination of these modifications on the chromatin encode a layer of information over and above the genetic constitution of a cell that is heritable and at the same time sensitive to genetic and environmental cues. The information contained in this epigenome regulates tissue-/state-specific expression of genes in different cell types.

The availability of complete genome sequence of two model plants, rice and *Arabidopsis* and the advances in high throughput techniques for studying functional genomics in a holistic manner have provided unprecedented opportunity to study the impact of processes like DNA methylation in modulating plant developmental processes. Cytosine DNA methyltransferases are the key enzymes that catalyze the transfer of a methyl group from S-adenosyl L-methionine (AdoMet) to N4 or C5 position of cytosines. The rice genome harbors a total of 10 genes that encode the conserved catalytic methyltransferase domain (M. Kapoor et al., unpublished results). These genes can be grouped along with *de novo* and maintenance methyltransferases identified in *Arabidopsis* and other organisms indicating that rice too utilizes the same set of DNA modifying enzymes for mediating epigenetic modification at the DNA level. While the biological roles of maintenance methyltransferase, *MET1*, de novo methyltransferase, *DRM2* and *DRM3* and chromo domain containing *CMT3* have been extensively studied in *Arabidopsis*, information about rice proteins is beginning to unfold (Finnegan and Kovac 2000; Lindroth et al. 2001; Cao and Jacobsen 2002; Chan et al. 2005; Xiao et al. 2006; Mathieu et al. 2007). To date, two methyltransferases, *OsMET1-1* (*OsMET1a*) and *OsMET1-2* (*OsMET1b*), have been cloned and characterized in rice (Teerawanichpan et al. 2004; Yamauchi et al. 2008). Expression of *OsMET1-2* was observed to be higher than that of *OsMET1-1* and, similar to animal *Dnmt1*, transcription of *OSMET1-2* produced alternatively spliced transcript forms that differed in the usage of 5' exons (Yamauchi et al. 2008). Functional analysis of *OsMET1* by RNAi approach has demonstrated that its inactivation does not affect de novo methylation at CpG locations in the genome. In addition, in vitro catalytic activity of the purified protein revealed its preference for hemi-methylated DNA, thereby, suggesting that *OsMET1* functions as the major maintenance methyltransferase in rice (Teerawanichpan et al. 2004; Miki and Shimamoto 2008).

The distribution and correlation of histone and DNA methylation with structural features of chromatin and regulation of gene transcription on two rice chromosomes, 4 and 10, was recently described by using tiling-path microarray (Li et al. 2008). DNA methylation along 77.5% of the length of these chromosomes was observed to be positively correlated with heterochromatin formation. In euchromatin, combinatorial interaction of DNA, H3K4me2 and H3K4me3 methylation was observed to be responsible for distinct expression states in cultured cells and differentiated shoot samples. As observed in other organisms, in rice too, gene body methylation was observed to have greater impact on transcriptional activity than promoter methylation. It was observed in rice that while cytosines at CG are methylated uniformly in genes in all cells, CNG and CNN methylation is more dynamic and dictates tissue-specific expression in different cell types.

Transposable elements, both DNA type (class II elements) and retrotransposons (class I elements) are known to contribute towards evolution of genomes and genes (for reviews, see Feschotte et al. 2002; Kazazian 2004). In rice, Tos17, a copia-like retrotransposon containing Long Terminal Repeats (LTR) and mPing, a miniature inverted-repeat DNA transposable element (MITE) are known to transpose randomly in the genome when activated by developmental or environmental cues (Hirochika et al. 1996; Jiang et al. 2003; Kikuchi et al. 2003; Miyao et al. 2003; Nakazaki et al. 2003). Transposition of both Tos17 (in *japonica*) and mPing (in *indica*) has been shown to be correlated with changes in cytosine DNA methylation patterns of the flanking sequences (Hirochika et al. 1996; Ngezahayo et al. 2009). Tos17 is present in 2–5 copies in the rice genome and is known to transpose into genic regions three times more frequently than in intergenic regions when activated under developmental or stress conditions (Hirochika et al. 1996; Miyao et al. 2003).

In rice, Tos17, a copia-like retrotransposon containing Long Terminal Repeats (LTR) is present in 2–5 copies and is known to transpose into genic regions three times more frequently than in intergenic regions when activated by developmental or environmental

cues (Hirochika et al. 1996; Miyao et al. 2003). This property of Tos17 has been exploited for functional genomics studies for studying gene function by gene disruption (Miyao et al. 2003; Hirochika et al. 2004). Transcriptional and transpositional activation of Tos17 under prolonged tissue culture conditions is accompanied by demethylation of its sequences that are otherwise methylated in mature plants (Hirochika et al. 1996). By a series of elegant experiments it has been shown that DNA methylation at Tos17 locus is modulated by the methylation state of underlying H3K9 that requires *SDG714*, a rice SET domain encoding gene, that functions as histone H3K9 methyltransferase (Ding et al. 2007). This gene is closely related to KYP/SUVH4, the major Su(var)3–9 class of histone methyltransferase in *Arabidopsis* (Jackson et al. 2002; Malagnac et al. 2002). SDG714 localization studies in *Arabidopsis* roots and transient expression in onion epidermal cells showed that SDG714 was specifically localized in nucleus where it was found to be enriched in the hetrochromatin region of the centromeres. Gene knockout mutants of *SDG714* displayed reduced levels of H3K9 dimethylation and a loss of both CG and CNG methylation at Tos17 locus. This was correlated with increase in transcription and copy number of Tos17 in the transformants. This is the first report in rice that has provided experimental evidence linking DNA and histone methylation with transcriptional and transpositional activation of a retrotransposon.

Many eukaryotes, including plants, possess gene silencing machinery in which small non-coding RNAs act as key players that link transcriptional gene silencing by DNA methylation with post transcriptional gene silencing via RNA degradation. Large body of information relating to expression and target genes of these small RNAs have been generated using the novel high-throughput deep-sequencing techniques that have revolutionized functional genomic studies. At least five classes of small RNA population have been characterized and these include microRNA (miRNA), small interfering RNA (siRNAs), heterochromatic RNA, trans-acting siRNA (ta-siRNA), natural antisense siRNA (nat-siRNA) and in metazoan, Piwi interacting RNAs (Vazquez et al. 2004; Meins et al. 2005; Vaucheret 2006; Zhang et al. 2006b; O'Donnell and Boeke 2007). Recently, natural antisense microRNA (nat-miRNA) were identified in rice and these were observed to be derived from processing of large intron containing precursors of antisense transcripts of miRNA genes (Lu et al. 2008a). While endogenous siRNAs synthesized by combinatorial activities of components of RNA interfering (RNAi) machinery such as Dicer-like, Argonautes and RNA-dependent RNA polymerases are known to direct chromatin modification and DNA methylation, miRNAs represent a novel class of non-coding RNA that have been implicated in regulating expression of genes involved in developmental and environmental stresses in rice and other plants (Zhang et al. 2006b; Liu et al. 2007; Nagasaki et al. 2007; Kapoor et al. 2008). miRNAs are small RNA molecules of about 21 nt in length that have the potential of base pairing with their target RNAs and mediate their cleavage or translational repression. Biogenesis of these molecules involves transcription from independent or clustered micro RNA genes located in intergenic or intronic regions of host genes to form pri-miRNA precursors (Cui et al. 2008). These molecules are then processed in the nucleus to form a partial stem loop precursor (pre-miRNA) that is further processed into small 21 nt RNAs by ribonuclease enzyme Dicers in animals and Dicer-like1 (*OsDCL1*) in rice and other plants (Liu et al. 2005). The single stranded miRNA then associate with Argonaute proteins in large protein complexes (RNA-induced Silencing Complexes, RISC) in the cytoplasm and recognize their target RNAs with perfect or near perfect complementarity and initiate their degradation or repression. Many conserved and non-conserved miRNA genes have been identified that are activated in response to abiotic stresses such as drought, salinity and heavy metals and phytohormone treatments (Sunkar et al. 2005; Liu et al. 2009; Zhu et al. 2008; Zhao et al. 2009). Rice genome encodes more number of miRNAs than the *Arabidopsis* genome ($\sim$350 in rice, $\sim$184 in *Arabidopsis*). Out of these, $\sim$90 miRNA are not conserved between the two plants suggesting species-specific roles of these regulatory RNA molecules. In silico studies on genomic distribution and promoter analysis of 212 rice miRNA genes have revealed that more than 90% of the genes (202 out of 212) possessed either single or multiple promoters that contained the conserved TATA box in their core promoters similar to pol II transcribed promoters of protein-coding genes (Cui et al. 2008). Recent studies have provided interesting insight into links between miRNA-mediated

regulation of developmental genes during embryogenesis and the molecular players of RNAi machinery in rice (Nagasaki et al. 2007). Loss of function mutants of *OsRDR6* (*SHL2*), *OsAGO7* (*SHL4*/*SHO2*) and *OsDCL4* (*SHO1*) show impaired Shoot Apical Meristem (SAM) formation and abnormal leaf development during embryogenesis. Microarray analysis of gene expression in wild type and mutant plants revealed that in the mutants down regulation of rice homeodomain-leucine zipper family (HD-ZIPIII) genes, *OSHB1* and *OSHB2*, was predominantly observed. This class of genes has previously been implicated in SAM initiation and maintenance during embryogenesis in *Arabidopsis* (Emery et al. 2003; Prigge et al. 2005). The negative regulation of *OSHB1* and *OSHB2* was further observed to be mediated by miR166 that over accumulated in each of *shl* and *sho* mutants. miR166 belongs to miR165/166 gene family and both *OSHB1* and *OSHB2* possess recognition sequence for binding this miRNA.

Binding of miRNAs with near perfect and perfect complementarity to their target molecules in plants has been exploited for development of molecular tools for comparative genomic studies, validation of gene function and for improvement of agronomic traits of food crops. Artificial miRNA (amiRNA) technology has been developed for silencing genes of interest in both rice and *Arabidopsis* (Schwab et al. 2006; Warthmann et al. 2008). This methodology utilizes an endogenous miRNA precursor that is cloned in a vector and then modified by replacing its stem-loop sequence with artificially designed miRNAs of known sequence using overlap PCR. Once introduced into plants by *Agrobacterium*-mediated transformation or by other standard methods, the modified miRNA precursor is processed along with other endogenous miRNA precursors resulting in generation of miRNAs with desirable sequences that will target the gene of interest and will mediate either its cleavage or translational repression.

Another significant contribution in rice functional epigenomics has been the identification and characterization of a germ-line specific Argonaute encoding gene, *MEL1* (MEIOSIS ARRESTED AT LEPTOTENE1). Argonautes are the effector molecules possessing PAZ and PIWI domains and are part of every RISC complex where they act as slicer molecules. These proteins in *Drosophila* (PIWI), *Caenorhabditis elegans* and mice (MIWI, MILI and MIWI2) are known to be involved in sexual reproduction and play roles in spermatogenesis and oocyte formation. *Mel1* mutants identified by screening seed-sterile mutants, which were generated by somatic culture, displayed abnormal meiosis (Hirochika et al. 1996; Yamazaki et al. 2001; Nonomura et al. 2007). Specifically, cells were observed to be arrested at leptotene stage of meiosis I during sporogenesis in anthers and female gametogenesis was also affected at pre-meiosis, meiosis and tetrad stages. In addition, loss of H3K9 dimethyaltion at pericentromeric positions was also observed.

Molecular breeding

The conventional plant breeding has contributed immensely towards improvement of yield and providing sustainability. In this era of genomics, molecular markers offer unprecedented opportunity for precision breeding. This can also help ensemble many desirable combinations of genes with a greater efficiency vis-à-vis conventional plant breeding. In the early years, when the molecular markers gained prominence, restriction fragment length polymorphism (RFLP) and random amplified polymorphic DNA (RAPD) markers were used routinely for many crop plants, including rice (Collard et al. 2008; Collard and Mackill 2008). They were subsequently converted into PCR-based markers called sequence tagged site (STS). In the due course of time, the simple sequence repeats (SSR) or microsatellite markers gained prominence since they were codominant, highly polymorphic and reproducible (Gupta and Varshney 2000). The availability of a high quality genome sequence of rice (IRGSP 2005) helped in mining a rather large number of SSR markers. The sheer number and the high density of SSRs make them highly suitable for molecular mapping and marker-assisted selection (MAS). The comparison of the genome sequences of the *japonica* and the *indica* rice cultivars has lead to the identification of single nucleotide polymorphisms (SNPs) (Feltus et al. 2004; Shen et al. 2004), the potential markers of choice in the years to come. More SNPs have also been identified recently by generating partial sequences of defined region of related genotypes of rice and drawing comparison with the

japonica and *indica* rice reference genomes (Monna et al. 2006; Shirasawa et al. 2007).

The enormity of the work involved in conventional breeding programmes and the complexity of the selection required, can indeed be tackled using these new tools judiciously, which are not only reliable but also inexpensive. MAS has many applications not only in rice breeding but also in genetic diversity assessment, identifying genotypes, marker-assisted backcrossing and gene pyramiding (Collard and Mackill 2008). For example, various molecular markers have been successfully deployed in the genetic diversity assessment of Indian aromatic rice (Jain et al. 2004), establishing the identity of the traditional basmati (Nagaraju et al. 2002), hybrid rice breeding (Cho et al. 2004), and in broadening the genetic base of the US rice varieties (Xu et al. 2004). Likewise, realizing the importance of bacterial blight and blast disease of rice, several efforts have been made for pyramiding the genes for resistance to these two diseases (Hittalmani et al. 2000; Sanchez et al. 2000; Davierwala et al. 2001; Zhang et al. 2006a; Perez et al. 2008).

The information on rice genome has also been used to clone agronomically useful genes by marker-assisted map-based approach. These include genes for tillering, dwarfism, salt tolerance, submergence tolerance, disease resistance, heading date, compatibility, shattering, grain yield and quality (Izawa et al. 2003; Han and Zhang 2008; Sakamoto and Matsuoka 2008; Fitzgerald et al. 2009; Huang et al. 2009; Wang et al. 2009a). Such genes and QTLs would be of great value for breeding to improve rice in the years to come.

Prospects

One of the primary aims of ongoing investigations in the area of rice genomics is to understand gene function and regulatory networks. Major limitation of functional redundancy needs to be overcome by multi-target mutation and gene silencing. Although a large number of insertion mutants are available in rice, more information is required about flanking sequence tags (FSTs) to determine their relationship to target genes. This could be helped by new approaches to genome sequencing at low cost and suitable DNA pooling. Also, this needs to be combined with TILLING and site-specific gene silencing to reach inaccessible genes. Such knowledge about rice genes would greatly impact research on other syntenic genomes of crop species. New layers of regulatory control represented in epigenomes should be unraveled and integrated with transcriptional and translational control circuits. This entails cell type and stimulus specific atlas of transcripts and proteins. The diversity of *Oryza* genomes and functional allelic variation needs to be incorporated in molecular breeding programs to generate improved phenotypes. Large scale screening of diverse germplasm, generation of high-density molecular markers like SNPs and their marriage with breeding efforts is required. The cost of using DNA markers need to be reduced tremendously to help their efficient use in breeding. The concerted effort of a large number of scientists world-wide (Zhang et al. 2008) is required for generating/analyzing enabling tools/resources, functional annotation, regulatory networks, interactome, diversity, bioinformatics, and genomics-assisted breeding. As we overcome practical impediments and integrate molecular biology to breeding activity (Collard et al. 2008), we hope to reap the benefit of genomics research for crop improvement.

Acknowledgments Our research is supported by DBT, DST and UGC, Government of India.

References

Agarwal P, Arora R, Ray S et al (2007) Genome-wide identification of C2H2 zinc-finger gene family in rice and their phylogeny and expression analysis. Plant Mol Biol 65:467–485

Armstead IP, Turner LB, Farrell M et al (2004) Synteny between a major heading-date QTL in perennial ryegrass (*Lolium perenne* L.) and the *Hd3* heading-date locus in rice. Theor Appl Genet 108:822–828

Arora R, Agarwal P, Ray S et al (2007) MADS-box gene family in rice: genome-wide identification, organization and expression profiling during reproductive development and stress. BMC Genomics 8:242

Bennetzen JL (2000) Comparative sequence analysis of plant nuclear genomes: microcolinearity and its many exceptions. Plant Cell 12:1021–1029

Bennetzen JL, Ma J (2003) The genome colinearity of rice and other grasses on the basis of genome sequence analysis. Curr Opin Plant Biol 6:128–133

Bennetzen JL, Ramakrishana W (2002) Numerous small rearrangements of gene content, order and orientation differentiate grass genomes. Plant Mol Biol 48:821–827

Bennetzen JL, Coleman C, Liu R et al (2004) Consistent overestimation of gene number in complex plant genomes. Curr Opin Plant Biol 7:732–736

Bolot S, Abrouk M, Masood-Quraishi U et al (2009) The 'inner circle' of the cereal genomes. Curr Opin Plant Biol 12:119–125

Bossolini E, Wicker T, Knobel PA et al (2007) Comparison of orthologous loci from small grass genomes *Brachypodium* and rice: implications for wheat genomics and grass genome annotation. Plant J 49:704–717

Brenner S, Johnson M, Bridgham J et al (2000) Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. Nat Biotechnol 18:630–634

Brunner S, Keller B, Feuillet C (2003) A large rearrangement involving genes and low-copy DNA interrupts the microcolinearity between rice and barley at the *Rph7* locus. Genetics 164:673–683

Cao X, Jacobsen SE (2002) Locus-specific control of asymmetric and CpNpG methylation by the DRM and CMT3 methyltransferase genes. Proc Natl Acad Sci USA 4:16491–16498

Chan SW, Henderson IR, Jacobsen SE (2005) Gardening the genome: DNA methylation in *Arabidopsis thaliana*. Nat Rev Genet 6:351–360

Chen M, Presting G, Barbazuk WB et al (2002) An integrated physical and genetic map of the rice genome. Plant Cell 14:537–545

Cho Y-I, Park C-W, Kwon S-W et al (2004) Key DNA markers for predicting heterosis in F1 hybrids of japonica rice. Breed Sci 54:389–397

Civán P, Svec M (2009) Genome-wide analysis of rice (*Oryza sativa* L. subsp. *japonica*) TATA box and Y Patch promoter elements. Genome 52:294–297

Collard BC, Mackill DJ (2008) Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philos Trans R Soc Lond B Biol Sci 363:557–572

Collard BCY, Cruz CMV, McNally KL et al (2008) Rice molecular breeding laboratories in the genomic era: current status and future considerations. Int J Plant Genomics, Article Id 524847

Cui X, Xu J, Asghar R et al (2005) Detecting single-feature polymorphisms using oligonucleotide arrays and robustified projection pursuit. Bioinformatics 21:3852–3858

Cui X, Xu XM, Mu DS et al (2008) Genomic analysis of rice microRNA promoters and clusters. Gene 431:61–66

Dardick C, Chen J, Richter T et al (2007) The rice protein kinase database for the rice kinome. Plant Physiol 143:579–586

Davierwala AP, Reddy AP, Lagu MD et al (2001) Marker assisted selection of bacterial blight resistance genes in rice. Biochem Genet 39:261–278

De Hoon M, Hayashizaki Y (2008) Deep cap analysis gene expression (CAGE): genome-wide identification of promoters, quantification of their expression, and network inference. Biotechniques 44:627–632

Devos KM (2005) Updating the 'crop circle'. Curr Opin Plant Biol 8:155–162

Ding Y, Wang X, Su L et al (2007) SDG714, a histone H3K9 methyltransferase, is involved in Tos17 DNA methylation and transposition in rice. Plant Cell 19:9–22

Ding X, Hou X, Xie K et al (2009) Genome-wide identification of BURP domain-containing genes in rice reveals a gene family with diverse structures and responses to abiotic stresses. Planta 230:149–163

Droc G, Périn C, Fromentin S et al (2009) OryGenesDB (2008) update: database interoperability for functional genomics of rice. Nucleic Acids Res 37:D992–D995

Dunford RP, Yano M, Kurata N et al (2002) Comparative mapping of the barley Ppd-Hi photoperiod response gene region which lies close to a junction between two rice linkage segments. Genetics 161:825–834

Edwards JD, Janda J, Sweeney MT et al (2008) Development and evaluation of a high-throughput, low-cost genotyping platform based on oligonucleotide microarrays in rice. Plant Methods 4:13

Eisen JA, Fraser CM (2003) Phylogenomics: intersection of evolution and genomics. Science 300:1706–1707

Emery JF, Floyd SK, Alvarez J et al (2003) Radial patterning of *Arabidopsis* shoots by classIII HD-ZIP and KANADI genes. Curr Biol 13:1768–1774

Endo M, Tsuchiya T, Saito H et al (2004) Identification and molecular characterization of novel anther-specific genes in *Oryza sativa* L. by using cDNA microarray. Genes Genet Syst 79:213–226

Feltus FA, Wan J, Schulze SR et al (2004) An SNP resource for rice genetics and breeding based on subspecies indica and japonica genome alignments. Genome Res 14:1812–1819

Feschotte C, Jiang N, Wessler SR (2002) Plant transposable elements: where genetics meets genomics. Nat Rev Genet 3:329–341

Feuillet C, Keller B (1999) High gene density is conserved at syntenic loci of small and large grass genomes. Proc Natl Acad Sci USA 96:1342–1353

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

Finnegan EJ, Kovac KA (2000) Plant DNA methyltransferases. Plant Mol Biol 43:189–201

Fischle W, Wang Y, Allis CD (2003) Binary switches and modification cassettes in histone biology and beyond. Nature 425:475–479

Fitzgerald MA, McCouch SR, Hall RD (2009) Not just a grain of rice: the quest for quality. Trends Plant Sci 14:133–139

Fukuoka S, Inoue T, Miyao A et al (1994) Mapping of sequence-tagged sites in rice by single strand conformation polymorphism. DNA Res 1:271–277

Furutani I, Sukegawa S, Kyozuka J (2006) Genome-wide analysis of spatial and temporal gene expression in rice panicle. Plant J 46:503–511

Goff SA, Ricke D, Lan TH et al (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science 296:92–100

Gowda M, Jantasuriyarat C, Dean RA et al (2004) Robust-LongSAGE (RL-SAGE): a substantially improved LongSAGE method for gene discovery and transcriptome analysis. Plant Physiol 134:890–897

Griffiths-Jones S, Saini HK, Van Dongen S et al (2008) miRBase: tools for microRNA genomics. Nucleic Acids Res 36:D154–D158

Gupta PK, Varshney RK (2000) The development and use of microsatellite markers for genetic analysis and plant

breeding with emphasis on bread wheat. Euphytica 113:163–185
Haas BJ, Volfovsky N, Town CD et al (2002) Full-length messenger RNA sequences greatly improve genome annotation. Genome Biol 3:0029.1–0029.12
Hackauf B, Rudd S, van der Voort JR et al (2008) Comparative mapping of DNA sequences in rye (*Secale cereale* L.) in relation to the rice genome. Theor Appl Genet 118:371–384
Han B, Zhang Q (2008) Rice genome research; current status and future perspectives. Plant Genome 1:71–76
Han F, Kleinhofs A, Ullrich SE et al (1998) Synteny with rice-analysis of barley malting quality QTLs and RPG4 chromosomal regions. Genome 41:373–380
Han F, Kilian A, Chen JP et al (1999) Sequence analysis of a rice BAC covering the syntenous barley Rpg1 region. Genome 42:1071–1076
Hazen SP, Kay SA (2003) Gene arrays are not just for measuring gene expression. Trends Plant Sci 8:413–416
Hirano K, Aya K, Hobo T et al (2008) Comprehensive transcriptome analysis of phytohormone biosynthesis and signaling genes in microspore/pollen and tapetum of rice. Plant Cell Physiol 49:1429–1450
Hirochika H, Sugimoto K, Otsuki Y et al (1996) Autonomous retrotransposons of rice involved in mutations induced by tissue culture. Proc Natl Acad Sci USA 93:7783–7788
Hirochika H, Guiderdoni E, An G et al (2004) Rice mutant resources for gene discovery. Plant Mol Biol 54:325–334
Hittalmani S, Parco A, Mew TV et al (2000) Fine mapping and DNA marker-assisted pyramiding of the three major genes for blast resistance in rice. Theor Appl Genet 100: 1121–1128
Hoecker N, Keller B, Muthreich N et al (2008) Comparison of maize (*Zea mays* L.) F1-hybrid and parental inbred line primary root transcriptomes suggests organ-specific patterns of nonadditive gene expression and conserved expression trends. Genetics 179:1275–1283
Huang X, Qian Q, Liu Z et al (2009) Natural variation at the DEP1 locus enhances grain yield in rice. Nat Genet 41:494–497
Imanishi T, Itoh T, Suzuki Y et al (2004) Integrative annotation of 21,037 human genes validated by full-length cDNA clones. PLoS Biol 2:856–875
International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. Nature 436: 793–800
Izawa T, Takahashi Y, Yano M (2003) Comparative biology comes into bloom: genomic and genetic comparison of flowering pathways in rice and *Arabidopsis*. Curr Opin Plant Biol 6:113–120
Jabbari K, Cruveiller S, Clay O et al (2004) The new genes of rice: a closer look. Trends Plant Sci 9:281–285
Jackson JP, Lindroth AM, Cao X et al (2002) Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. Nature 416:556–560
Jain S, Jain RK, McCouch SR (2004) Genetic analysis of Indian aromatic and quality rice (*Oryza sativa* L.) germplasm using panels of fluorescently-labeled microsatellite markers. Theor Appl Genet 109:965–977
Jain M, Nijhawan A, Arora R et al (2007) F-box proteins in rice. Genome-wide analysis, classification, temporal and spatial gene expression during panicle and seed development, and regulation by light and abiotic stress. Plant Physiol 143:1467–1483
Jenuwein T, Allis CD (2001) Translating the histone code. Science 293:1074–1080
Jiang N, Bao Z, Zhang X et al (2003) An active DNA transposon family in rice. Nature 421:163–167
Jiao Y, Tausta SL, Gandotra N et al (2009) A transcriptome atlas of rice cell types uncovers cellular, functional and developmental hierarchies. Nat Genet 41:258–263
Jongeneel CV, Delorenzi M, Iseli C et al (2005) An atlas of human gene expression from massively parallel signature sequencing (MPSS). Genome Res 15:1007–1014
Jung KH, An G, Ronald PC (2008) Towards a better bowl of rice: assigning function to tens of thousands of rice genes. Nat Rev Genet 9:91–101
Kapoor M, Arora R, Lama T et al (2008) Genome-wide identification, organization and phylogenetic analysis of Dicer-like, Argonaute and RNA-dependent RNA polymerase gene families and their expression analysis during reproductive development and stress in rice. BMC Genomics 9:451
Kathuria H, Giri J, Tyagi H et al (2007) Advances in transgenic rice biotechnology. Crit Rev Plant Sci 26:65–103
Kawasaki S, Borchert C, Deyholos M et al (2001) Gene expression profiles during the initial phase of salt stress in rice. Plant Cell 13:889–905
Kazazian HH Jr (2004) Mobile elements: drivers of genome evolution. Science 303:1626–1632
Keller B, Feuillet C (2000) Colinearity and gene density in grass genomes. Trends Plant Sci 5:246–251
Kikuchi S, Satoh K, Nagata T et al (2003) Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice. Science 301:376–379
Kikuchi S, Wang G-L, Li L (2007) Genome-wide RNA expression profile in rice. In: Upadhyaya N (ed) Rice functional genomics: challenges, progress and prospects. Springer, New York, pp 31–54
Kim KM, Cho SK, Shin SH et al (2005) Analysis of differentially expressed transcripts of fungal elicitor and wound-treated wild rice (*Oryza grandiglumis*). J Plant Res 118:347–354
Krishnan A, Guiderdoni E, An G et al (2009) Mutant resources in rice for functional genomics of the grasses. Plant Physiol 149:165–170
Larmande P, Gay C, Lorieux M et al (2008) Oryza Tag Line, a phenotypic mutant database for the Genoplante rice insertion line library. Nucleic Acids Res 36:D1022–D1027
Leister D, Kurth J, Laurie DA et al (1998) Rapid reorganization of resistance gene homologues in cereal genomes. Proc Natl Acad Sci USA 95:370–375
Li L, Wang X, Xia M et al (2005) Tiling microarray analysis of rice chromosome 10 to identify the transcriptome and relate its expression to chromosomal architecture. Genome Biol 6:R52
Li L, Wang X, Stolc V et al (2006) Genome-wide transcription analyses in rice using tiling microarrays. Nat Genet 38:124–129
Li L, Wang X, Sasidharan R et al (2007) Global identification and characterization of transcriptionally active regions in the rice genome. PLoS ONE 2:e294

Li X, Wang X, He K et al (2008) High-resolution mapping of epigenetic modifications of the rice genome uncovers interplay between DNA methylation, histone methylation and gene expression. Plant Cell 20:259–276

Lindroth AM, Cao X, Jackson JP et al (2001) Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation. Science 292:2077–2080

Lister R, Gregory BD, Ecker JR (2009) Next is now: new technologies for sequencing of genomes, transcriptomes, and beyond. Curr Opin Plant Biol 12:107–118

Liu J, Hara C, Umeda M et al (1995) Analysis of randomly isolated cDNAs from developing endosperm of rice (*Oryza sativa* L.): evaluation of expressed sequence tags, and expression levels of mRNAs. Plant Mol Biol 29: 685–689

Liu B, Li P, Li X et al (2005) Loss of function of OsDCL1 affects microRNA accumulation and causes developmental defects in rice. Plant Physiol 139:296–305

Liu B, Chen Z, Song X et al (2007) *Oryza sativa* Dicer-like4 reveals a key role for small interfering RNA silencing in plant development. Plant Cell 19:2705–2718

Liu Q, Zhang YC, Wang CY et al (2009) Expression analysis of phytohormone-regulated microRNAs in rice, implying their regulation roles in plant hormone signaling. FEBS Lett 583:723–728

Lu C, Jeong DH, Kulkarni K et al (2008a) Genome-wide analysis for discovery of rice microRNAs reveals natural antisense microRNAs (nat-miRNAs). Proc Natl Acad Sci USA 105:4951–4956

Lu T, Huang X, Zhu C et al (2008b) RICD: a rice *indica* cDNA database resource for rice functional genomics. BMC Plant Biol 8:118

Lu F, Ammiraju JS, Sanyal A et al (2009) Comparative sequence analysis of MONOCULM1-orthologous regions in 14 *Oryza* genomes. Proc Natl Acad Sci USA 106: 2071–2076

Malagnac F, Bartee L, Bender J (2002) An Arabidopsis SET domain protein required for maintenance but not establishment of DNA methylation. EMBO J 21:6842–6852

Mathieu O, Reinders J, Caikovski M et al (2007) Transgenerational stability of the *Arabidopsis* epigenome is coordinated by CG methylation. Cell 130:851–862

Matsumoto T, Wu J, Antonio B et al (2008) Development in rice genome research based on accurate genome sequence. Int J Plant Genomics, Article Id 348621

Matsumura H, Bin Nasir KH, Yoshida K et al (2006) Super-SAGE array: the direct use of 26-base-pair transcript tags in oligonucleotide arrays. Nat Methods 3:469–474

Meins F Jr, Si-Ammour A, Blevins T et al (2005) RNA silencing systems and their relevance to plant development. Annu Rev Cell Dev Biol 21:297–318

Meyer S, Pospisil H, Scholten S (2007) Heterosis associated gene expression in maize embryos 6 days after fertilization exhibits additive, dominant and overdominant pattern. Plant Mol Biol 63:381–391

Mihara M, Itoh T, Izawa T (2009) SALAD database: a motif-based database o f protein annotations for plant comparative genomics. Nucleic Acids Res Database Issue

Miki D, Shimamoto K (2008) *De novo* DNA methylation induced by siRNA targeted to endogenous transcribed sequences is gene-specific and *OsMet1*-independent in rice. Plant J 56:539–549

Misra S, Crosby M, Mungall C et al (2002) Annotation of the *Drosophila melanogaster* euchromatic genome: a systematic review. Genome Biol 3:0083.1–0083.22

Miyao A, Tanaka K, Murata 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

Monna L, Miyao A, Inoue T 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

Monna L, Ohta R, Masuda H et al (2006) Genome-wide searching of single-nucleotide polymorphisms among eight distantly and closely related rice cultivars (*Oryza sativa* L.) and a wild accession (*Oryza rufipogon* Griff.). DNA Res 13:43–51

Nagaraju J, Kathirvel M, Kumar RR et al (2002) Genetic analysis of traditional and evolved Basmati and non-Basmati rice varieties by using fluorescence-based ISSR-PCR and SSR markers. Proc Natl Acad Sci USA 99:5836–5841

Nagasaki H, Itoh J, Hayashi K et al (2007) The small interfering RNA production pathway is required for shoot meristem initiation in rice. Proc Natl Acad Sci USA 104:14867–14871

Nakano M, Nobuta K, Vemaraju K et al (2006) Plant MPSS databases: signature-based transcriptional resources for analyses of mRNA and small RNA. Nucleic Acids Res 34:D731–D735

Nakazaki T, Okumoto Y, Horibata A et al (2003) Mobilization of a transposon in the rice genome. Nature 421:170–172

Ngezahayo F, Xu C, Wang H et al (2009) Tissue culture-induced transpositional activity of *mPing* is correlated with cytosine methylation in rice. BMC Plant Mol Biol 9:91

Nijhawan A, Jain M, Tyagi AK et al (2008) A genomic survey and gene expression analysis of basic leucine zipper (bZIP) transcription factor family in rice. Plant Physiol 146:333–350

Nobuta K, Venu RC, Lu C et al (2007) An expression atlas of rice mRNAs and small RNAs. Nat Biotechnol 25:473–477

Nonomura K, Morohoshi A, Nakano M et al (2007) A germ cell-specific gene of the ARGONAUTE family is essential for the progression of premeiotic mitosis and meiosis during sporogenesis in rice. Plant Cell 19:2583–2594

O'Donnell KA, Boeke JD (2007) Mighty Piwis defend the germline against genome intruders. Cell 129:37–44

Ouyang S, Zhu W, Hamilton J et al (2007) The TIGR rice genome annotation resource: improvements and new features. Nucleic Acids Res 35:D883–D887

Ouyang Y, Chen J, Xie W et al (2009) Comprehensive sequence and expression profile analysis of Hsp20 gene family in rice. Plant Mol Biol 70:341–357

Ozsolak F, Platt AR, Jones DR et al (2009) Direct RNA sequencing. Nature 461:814–818

Paterson AH, Bowers JE, Chapman BA (2004) Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. Proc Natl Acad Sci USA 101:9903–9908

Paterson AH, Bowers JE, Bruggmann R et al (2009) The *Sorghum bicolor* genome and the diversification of grasses. Nature 457:551–556

Perez LM, Redona ED, Mendioro MS et al (2008) Introgression of *Xa4*, *Xa7* and *Xa21* for resistance to bacterial blight in thermosensitive genetic male sterile rice (*Oryza sativa* L.) for the development of two-line hybrids. Euphytica 164:627–636

Picault N, Chaparro C, Piegu B et al (2009) Identification of an active LTR retrotransposon in rice. Plant J 58:754–765

Prigge MJ, Otsuga D, Alonso JM et al (2005) Class III homeodomain-leucine zipper gene family members have overlapping, antagonistic, and distinct roles in Arabidopsis development. Plant Cell 17:61–76

Rabbani MA, Maruyama K, Abe H et al (2003) Monitoring expression profiles of rice genes under cold, drought, and high-salinity stresses and abscisic acid application using cDNA microarray and RNA gel-blot analyses. Plant Physiol 133:1755–1767

Ray S, Agarwal P, Arora R et al (2007) Expression analysis of calcium-dependent protein kinase gene family during reproductive development and abiotic stress conditions in rice (*Oryza sativa* L. ssp. indica). Mol Genet Genomics 278:493–505

Reinartz J, Bruyns E, Lin JZ et al (2002) Massively parallel signature sequencing (MPSS) as a tool for in-depth quantitative gene expression profiling in all organisms. Brief Funct Genomics Proteomics 1:95–104

Rensink WA, Buell CR (2005) Microarray expression profiling resources for plant genomics. Trends Plant Sci 10: 603–609

Rice Annotation Project (2007) Curated genome annotation of *Oryza sativa* ssp. *japonica* and comparative genome analysis with *Arabidopsis thaliana*. Genome Res 17:175–183

Rice Annotation Project (2008) The rice annotation project database (RAP-DB): 2008 update. Nucleic Acids Res 36:D1028–D1033

Sakamoto T, Matsuoka M (2008) Identifying and exploiting grain yield genes in rice. Curr Opin Plant Biol 11:209–214

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

Sanchez AC, Brar DS, Huang N et al (2000) Sequence tagged site marker-assisted selection for three blight resistant genes in rice. Crop Sci 40:792–797

Sasaki T, Song J, Koga-Ban Y et al (1994) Toward cataloguing all rice genes: large-scale sequencing of randomly chosen rice cDNAs from a callus cDNA library. Plant J 6:615–624

Satoh K, Doi K, Nagata T et al (2007) Gene organization in rice revealed by full-length cDNA mapping and gene expression analysis through microarray. PLoS ONE 2:e1235

Schena M, Shalon D, Davis RW et al (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270:467–470

Schwab R, Ossowski S, Riester M et al (2006) Highly specific gene silencing by artificial microRNAs in Arabidopsis. Plant Cell 18:1121–1133

Shen YJ, Jiang H, Jin JP et al (2004) Development of genome-wide DNA polymorphism database for map-based cloning of rice genes. Plant Physiol 135:1198–1205

Shim KS, Cho SK, Jeung JU et al (2004) Identification of fungal (*Magnaporthe grisea*) stress-induced genes in wild rice (*Oryza minuta*). Plant Cell Rep 22:599–607

Shimono M, Sugano S, Nakayama A et al (2007) Rice WRKY45 plays a crucial role in benzothiadiazole-inducible blast resistance. Plant Cell 19:2064–2076

Shirasawa K, Maeda H, Monna L et al (2007) The number of genes having different alleles between rice cultivars estimated by SNP analysis. Theor Appl Genet 115:1067–1074

Singh NK, Raghuvanshi S, Srivastava SK et al (2004) Sequence analysis of the long arm of rice chromosome 11 for rice-wheat synteny. Funct Integr Genomics 4:102–117

Singh NK, Dalal V, Batra K et al (2007) Single-copy genes define a conserved order between rice and wheat for understanding differences caused by duplication, deletion, and transposition of genes. Funct Integr Genomics 7:17–35

Skvortsov D, Abdueva D, Stitzer M et al (2007) Using expression arrays for copy number detection: an example from *E. coli*. BMC Bioinformatics 8:203

Soderlund C, Haller K, Pampanwar V et al (2006) MGOS: a resource for studying *Magnaporthe grisea* and *Oryza sativa* interactions. Mol Plant Microbe Interact 19:1055–1061

Song BK, Hein I, Druka A et al (2008) The 172-kb genomic region of the *O. rufipogon* yld1.1 locus: comparative sequence analysis with *O. sativa* ssp. japonica and *O. sativa* ssp. indica. Funct Integr Genomics 9:97–108

Stein LD, Mungall C, Shu S et al (2002) The generic genome browser: a building block for a model organism system database. Genome Res 12:1599–1610

Stolc V, Li L, Wang X et al (2005) A pilot study of transcription unit analysis in rice using oligonucleotide tiling-path microarray. Plant Mol Biol 59:137–149

Strahl BD, Allis CD (2000) The language of covalent histone modifications. Nature 403:42–45

Sunkar R, Girke T, Jain PK et al (2005) Cloning and characterization of microRNAs from rice. Plant Cell 17:1397–1411

Suwabe K, Suzuki G, Takahashi H et al (2008) Separated transcriptomes of male gametophyte and tapetum in rice: validity of a laser microdissection (LM) microarray. Plant Cell Physiol 49:1407–1416

Swanson-Wagner RA, Jia Y, DeCook R et al (2006) All possible modes of gene action are observed in a global comparison of gene expression in a maize F1 hybrid and its inbred parents. Proc Natl Acad Sci USA 103:6805–6810

Teerawanichpan P, Chandrasekharan MB, Jiang Y et al (2004) Characterization of two rice DNA methyltransferase genes and RNAi-mediated reactivation of a silenced transgene in rice callus. Planta 218:337–349

The Rice Chromosome 11 and 12 Sequencing Consortia (2005) The sequence of rice chromosomes 11 and 12, rich in disease resistance genes and recent gene duplications. BMC Biol 3:20

Thongjuea S, Ruanjaichon V, Bruskiewich R et al (2009) RiceGeneThresher: a web-based application for mining

genes underlying QTL in rice genome. Nucleic Acids Res 37:D996–D1000

Throude M, Bolot S, Bosio M et al (2009) Structure and expression analysis of rice paleo duplications. Nucleic Acid Res 37:1248–1259

Uchimiya H, Kidou S, Shimazaki T et al (1992) Random sequencing of cDNA libraries reveals a variety of expressed genes in cultured cells of rice (*Oryza sativa* L.). Plant J 2:1005–1009

Vaucheret H (2006) Post-transcriptional small RNA pathways in plants: mechanisms and regulations. Genes Dev 20:759–771

Vazquez F, Vaucheret H, Rajagopalan R et al (2004) Endogenous trans-acting siRNAs regulate the accumulation of Arabidopsis mRNAs. Mol Cell 16:69–79

Velculescu VE, Zhang L, Vogelstein B et al (1995) Serial analysis of gene expression. Science 270:484–487

Venu RC, Jia Y, Gowda M et al (2007) RL-SAGE and microarray analysis of the rice transcriptome after *Rhizocotonia solani* infection. Mol Genet Genomics 278:421–431

Vij S, Tyagi AK (2008) A20/AN1 zinc-finger domain-containing proteins in plants and animals represent common elements in stress response. Funct Integr Genomics 8:301–307

Vij S, Gupta V, Kumar D et al (2006) Decoding the rice genome. BioEssays 28:421–432

Vij S, Giri J, Dansana P et al (2008) The receptor-like cytoplasmic kinase (OsRLCK) gene family in rice: organization, phylogenetic relationship, and expression during development and stress. Mol Plant 1:732–750

Wang SM (2007) Understanding SAGE data. Trends Genet 23:42–50

Wang Z, Liang Y, Li C et al (2005) Microarray analysis of gene expression involved in anther development in rice (O*ryza sativa* L.). Plant Mol Biol 58:721–737

Wang J, Nakazaki T, Chen S et al (2009a) Identification and characterization of the erect-pose panicle gene EP conferring high grain yield in rice (*Oryza sativa* L.). Theor Appl Genet 119:85–91

Wang Z, Gerstein M, Snyder M (2009b) RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10:57–63

Warthmann N, Chen H, Ossowski S et al (2008) Highly specific gene silencing by artificial miRNAs in rice. PLoS One 3:e1829

Wei G, Tao Y, Liu G et al (2009) A transcriptomic analysis of superhybrid rice LYP9 and its parents. Proc Natl Acad Sci USA 106:7695–7701

Wortman JR, Haas BJ, Hannick LI et al (2003) Annotation of the *Arabidopsis* genome. Plant Physiol 132:461–468

Wu J, Maehara T, Shimokawa T et al (2002) A comprehensive rice transcript map containing 6591 expressed sequence tag sites. Plant Cell 14:525–535

Xiao W, Custard KD, Brown RC et al (2006) DNA methylation is critical for *Arabidopsis* embryogenesis and seed viability. Plant Cell 18:805–814

Xu Y, Beachell H, McCouch SR (2004) A marker-based approach to broadening the genetic base of rice in the USA. Crop Sci 44:1947–1959

Yamaguchi T, Nakayama K, Hayashi T et al (2004) cDNA microarray analysis of rice anther genes under chilling stress at the microsporogenesis stage revealed two genes with DNA transposon Castaway in the 5′-flanking regions. Biosci Biotechnol Biochem 68:1315–1323

Yamakawa H, Hirose T, Kuroda M et al (2007) Comprehensive expression profiling of rice grain filling-related genes under high temperature using DNA microarray. Plant Physiol 144:258–277

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

Yamauchi T, Moritoh S, Johzuka-Hisatomi Y et al (2008) Alternative splicing of the rice OsMET1 genes encoding maintenance DNA methyltransferase. J Plant Physiol 165:1774–1782

Yamazaki M, Tsugava H, Miyao A et al (2001) The rice retrotransposon Tos17 prefers low-copy number sequences as integration targets. Mol Genet Genomics 265:336–344

Yang GX, Jan A, Shen SH et al (2004) Microarray analysis of brassinosteroids- and gibberellin-regulated gene expression in rice seedlings. Mol Genet Genomics 271:468–478

Yang Z, Gao Q, Sun C et al (2009) Molecular evolution and functional divergence of HAK potassium transporter gene family in rice (*Oryza sativa* L.). J Genet Genomics 36:161–172

Yazaki J, Kishimoto N, Nakamura K 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

Yazaki J, Kishimoto N, Nagata Y et al (2003) Genomics approach to abscisic acid- and gibberellin-responsive genes in rice. DNA Res 10:249–261

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

Yu J, Wang J, Lin W et al (2005) The genomes of *Oryza sativa*: a history of duplications. PLoS Biol 3:e38

Yuan Q, Ouyang S, Wang A et al (2005) The Institute for Genomic Research Osa1 rice genome annotation database. Plant Physiol 138:18–26

Zhang J, Li X, Jiang G et al (2006a) Pyramiding of Xa7 and Xa21 for the improvement of disease resistance to bacterial blight in hybrid rice. Plant Breed 125:600–605

Zhang B, Pan X, Cannon CH et al (2006b) Conservation and divergence of plant microRNA genes. Plant J 46:243–259

Zhang Q, Li J, Xue Y et al (2008) Rice 2020; a call for international coordinated effort in rice functional genomics. Mol Plant 1:715–719

Zhao W, Wang J, He X et al (2004) BGI-RIS: an integrated information resource and comparative analysis workbench for rice genomics. Nucleic Acids Res 32:D377–D382

Zhao B, Ge L, Liang K et al (2009) Members of miR-169 family are induced by high salinity and transiently inhibit the NF-YA transcription factor. BMC Mol Biol 10:29

Zhou J, Wang X, Jiao Y et al (2007) Global genome expression analysis of rice in response to drought and high-salinity stresses in shoot, flag leaf, and panicle. Plant Mol Biol 63:591–608

Zhu Q-H, Spriggs A, Matthew L et al (2008) A diverse set of microRNAs and microRNA-like small RNAs in developing rice grains. Genome Res 18:1456–1465

Zwick MS, Islam-Faridi MN, Czeschin DG et al (1998) Physical mapping of the liguleless linkage group in *Sorghum bicolor* using rice RFLP-selected sorghum BACs. Genetics 148:1983–1992