Title

Towards a better bowl of rice: assigning function to tens of thousands of rice genes.

Journal

Nature reviews. Genetics, 9(2)

ISSN

1471-0064

Authors

Jung, Ki-Hong
An, Gynheung
Ronald, Pamela C

Publication Date

2008-02-01

Peer reviewed

Towards a better bowl of rice: assigning function to tens of thousands of rice genes

Ki-Hong Jung, Gynheung An‡ and Pamela C. Ronald**

Abstract | Rice, one of the most important food crops for humans, is the first crop plant to have its genome sequenced. Rice whole-genome microarrays, genome tiling arrays and genome-wide gene-indexed mutant collections have recently been generated. With the availability of these resources, discovering the function of the estimated 41,000 rice genes is now within reach. Such discoveries have broad practical implications for understanding the biological processes of rice and other economically important grasses such as cereals and bioenergy crops.

Pseudomolecules
Virtual contiguous sets of clones constructed by resolving discrepancies between overlapping F-factor-based bacterial artificial chromosome (BAC) and P1-derived artificial chromosome (PAC) clones, trimming the overlapping regions at junction points in which the phase 3 BAC–PAC sequences are preferably used, and linking the unique sequences to form a contiguous sequence.

**Department of Plant Pathology, 1 Shields Avenue, UC Davis, Davis, California 95616, USA.*
‡Department of Life Sciences, Pohang University of Science and Technology, Pohang, Republic of Korea 790-784.
Correspondence to P.C.R.
e-mail: pcronald@ucdavis.edu
doi:10.1038/nrg2286
Published online
27 December 2007

Taxonomically, all flowering plants belong to one of two major groups: the monocotyledonous (monocot) or dicotyledonous (dicot) species. Dicots include broad-leaved herbs and trees, as well as *Arabidopsis thaliana*, a species in the mustard family, and the first plant to have its genome sequenced. The monocots include cereal crops in the family *Poaceae* such as rice, wheat, maize, barley, sorghum and oat. These crops provide the bulk of the calorific intake of the world's population[1]. At a compact 389 Mb, the rice genome is one-sixth the size of the maize genome and 40 times smaller than the wheat genome[2], making rice an excellent model for the study of cereal genomes[3–10]. Rice also serves as a model for studies of perennial grasses such as switchgrass and *Miscanthus*, which show promise as feedstocks for biofuel production[11].

Most rice cultivars can be placed within two subspecies of rice: *Oryza sativa* ssp. *japonica* and *Oryza sativa* ssp. *indica*, which differ in physiological and morphological traits[12]. *Indica* rice is usually found in the lowlands of tropical Asia, whereas *japonica* rice is typically found in the upland hills of southern China, northeast Asia, southeast Asia and Indonesia, as well as in regions outside Asia (Africa, North America, Europe and South America)[13]. The current map-based rice genome sequence assembly (372.1 Mb) covers over 95% of the *japonica* genome. The remaining 5% includes 38 physical gaps within the 12 pseudomolecules and gaps at 10 centromeres and 10 telomeres[2,14]. The genome sequence has been subjected to extensive annotation using *ab initio* gene prediction, comparative genomics and various other computational methods[2,15,16]. The Institute for Genomic Research (TIGR) Rice Genome Annotation database and resource currently lists 56,278 genes (loci) (TIGR rice annotation release 5, 2007)[14]. Because 6,498 of these loci encode 10,432 alternative splicing isoforms, the total number of transcripts (or gene models) is 66,710. If the 15,232 transposable element (TE)-related gene models are removed, the total number of rice non-TE-related gene models is currently estimated to be 41,478. The rice genome annotation can also be obtained from the Rice Annotation Project database (RAP-DB)[17,18]. This RAP annotation has been incorporated into other databases such as the NCBI map viewer, the DNA Data Bank of Japan and the European Molecular Biology Laboratory.

A total of 33,882 of these gene models have been empirically validated through methods that characterize RNA transcripts. These include ESTs, full-length cDNA (FL-cDNA) sequences, whole-genome tiling microarrays, gene-expression arrays, serial analysis of gene expression and massively parallel signature sequencing (MPSS). In addition to validating hypothetical gene models, these data have also led to the identification of thousands of new genes[16,19–21].

However, despite the availability of the finished genome sequence and of tools for rice genome analysis, the number of genes that have been functionally characterized in rice lags far behind that of the dicot *A. thaliana*. Extensive efforts so far have revealed the function of only a handful of rice genes and most of these have been identified through laborious map-based cloning (TABLE 1). Although map-based cloning and candidate-gene validation has been facilitated by the availability

Table 1 | **Examples of agriculturally important genes isolated from rice**

Locus or gene	Function	Identification method	Ref(s)
Xa21	Bacterial resistance	Map-based cloning	86
Sub1	Submergence tolerance	Map-based cloning	53
Moc1	Tillering number control	Map-based cloning	87
Pi9	Fungal resistance	Map-based cloning	88
Pi2	Fungal resistance	Map-based cloning	89
Gid1, *Gid2* or *Slr1*	Gibberellin signalling pathway	Map-based cloning	90–92
Sd1	Gibberellin synthesis	Map-based cloning	93
Lsi1	Silicon transport	Map-based cloning	94
qSH1	Grain abscission control	Map-based cloning	95
Spl18	Fungal resistance	Activation tagging	96
Fon1	Tillering number control and the number of seeds	T-DNA	97
Lhs	Floral organ formation and seed setting	T-DNA	98
Udt1	Early anther development	T-DNA	71
Xb3	Bacterial resistance	Yeast two-hybrid	79
NH1	Bacterial resistance	Yeast two-hybrid	99
NRR	Bacterial susceptibility	Yeast two-hybrid	99

Fon1, floral organ number 1; *Gid1*, Gibberellin-insensitive dwarf protein 1; *Gid2*, GA-insensitive dwarf protein 2; *Lhs*, Leafy hull sterile; *Lsi1*, Low silicon rice 1; *Moc1*, Monoculm 1; *NH1*, NPR1 homologue 1; *NRR*, Negative regulator of disease resistance; *Pi2*, pistillate florets 2; *Pi9*, *Magnaporthe grisea* resistance 9; *qSH1*, QTL of seed shattering in chromosome 1; *Sd1*, Semi-dwarf 1; *Slr1*, Slender rice 1; *Spl18*, Spotted leaf 18; *Sub1*, Submergence tolerance 1; *Udt1*, Undeveloped Tapetum 1; *Xa21*, *Xanthomonas oryzae* pv. *oryzae* resistance 21; *Xb3*, Xa21-binding protein 3.

of the finished genome sequence and improvements in *Agrobacterium tumefaciens*-mediated transformation methods, it is still relatively slow. This is because it requires intensive crosses between two genetically distinct rice varieties or subspecies, and genotyping of 500–30,000 F_2 mutants for fine-mapping-analysis studies, which require extensive greenhouse and field space[22].

With duplicated genomic segments estimated to cover 27–65.7% of the rice genome[14,23], rice seems to encode more genes that have a redundant function as compared with *A. thaliana*. This high level of redundancy in the rice genome complicates mutant analysis[24,25]. To speed the pace of gene-function discovery in rice, innovative, integrated and efficient utilization of functional genomic technologies is essential. In this Review, we describe new tools that are available for rice functional genomics analysis with emphasis on publicly available genome-wide gene-indexed mutant collections, rice gene expression microarray and genome tiling array platforms. We also discuss experimental approaches that can be used to elucidate the function of genes that are members of multi-gene families with redundant functions. We conclude with a description of how the integration of multiple tools can facilitate functional analysis. We do not describe rice chromosome organization, the *Oryza* Map Alignment Project, methods of forward and reverse genetics, proteomics approaches or annotation resources for rice because these have already been well reviewed[26–30].

Map-based cloning
A process of identifying the gene responsible for a mutant phenotype by defining a small physical interval through linkage analysis and then systematically testing all candidate genes residing in the interval.

Available rice microarray platforms

Microarray technology allows biologists to measure the expression levels of thousands of genes in a single experiment[31] and to identify transcriptionally active regions (TARs) in the genome. This technology can also be used for genome-wide polymorphism surveys and the identification of mutations[32,33].

Several rice array platforms for the two rice subspecies have been reported and their characteristics are summarized in TABLE 2. The *Oryza sativa* Genome Oligo Set (Version 1.0; 61K) was designed by the Beijing Genomics Institute (BGI) and was based on draft *indica* and *japonica* sequences. The University of California, Davis, USA, led a National Science Foundation (NSF) supported effort to design, print and validate a 45k (45,116) oligonucleotide array based on 61,419 gene-model predictions from TIGR's osa1 version 3.0 release. The NSF Rice Oligonucleotide Array Project details the NSF45k arrays and the input gene set, the chosen oligos and their shared targets. Commercial arrays are also available. The GeneChip rice genome array, designed by Affymetrix and produced using a direct synthesis method, contains approximately 48,564 transcripts and 1,260 transcripts from the *japonica* and *indica* cultivars, respectively. Agilent has constructed a 22,000-element Agilent Rice Oligo Microarray Kit based on rice FL-cDNAs and recently announced a 44k version[34]. The NSF Rice Oligonucleotide Array Project provides a Rice Multi-Platform Microarray Search tool that allows users to search across the different rice oligo microarray platform types to determine which probes from each platform map to a particular gene target. Although this tool facilitates the analysis of array data, new tools are still needed because the extent of differential expression (for example, NSF45k, BGI and Agilent arrays) and gene-expression levels (for example, the Affymetrix array) differs between data sets and requires normalization before direct comparisons can be made. Furthermore, differences in sampling or growth stage of the samples can also affect comparisons of the array data.

Because many eukaryotic genome species contain a large number of alternatively spliced transcripts, some genes cannot be uniquely identified using any of the currently available array systems. A notable example is the rice gene *Thic* (*LOC_Os03g47610*), which encodes a putative thiamine biosynthesis protein with eight alternatively spliced transcripts. It is difficult to design unique oligos for each individual transcript no matter how far the design parameters are relaxed. Therefore, for some genes, microarray analysis cannot distinguish expression levels of alternatively spliced transcripts. Owing to this limitation, developers of rice microarray platforms use new methods to approach this issue. As shown in TABLE 2, the NSF45k array (NCBI Gene Expression Omnibus (GEO) platform accession numbers GPL4105 and GPL4106) includes 6,544 oligos that were computationally designed to match 15,003 multiple or alternatively spliced transcripts. The Affymetrix GeneChip includes 9,550 probe sets corresponding to

Table 2 | **Summary of available rice oligoarray platforms**

Platform	Number of probes (oligo length, nt)	TIGR V5 gene models*			Unique oligos or probe sets/gene models matched‖	Non-unique oligos or probe sets/gene models matched¶
		Non-TE‡	TE§	Total		
NSF45k	43,311 (50–70)	41,228	5,075	46,303	32,975/32,712	6,544/15,003
Affymetrix	610,665# (25)	43,794	3,759	47,533	34,535/29,334	9,550/19,660
Yale/BGI	60,727 (70)	35,438	5,311	40,749	25,227/22,195	8,320/19,815
Agilent 44k	40,901 (60)	Not analysed	Not analysed	36,021	24,535/18,574	12,544/17,447
Yale/Nimblegen tiling array[38]	12,254,374 (36)**	40,257‡‡	5,719‡‡	45,976‡‡	Not analysed	Not analysed

*Number of TIGR V5 gene models represented on each platform. ‡Number of non-TE-related gene models represented on each platform. §Number of TE-related gene models represented on each platform. ‖Number of oligos or probe sets that match a single TIGR V5 gene model and the numbers of TIGR V5 gene models that are targeted by unique oligos or probe sets (R. Buell, personal communication). ¶Number of oligos or probe sets that match more than one TIGR V5 gene model and the numbers of TIGR V5 gene model that are targeted by non-unique oligos or probe sets (R. Buell, personal communication). #There are 55,515 probe sets, each consisting of 11 probes. Mapping of Affymetrix probes required at least 7 of the 11 probes within a probe set to map to a model. **An average of 10 nucleotides separates adjacent probes. The probes tile both DNA strands of the non-repetitive sequences of the genome and were synthesized in a set of 32 arrays[38]. ‡‡In this case, the number of gene models was generated using TIGR V3 gene models[38]. BGI, Beijing Genomics Institute; NSF, National Science Foundation; nt, nucleotides; TE, transposable element; TIGR, The Institute for Genomic Research; V3, version 3; V5, version 5.

19,660 transcripts, the BGI array includes 8,320 oligos corresponding to 19,815 transcripts and the Agilent 44k array includes 12,544 oligos corresponding to 17,447 transcripts. The expression pattern of oligos targeting multiple transcripts can be compared with an oligo representing a single transcript of the same set, so that alternatively spliced transcripts at the same locus can be distinguished. Even though such an approach cannot discriminate all alternatively spliced transcripts, they do contribute to enhancing the interpretation of microarray data.

Tiling arrays and MPSS

Genome tiling arrays, a recent advance in microarray technology, are not based on predicted or known genes. Instead, the target genome is represented by oligonucleotide probes that 'tile' a continuous path along each chromosome[35,36]. A genome sequence can be covered by a manageable number of arrays, depending on the probe density. For example, in a recent set of experiments, only 32 tiling microarrays were needed to cover the non-repetitive sequence of the rice genome[37]. These arrays contain 13,078,888 individual 36-mer oligonucleotide probes spaced by 10 nucleotides[37].

Because hybridization of these tiling arrays with fluorescently labelled cDNA can reveal transcription of any genomic region, these arrays can be used to empirically validate predicted gene models and to identify novel transcription units[37,38]. For example, in addition to detecting transcription of 81.9% of the annotated gene models, the rice tiling array study also identified 15,472 transcribed intergenic regions, 9,023 antisense regions and 857 intronic regions[38]. Some TARs generate natural small interfering RNAs (siRNAs) derived from paired sense–antisense transcripts[38].

Another method that is used to validate predicted gene models and to identify novel genes is MPSS, which reveals short sequence signatures of cDNA libraries. This approach was used to develop a comprehensive expression atlas of rice sequences (Rice MPSS)[39]. This study revealed 46,971,553 mRNA transcripts from 22 libraries, and 2,953,855 small RNAs from 3 libraries. This approach also revealed widespread transcription throughout the genome, including sense expression of at least 25,500 annotated genes and antisense expression of nearly 9,000 annotated genes. An additional set of 15,000 mRNA signatures mapped to unannotated genomic regions. The majority of the small RNA data were derived from repetitive sequences and intergenic regions, and numerous clusters of highly regulated small RNAs were observed[40]. One drawback to MPSS is that the high cost limits the number of biological replicates that can be performed. Therefore, the quantitative significance of the resulting gene-expression profiles cannot yet be statistically validated.

Development of rice gene-indexed mutants

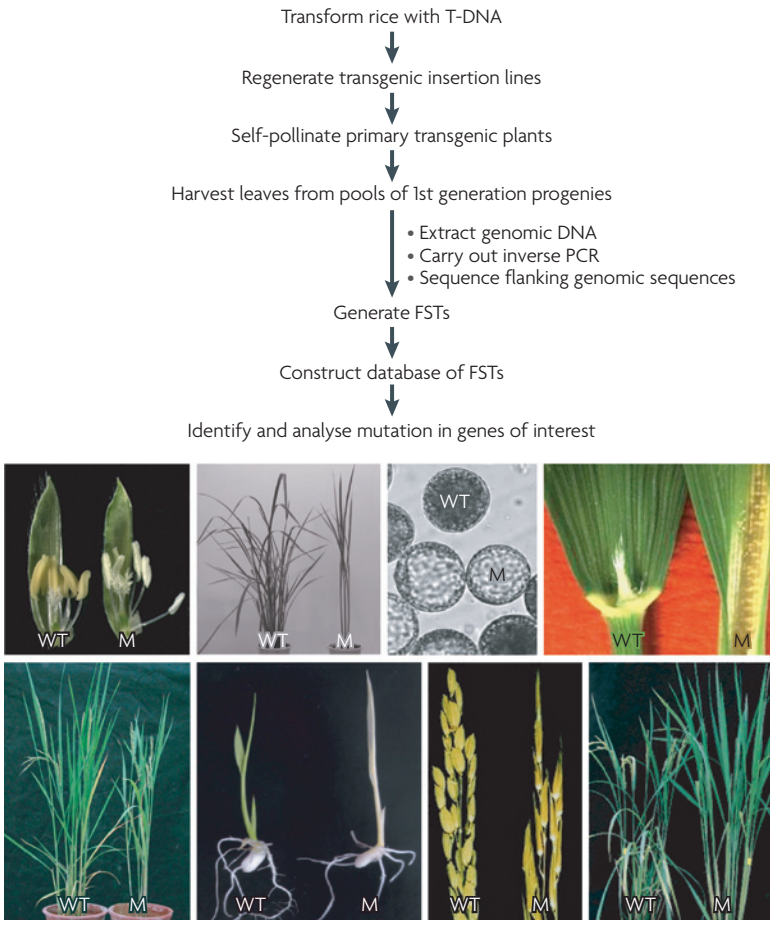
Several experimental approaches have been undertaken to develop rice lines in which genes are randomly tagged by DNA insertion elements[41–44]. Such mutant populations, which include gene knockout and gene overexpression lines, are useful for determining gene function based on phenotypes.

DNA elements that can insert randomly within chromosomes to disrupt gene function (to create loss-of-function mutants) include the T-DNA of *A. tumefaciens*, heterologous transposons (Ds and dSpm) and the Tos*17* retrotransposon[41,42,44–46] (BOX 1; TABLE 3). The basic strategy for the creation and screening of T-DNA insertional mutants and generation of flanking sequence tags (FSTs) is shown in BOX 1. So far, 172,500 FSTs have been generated. Using the FST database, we found that 27,551 (48%) of the 57,142 rice loci (given by the International Rice Genome Sequencing Project and the current TIGR rice genome pseudomolecules release) contain insertions in their genic regions or 5′ untranslated regions (as shown by the RiceGE: database sources, details and summary).

Another type of insertion population consists of lines that carry gain-of-function phenotypes. Such 'activation tagged' lines carry gene cassettes containing a strong enhancer element near one end, which can boost the expression of genes within a few kilobase pairs of chromosomal DNA. These tags can activate genes that are

Box 1 | **Creation and screening of T-DNA insertional mutants**

Insertional mutagenesis is a rapid method for mutating genes that can later be easily identified based on knowledge of the DNA tag. In this scheme (see figure), embryonic calli are co-cultivated with *Agrobacterium tumefaciens* carrying a T-DNA vector for random insertional mutagenesis (2–3 months). After selecting transgenic lines with selectable markers (for example, *hph* or *bar*), the lines are generated and transplanted to the greenhouse (2–3 months). Primary transgenic lines produce seeds by self-pollination (4–5 months). Because the sequence of the inserted element is known, the gene in which the insertion has occurred can be recovered, using various cloning (plasmid rescue) or PCR-based strategies (inverse PCR or thermal asymmetric interlaced (TAIL) PCR). Because 80–90% of primary transgenic lines have seeds, extraction of DNA from pooled leaves from first generation progeny can be used for the isolation of flanking genomic sequences of inserted T-DNA. The phenotype of the T-DNA insertional mutant allele can then be characterized in depth (which takes more than a year). All rice flanking sequence tags (FSTs) are publicly available at the Rice Functional Genomic Express Database (RiceGE) developed by the Salk Institute. Currently, 172,500 FSTs have been generated through the efforts of researchers around the world (11 institutes in 7 countries) who have created insertional mutations in 27,551 genes corresponding to 57,142 gene models (International Rice Genome Sequencing Project). For these calculations, insertions in 5′ UTR regions upstream from the ATG were included, whereas those in promoter and 3′ UTR regions were excluded. Knockout lines in genes of interest can be identified by carrying out a simple blast search using The Institute for Genomic Research (TIGR) locus name. The seeds of the mutant lines are provided by individual suppliers. Of these, the Plant Functional Laboratory T-DNA insertional mutant pool (POSTECH, South Korea) is the largest and has so far generated mutations in 20,460 genes and generated 82,520 flanking sequence tags (FSTs). Orders can be placed through the POSTECH rice T-DNA insertion sequence database. Tagged T-DNA flanking genomic sequence tags have been submitted to the RiceGE. WT, wild type; M, mutant.



located up to 10 kb away from the enhancer sequence independent of the direction of transcription, resulting in ectopic or increased expression of the targeted gene. The POSTECH Plant Functional Genomics Laboratory has produced more than 47,900 lines carrying tetramerized cauliflower mosaic virus 35S enhancer sequences and has generated 27,621 FSTs from these lines[42]. Over 90% of tested lines show activation of target genes at mature leaf stages[42].

The activation tagging approach can address the problem of gene redundancy that is associated with traditional screens for loss-of-function mutations. Another advantage of activation tagging is that it requires less effort to generate the lines as compared with overexpression analyses. Because multiple genes near the target gene can be simultaneously activated by the enhancer, phenotypes identified by this system need to be confirmed by conventional gain-of-function or other analyses.

Another gain-of-function approach called the 'FL-cDNA over-expresser (FOX) gene-hunting system' has recently been developed. So far, over 13,980 unique FL-cDNA clones that have been expressed in rice under the control of the maize ubiquitin 1 (MUB1) promoter[47]. However, one drawback of the FOX system is that ectopic expression of FL-cDNAs using the maize *MUB1* promoter can sometimes trigger unintended phenotypic changes. To overcome these problems, the FOX system can be modified for use with tissue-specific or inducible promoters.

The combined total of the five mutant populations listed in TABLE 3 is approximately 500,000 lines, with one to ten copies of mutations per line. If the insertions were evenly distributed across the genome, this FST collection would be predicted to target insertions in 99% of rice loci.

The insertion of DNA elements into coding regions often leads to complete loss of gene function. If the gene carries out an essential function then the mutation will be lethal, preventing subsequent phenotypic studies. To overcome this limitation, researchers from Fred Hutchinson Cancer Research Center developed a process for 'targeting induced local lesions in genomes', or TILLING. TILLING is useful as a supplementary tool to inactivate genes for which insertions are not available, or to obtain partial loss-of-function mutations so that an allelic series can be evaluated. In contrast to insertion mutagenesis, point mutations often lead to mild defects so that the function of essential genes can be evaluated[48]. In addition, TILLING produces non-transgenic stocks that can be used for field testing in parts of the world where transgenic field testing is restricted[52]. TILLING populations are generated by chemical mutagenesis (for example, ethyl methane sulphonate[49]), and mutants are screened using SNP detection assays, such as mismatch cleavage in heteroduplexes[50,51]. The method has been recently demonstrated in rice[52]. One drawback of TILLING is that, in order to detect the point mutations, PCR primers must be designed that can successfully amplify the target region.

Table 3 | **Summary of rice insertional mutant resources**

	Mutagen				
	T-DNA	**Ac/Ds, Spm/dSpm**	**T-DNA with enhancer**	**FOX system**	**Tos17**
Method*	*Agrobacterium*	*Agrobacterium*, crossing or selfing	*Agrobacterium*	*Agrobacterium*	Tissue culture
Copies per line	1–2	1–7	1–2	1	5–10
Distribution	Genic regions preferred	Genic regions highly preferred	Genic regions preferred	Not analysed	Genic regions preferred
Hot spots	No	Yes	No	No	Yes
FSTs (total 172,550)‡	141,254‡	13,309‡	46,083§	8,225‖	17,937‡
Target site of insertion	Genic	Genic	Genic or intergenic	Full-length cDNA	Genic except intron
Number of non-redundant insertions in genic regions (total 27,551)‡	~20,460‡	~2,433‡	Not analysed	5,462	3,719‡
Reference(s)	44, 46, 100, 101	41, 101, 102	44, 46, 96, 101, 103	47	44, 104

*Indicates the method for generating insertional mutants. ‡Indicates number of available flanking sequence tags. All data were generated from the rice functional genomic database at RiceGE. These data include the 46,083 FSTs from the POSTECH T-DNA enhancer lines, which was updated on August 23, 2007. §These lines were generated at the POSTECH Plant Functional Genomics Laboratory[42]. ‖Indicates the number of lines which identified the full-length cDNAs by PCR reaction. Ac/Ds, Spm/dSpm, heterologous transposons; FOX, full-length cDNA over-expresser; FST, flanking sequence tag; Tos17, a retrotransposon.

Deletion detection using tiling or microarrays

As indicated above, map-based cloning is labour intensive and time consuming[53] — especially for mutants with phenotypes that require intricate and lengthy analyses, such as submergence tolerance.

These factors have motivated several groups to explore the use of oligonucleotide tiling arrays to identify genomic deletions. If a genomic deletion is responsible for a particular mutant phenotype, then hybridization of genomic DNA from the mutant line to a set of whole-genome probes can be used to rapidly identify the probe or probes corresponding to a genomic deletion that is present in the mutant line. This method was used to successfully clone an *A. thaliana* sodium over-accumulation mutant gene carrying a 523-bp genomic deletion[54,55]. Cosegregation, complementation and comparative analyses among different salt-sensitive mutants were used to confirm that a deletion within the *A. thaliana* *HKT1* gene was responsible for the observed phenotypes of a fast-neutron irradiated mutant (sodium over-accumulation in shoots and leaf sodium sensitivity). A similar strategy was used to identify the *DMI3* gene from *Medicago truncatula* that encodes a calcium–calmodulin-dependent protein kinase[56,57]. These results demonstrate that oligonucleotide microarray-based cloning is an efficient and powerful tool to rapidly clone genetic deletion mutants that are otherwise difficult to phenotype for mapping, such as metabolic or cell-signalling mutants. Several groups are now using Affymetrix, the NSF45K or tiling arrays to identify genes that are missing from rice deletion mutants induced by fast-neutron mutagenesis[49,58].

Gene silencing

The presence of large gene families in rice, and the varying levels of functional redundancy associated with such families, creates a considerable challenge to the functional analysis of individual genes. For example, knockouts of a single gene within a gene family often produce little or no observable phenotype[59]. Newer technologies such as RNAi provide enhanced capability to study gene families. RNAi has been used to effectively knockdown multiple genes simultaneously with inverted-repeat constructs that target unique or conserved regions of multiple genes[60].

MicroRNAs (miRNAs), which are derived from irregular stem-loop structures, interact with target mRNAs by sequence complementarity, often reducing expression of the target gene[61]. Plant miRNAs direct mRNA cleavage of a single or small number of targets with near-perfect complementarity. Weigel and colleagues have recently designed an artificial miRNA (amiRNA) strategy to target one or more genes[61]. In this method, site-directed mutagenesis of endogenous miRNA precursors is performed to create amiRNAs. In *A. thaliana* experiments, the researchers found that all amiRNA-overexpressing plants targeting single genes caused phenotypic changes similar to those seen in plants with mutations in those genes. When multiple transcripts were targeted, the degree of downregulation of multiple targets varied. Therefore, silencing of multiple gene targets is possible, but for technical or biological reasons it requires follow-up studies to understand which subset of gene products have been depleted. Computational analysis predicts that at least 20 miRNA families are shared between *A. thaliana* and rice, and 13 additional miRNA families in rice have been identified[62]. Thus, amiRNAs will probably be a useful tool for functional genomics in rice. However, these technologies have practical limitations on the number of genes that can be simultaneously silenced and still require rational selection of gene targets[60].

Transient assay systems

To validate the relevance of highly prioritized candidate genes for a particular biological function, the rice

Allelic series
An allele is one or more alternative forms of a DNA sequence. To create an allelic series, molecular geneticists create mutations in a gene of interest and analyse the resulting phenotypes. Such allelic series are useful for determining gene function.

Cosegregation
The tendency for closely linked genes and genetic markers to segregate together.

protoplast transient assay system is a useful tool. It can be used to quickly examine subcellular localization, detect protein–protein interactions and measure promoter activity. Protoplast assays are relatively high-throughput, making it possible for researchers to reduce the number of candidate genes to be analysed using more laborious approaches such as transgenic or mutant analysis[63–66]. Furthermore, data derived from protoplasts are likely to be more biologically relevant than data obtained from heterologous cell systems such as bacteria and yeast[63].

Subcellular localization of target proteins can be determined by fusing the gene of interest with a marker such as GFP and then transiently expressing the fusion protein in protoplasts using electroporation or polyethylene glycol (PEG)-mediated transformation[65]. Transformation efficiencies of mesophyll protoplasts reach 60–70% — enough to obtain reliable and reproducible data[64]. Protoplasts can also be used to assay protein–protein interactions in living cells using the bimolecular fluorescence complementation (BiFC) system. In this system, the GFP (or variants such as yellow fluorescent protein (YFP)) is split into two halves. Neither fragment alone can fluoresce, but if each of the two non-fluorescent fragments is fused to two interacting partners, then fluorescence will be restored. The advantage of the BiFC method over other methods of visualizing protein–protein interactions is that it can confirm interactions *in vivo* as well as allowing visualization of the cellular localization of the complex. A protoplast-based BiFC system using complementation of two split YFP fragments was recently used to demonstrate homodimerization of SPIN1, an RNA-binding protein in rice cells[65].

Transcriptional activity of target genes can be measured by transient expression of promoter–reporter fusions. Genes can also be silenced in protoplasts and the downstream effects monitored. For example, we reported that transformation of protoplasts with an siRNA-targeting luciferase resulted in a significant level of silencing after only three hours, with an <83% decrease in expression[64]. In *A. thaliana*, conclusions derived from transient-expression assays have been validated by studies of transgenic or mutant lines, suggesting that rice protoplast studies will also be relevant to whole plant studies[63,67].

Mutant and transcriptome analysis

The use of the functional genomic tools discussed above is now generating large amounts of data, much of which has not been efficiently utilized and translated into biological insights. Despite the public availability of more than 300 array hybridizations, basic information linking physiological changes to well-defined transcriptional responses is lacking for even some of the most fundamental plant processes such as light responses and photosynthesis. Thus, there is a clear need to develop approaches that will move researchers beyond the initial stages of gene-expression analyses to efficiently and accurately assign function to the diverse components of rice signalling pathways.

Integrated and flexible informatic strategies that combine data from diverse sources are needed to enable systems-level analyses of functional genomic data. In this section, we discuss two strategies that can facilitate characterization of rice gene function and identify their role in rice signalling pathways.

Pathways can be reconstructed from microarray data that was derived from mutant lines carrying targeted mutations and/or time-course data[68,69]. Such experiments allow for the identification of cause–consequence relationships for genes with expression controlled by other genes. The results of such expression-profiling analyses can then be mapped onto known biological pathways to gain further insight into the pathway (for example, the GRAMENE pathway module).

One approach is to use a combination of whole-genome transcriptome, predicted pathway and insertional mutant analyses to identify and characterize candidate genes (FIG. 1). To test this strategy we examined 52 genes encoding proteins that are candidates for catalysing 8 key steps in the photorespiratory pathway. Gene family members were assigned to one of the 8 steps in the reconstructed pathway. We then used publicly available data comparing 7 light and dark treatments to identify the most highly expressed light-responsive gene family member for each of the 8 steps in the pathway (FIG. 1a–c). In this way, we were able to reduce the number of candidate genes in the pathway from 52 to 11, excluding 6 genes for which no data was available. Analysis of this data indicated that steps 4 and 8 are likely to be encoded by unique genes, whereas the other 6 steps are likely to be encoded by multi-gene family members. For these 6 steps, one gene family member showing the highest differential expression in step 1 (1-1), step 5 (5-1), step 6 (6-1) and step 7 (7-1) was predicted to be responsible for catalytic function. This prediction is supported by the fact that a Tos*17* insertional mutant (NC2658, NIAS Tos17 insertion mutant database) in rice gene 5-1 (*LOC_Os03g52840*) encoding a serine hydroxymethyltransferase conferred a variegated leaf phenotype (FIG. 1d). A similar variegated leaf phenotype was observed in mutants of the *A. thaliana* orthologue[70]. These experiments confirm a major role for gene 5-1 in the photorespiratory pathway.

Another approach to identify candidate genes that might have a function in a partially characterized biological pathway is shown in FIG. 2. This strategy takes advantage of the publicly available gene-expression data present in the NCBI GEO. These data, from over 300 hybridizations, include genes expressed during different developmental stages, in different tissues, in mutants and in response to biotic and abiotic stress [71–75]. In this case, we examined genes that were expressed in a male sterile mutant carrying an insertion in the gene *Undeveloped tapetum 1* (*Udt1*), which encodes a basic helix-loop-helix transcription factor[71]. Expression of *Udt1* is required for the differentiation of secondary parietal cells into mature tapetal cells and is expressed specifically in the anther during meiosis. First, we used whole-genome expression profiling to identify transcription factors that are specifically expressed in

Protoplast
A plant cell with the cell wall removed. Transient assays using protoplasts are effective for processing large quantities of genetic data coming out of high-throughput assays.

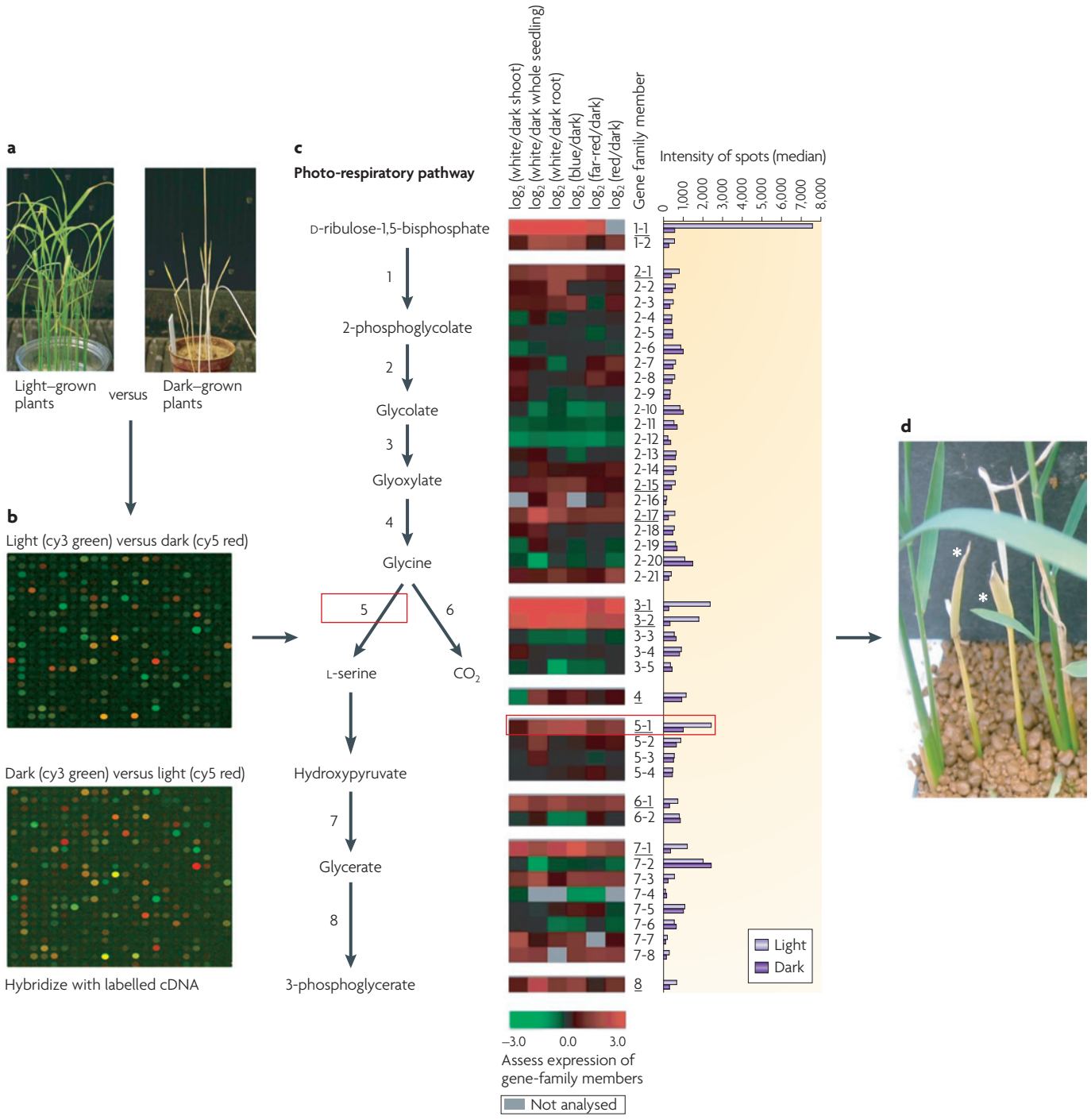


Figure 1 | **A combination of whole-genome transcriptome and predicted-pathway analysis can be used to efficiently determine gene function. a**, **b** | Whole-genome expression profiling was carried out to identify light-responsive genes[72]. **c** | Fifty two light-responsive gene-family members encoding enzymes catalysing one of the 8 steps in the photorespiration pathway as predicted by Ricecyc1.2 software in the GRAMENE database were identified. Microarray data for 46 of these genes were extracted from the Beijing Genomics Institute (BGI) data set. Gene-family members associated with 6 of the steps in the pathway are labelled as follows: 1-1, 1-2 for step 1; 2-1, 2-2 up to 2-21 for step 2; 3-1, 3-2 up to 3-5 for step 3; 5-1, 5-2, 5-3, 5-4 for step 5; 6-1, 6-2 for step 6; and 7-1, 7-2 up to 7-8 for step 7. On the bar chart, the average normalized level of expression of each gene in the six light treatments is indicated by a light blue bar; average gene expression in the dark condition is indicated by a purple bar. **d** | The function of highly expressed family members were examined using publicly available collections of insertion mutants, such as those depicted. These collections are available from the Rice Functional Genomic Express Database. A plant with a *Tos17* insertion (NC2658) in gene 5-1 (indicated by a red box in **c**) is marked with an asterisk. This analysis confirmed that gene 5-1 has a major role in the photorespiratory pathway, and demonstrates that a combination of transcriptome, pathway and phenotypic analyses is an efficient method to validate gene function.

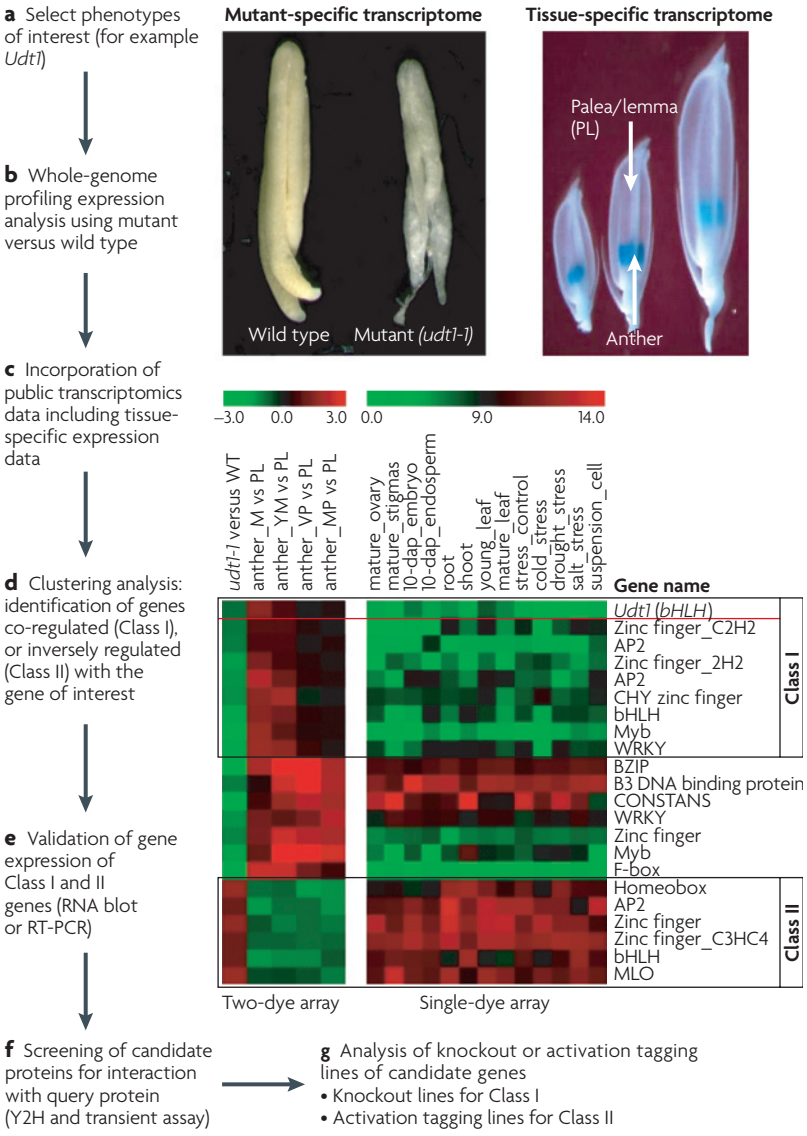


Figure 2 | **Elucidation of the function of candidate genes using publicly available transcriptomics data in combination with insertion-mutant collections. a** | Select the phenotype of interest. In this case it is a male sterile mutant carrying an insertion in the gene *Undeveloped tapetum 1* (*Udt1*), which encodes a basic helix-loop-helix transcription factor required for early tapetum development. **b** | Carry out whole-genome profiling analyses to identify genes expressed in the *udt1-1* mutant versus wild-type lines. **c** | Incorporate public transcriptomics data, including tissue-specific expression profiles to identify other genes that are likely to be involved in controlling tapetum development. For example, because *Udt1* is expressed specifically in the anther during early development[71], analysis of genes expressed in these anthers will help identify other genes with potential roles in controlling the development of tapetal cells. **d** | Cluster candidate genes based on expression profiles. Using the MultiExperiment Viewer 4.0 we identified genes that are co-regulated (Class I) or inversely regulated (Class II) with the gene of interest (*Udt1*, underlined in red). **e** | This prescreening approach allows biological studies to be carried out at greater depth on a few highly prioritized candidate genes, which can then be validated using RT-PCR. **f** | Physical interactions of the encoded proteins with UDT1 can be assessed using protein–protein interaction tools (yeast two-hybrid (Y2H) and split yellow fluorescent protein). **g** | The functions of highly prioritized Class I candidate genes can be determined by assessing anther-defective phenotypes of lines carrying insertion mutations and those of Class II candidate genes can be determined by assessing phenotypes of activation tagged lines. This strategy illustrates the benefits of combining computational comparisons of diverse publicly available microarray data sets with other genomic tools to validate the function of candidate genes. Photo containing Mutant *udt1-1* reproduced with permission from REF 71. © (2005) the American Society of Plant Biologists.

the *Udt1* wild type but not in the mutant line. We then assessed gene-expression profiles of wild-type anthers at meiosis and at the young microspore stage relative to the palea and/or lemma (tissues that make up the hulls of the rice spikelet) and other publicly available microarray data that assay developmental expression patterns. We then further classified these candidate genes using publicly available clustering software (TIGR multiexperiment viewer 4.0)[76] (FIG. 2). Clustering analysis makes it possible to arrange genes according to similarity in patterns of gene expression[76]. Because microarray analyses that rely on a single data set often identify genes governing biological variations that are not directly related to the mutant phenotype, this type of clustering analysis helps eliminate false positives and reduces the number of candidate genes. Co-expression of genes of known function (for example, *Udt1*) with poorly characterized or novel genes suggests that the genes can operate in the same pathway[76]. These analyses lead to the identification of genes that are co-regulated with *Udt1* (Class I), or inversely regulated with *Udt1* (Class II).

The function of Class I candidate genes can be assessed by examining anther phenotypes in lines carrying insertion mutations of these genes. Conversely, the function of Class II candidate genes can be assessed by examining the phenotypes of lines in which the candidate genes are activated (FIG. 2). This strategy illustrates the potential benefits of using computational comparisons of diverse, publicly available microarray data sets. This pre-screening approach allows biological studies to be carried out in greater depth on a few highly prioritized candidate genes, which can then be silenced in protoplasts or whole plants. These silenced lines can be assayed for biological phenotypes and for alterations of gene expression of other candidate genes. Such changes in gene expression would further support a role for these candidate genes in a shared biological pathway[64,65,77]. The physical relationship of the encoded proteins can be assayed using the yeast two-hybrid system[78,79].

Phylogenomics

In the absence of phenotypic information, functional information can be inferred from comparative genomic or systems biological studies that incorporate bioinformatic, genomic, gene expression and proteomic data. These approaches are hampered by current database formats that typically permit displays of only one gene or one field at a time, and are therefore not amenable to simultaneous comparisons of multiple data sets and multi-gene families. The 'scattered' nature of genomic data across multiple databases creates further challenges to data integration.

A new field of study that is, at least in part, resolving these limitations is phylogenomics, a field of study that puts genomic data in a phylogenetic context[80]. Phylogenetic trees provide a platform to sort and categorize genes into groups based on sequence similarity and are particularly valuable when studying large gene families. Consequently, phylogenetic trees provide a useful foundation for functional predictions

Box 2 | **Association mapping**

Another method to assign biological roles to genes of unknown function is to use a population genomics approach. Such studies, known as association mapping, can be used to study how allelic sequence variation among individuals results in phenotypic differences. Recently, an *Arabidopsis thaliana* Affymetrix genotyping array containing 250,000 SNPs was used successfully for genome-wide association mapping[84]. In rice, a similar study is underway to identify SNPs from across the whole genome of 20 rice varieties[85]. The study will focus on the genetic basis underlying important agricultural traits such as the nutritional value and disease resistance of these diverse rice varieties. Such rice varieties are a rich resource of diverse traits, and analysis of their genomic variations will provide valuable information regarding phenotypic variation between different rice strains.

based on limited phenotypic data. They also provide a context to identify members within gene families that have unique properties, such as the presence of novel domains, functional motifs or expression patterns. Unlike association mapping, phylogenomics does not rely on sequence information from numerous phenotypically characterized populations. Thus, phylogenomic analyses provide a rapid and logical basis for rational selection of gene candidates for further detailed functional studies[81]. Moreover, transcriptional analysis combined with phylogenetic analysis can further enhance the power of this approach to identify genes with relevance to a particular biological function[80].

An example of this type of analysis is the Rice Kinase Database (RKD), which was created to provide a logical format to analyse diverse sets of genomic information in a phylogenetic context[82]. The RKD displays user-selected genomic and functional genomic fields on a phylogenetic tree, with links to chromosomal and protein–protein interaction maps. Rather than analysing kinases one by one, the RKD allows simultaneous visualization of entire kinase groups, families and subfamilies. This format allowed us to identify features of rice receptor kinases that are specifically associated with pathogen recognition (the 'non-RD' motif)[81]. This database also allowed for the rational selection of kinases for use in a large-scale kinase proteomic screen[26]. The ability to integrate and analyse growing, functional genomic data sets in a logical and user-defined fashion will be essential to establishing a more global view of the role that kinases have in signalling.

Future directions

An intensive effort to analyse the function of the tens of thousands of predicted rice genes is now being undertaken. Integration of data from diverse transcriptomic, proteomic and computational approaches is needed to understand the function of these genes and of regulatory regions in non-coding RNAs and other newly or poorly annotated regions of the genome. The number of genes that can be characterized using current methods will be, at most, 10% of the genome based on estimates in *A. thaliana*[83].

Here, we have described the major limitations in current rice functional genomics approaches and suggested possible solutions. First, functional redundancy within gene families makes it difficult to determine function of an insertional mutant. This problem can often be overcome by generating double and/or triple mutants or by silencing multiple genes. Second, somaclonal variations frequently mask the true phenotypes; generating homozygous progenies is an effective tool to solve this problem. We have now produced homozygous progenies of around 1,000 genes (for example, E3 ligase, transcription factors and receptor-like kinase genes). International cooperation is required to expand the number of homozygous lines on a genomic scale. Such populations can also be used to generate double and/or triple mutants. Third, Nipponbare, Dongjin and Hwayoung varieties have been the primary germplasm sources used to generate random insertional mutants. Unfortunately, these rice varieties are difficult to grow and manage under greenhouse and growth-chamber conditions. The rice community therefore needs to generate such populations in rice varieties that can be more easily grown and subjected to genetic tests. Kitaake is such a variety. Not only is Kitaake efficiently transformed by *A. tumefaciens*-mediated T-DNA approaches, but Kitaake has a shorter life cycle (circa 9 weeks) than other varieties. We have generated several thousand T-DNA insertional mutants and overexpressed or silenced several hundred genes in Kitaake[26]. Fourth, accessibility to genetic resources such as rice insertional mutants is still restricted. International cooperation is therefore needed to create easier access to these materials. For example, phenotypic descriptions of all lines having FSTs will be useful for researchers attempting to associate gene function with gene-expression profiles, as shown in FIG. 1. Finally, there is a limitation in comparing gene-expression profiles among the four array platforms because each platform possesses different characteristics, such as distinct oligo identifiers and gene annotations. Although the NSF Rice Multi-Platform Microarray Search tool is a step towards comparing and using array data from multiple platforms, improvements are still needed. For example, if all gene-expression profiles were derived from a common, publicly available and affordable platform, the convenience of usage and consistency among data sets would be significantly enhanced. In addition, hundreds of experimental array data sets have not yet been published. Sharing these unpublished array data will contribute to accelerating rice functional genomics approaches. Population genomic approaches such as association mapping (BOX 2) can be used to study how allelic sequence variation among individuals results in phenotypic differences.

The long-term goal of the International Rice Functional Genomics Consortium is to determine the function of all the rice genes. Rice genes which show low sequence similarity with *A. thaliana* genes or which have different expression profiles with *A. thaliana* orthologous genes are of great interest to the research community. Worldwide collaboration will be necessary to pursue these goals. The availability of common resources will allow broader access to, and promote sharing of, the rice genetic information that is crucial to functional genomics research and the improvement of rice crops.

Somaclonal variation
Describes the genetic variation sometimes observed in plants that have passed through plant tissue culture. Chromosomal rearrangements are an important source of this variation.

1. Poehlman, J. M. *Genetics and plant breeding* (AVI Publishing Company, Westport, 1983).
2. IRGSP. The map-based sequence of the rice genome. *Nature* **436**, 793–800 (2005).
This paper reports that the map-based sequence of the whole rice genome provides more detailed features of the rice genome compared to previous draft sequences.
3. Hoshikawa, K. in *Science of the Rice Plant*. (eds Matsuo, T. & Hoshikawa, K.) 91–132 (Food and Agriculture Policy Research Center, 1993).
4. Paterson, A. H., Bowers, J. E., Peterson, D. G., Estill, J. C. & Chapman, B. A. Structure and evolution of cereal genomes. *Curr. Opin. Genet. Dev.* **13**, 644–650 (2003).
5. Devos, K. M. & Gale, M. D. Genome relationships: the grass model in current research. *Plant Cell* **12**, 637–646 (2000).
6. Hiei, Y., Komari, T. & Kubo, T. Transformation of rice mediated by *Agrobacterium tumefaciens*. *Plant Mol. Biol.* **35**, 205–218 (1997).
This is the first report detailing the use of the *A. tumefaciens*-mediated T-DNA transformation method in rice.
7. Gale, M. D. & Devos, K. M. Comparative genetics in the grasses. *Proc. Natl Acad. Sci. USA* **95**, 1971–1974 (1998).
8. Goff, S. A. Rice as a model for cereal genomics. *Curr. Opin. Plant Biol.* **2**, 86–89 (1999).
9. Hiei, Y., Ohta, S., Komari, T. & Kumashiro, T. Efficient transformation of rice (*Oryza sativa*, L.) mediated by *Agrobacterium* and sequence analysis of the boundaries of the T-DNA. *Plant J.* **6**, 271–282 (1994).
10. Shimamoto, K. & Kyozuka, J. Rice as a model for comparative genomics of plants. *Annu. Rev. Plant Biol.* **53**, 399–419 (2002).
11. Kellogg, E. A. Evolutionary history of the grasses. *Plant Physiol.* **125**, 1198–1205 (2001).
12. Londo, J. P., Chiang, Y. C., Hung., K. H., Chiang, T. Y. & Schaal, B. A. Phylogeography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice, *Oryza sativa*. *Proc. Natl Acad. Sci. USA* **103**, 9578–9583 (2006).
13. Khush, G. S. Origin, dispersal, cultivation and variation of rice. *Plant Mol. Biol.* **35**, 25–34 (1997).
14. Ouyang, S. *et al.* The TIGR Rice Genome Annotation Resource: improvements and new features. *Nucleic Acids Res.* **35**, D883–887 (2007).
15. Sasaki, T. *et al.* The genome sequence and structure of rice chromosome 1. *Nature* **420**, 312–316 (2002).
16. Yuan, Q. *et al.* The Institute for Genomic Research Osa1 rice genome annotation database. *Plant Physiol.* **138**, 18–26 (2005).
This paper describes a rice genome annotation database (Osa1), which provides structural and functional annotation using *O. sativa* ssp. *japonica* cv. Nipponbare from the International Rice Genome Sequencing Project.
17. Itoh, T. *et al.* Curated genome annotation of *Oryza sativa* ssp. *japonica* and comparative genome analysis with *Arabidopsis thaliana*. *Genome Res.* **17**, 175–183 (2007).
This paper describes a set of 32,127 FL-cDNA clones corresponding to approximately 21,000 transcription units of the *japonica* rice genome that are available from the Knowledge-based *Oryza* Molecular Biological Encyclopedia.
18. Ohyanagi, H. *et al.* The Rice Annotation Project Database (RAP-DB): hub for *Oryza sativa* ssp. *japonica* genome information. *Nucleic Acids Res.* **34**, D741–D744 (2006).
19. The map-based sequence of the rice genome. *Nature* **436**, 793–800 (2005).
20. Kikuchi, S. *et al.* Collection, mapping, and annotation of over 28,000 cDNA clones from *japonica* rice. *Science* **301**, 376–379 (2003).
21. Juretic, N., Bureau, T. E. & Bruskiewich, R. M. Transposable element annotation of the rice genome. *Bioinformatics* **20**, 155–160 (2004).
22. Lukowitz, W., Gillmor, C. S. & Scheible, W. R. Positional cloning in *Arabidopsis*. Why it feels good to have a genome initiative working for you. *Plant Physiol.* **123**, 795–805 (2000).
23. Yu, J. *et al.* The genomes of *Oryza sativa*: a history of duplications. *PLoS Biol.* **3**, e38 (2005).
24. Shiu, S. H. *et al.* Comparative analysis of the receptor-like kinase family in *Arabidopsis* and rice. *Plant Cell* **16**, 1220–1234 (2004).
25. Tian, C., Wan, P., Sun, S., Li, J. & Chen, M. Genome-wide analysis of the GRAS gene family in rice and *Arabidopsis*. *Plant Mol. Biol.* **54**, 519–532 (2004).
26. Rohila, J. S. *et al.* Protein–protein interactions of tandem affinity purification-tagged protein kinases in rice. *Plant J.* **46**, 1–13 (2006).
27. Jiang, J., Birchler, J. A., Parrott, W. A. & Dawe, R. K. A molecular view of plant centromeres. *Trends Plant Sci.* **8**, 570–575 (2003).
28. Rensink, W. A. & Buell, C. R. Microarray expression profiling resources for plant genomics. *Trends Plant Sci.* **10**, 603–609 (2005).
This Review focuses on recent advances in the application of microarrays in plant genomic research and in gene-expression databases available for plants.
29. Alonso, J. M. & Ecker, J. R. Moving forward in reverse: genetic technologies to enable genome-wide phenomic screens in *Arabidopsis*. *Nature Rev. Genet.* **7**, 524–536 (2006).
30. Wing, R. A. *et al.* The *Oryza* map alignment project: the golden path to unlocking the genetic potential of wild rice species. *Plant Mol. Biol.* **59**, 53–62 (2005).
31. Lipshutz, R. J., Fodor, S. P., Gingeras, T. R. & Lockhart, D. J. High density synthetic oligonucleotide arrays. *Nature Genet.* **21**, 20–24 (1999).
32. Ramsay, G. DNA chips: state-of-the-art. *Nature Biotechnol.* **16**, 40–44 (1998).
33. Borevitz, J. O. *et al.* Genome-wide patterns of single-feature polymorphism in *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA* **104**, 12057–12062 (2007).
34. Shimono, M. *et al.* Rice WRKY45 plays a crucial role in benzothiadiazole-inducible blast resistance. *Plant Cell* **19**, 2064–2076 (2007).
35. Mockler, T. C. *et al.* Applications of DNA tiling arrays for whole-genome analysis. *Genomics* **85**, 1–15 (2005).
36. Johnson, J. M., Edwards, S., Shoemaker, D. & Schadt, E. E. Dark matter in the genome: evidence of widespread transcription detected by microarray tiling experiments. *Trends Genet.* **21**, 93–102 (2005).
37. Li, L. *et al.* Genome-wide transcription analyses in rice using tiling microarrays. *Nature Genet.* **38**, 124–129 (2006).
This paper describes a full-genome transcription analysis of the *indica* rice subspecies using high-density oligonucleotide tiling microarrays.
38. Li, L. *et al.* Global identification and characterization of transcriptionally active regions in the rice genome. *PLoS ONE* **2**, e294 (2007).
This paper describes TILLING of the rice genome and newly identified transcription units.
39. Nakano, M. *et al.* Plant MPSS databases: signature-based transcriptional resources for analyses of mRNA and small RNA. *Nucleic Acids Res.* **34**, D731–D735 (2006).
40. Nobuta, K. *et al.* An expression atlas of rice mRNAs and small RNAs. *Nature Biotechnol.* **25**, 473–477 (2007).
This paper describes the first deep sequence data for small RNAs in a crop plant.
41. Kumar, C. S., Wing, R. A. & Sundaresan, V. Efficient insertional mutagenesis in rice using the maize En/Spm elements. *Plant J.* **44**, 879–892 (2005).
42. Jeong, D. H. *et al.* Generation of a flanking sequence-tag database for activation-tagging lines in *japonica* rice. *Plant J.* **45**, 123–132 (2006).
This paper reports the generation of 47,932 T-DNA tag lines in *japonica* rice using activation-tagging vectors that contain tetramerized 35S enhancer sequences.
43. Miyao, A. *et al.* Target site specificity of the Tos*17* retrotransposon shows a preference for insertion within genes and against insertion in retrotransposon-rich regions of the genome. *Plant Cell* **15**, 1771–1780 (2003).
44. An, G., Lee, S., Kim, S. H. & Kim, S. R. Molecular genetics using T-DNA in rice. *Plant Cell Physiol.* **46**, 14–22 (2005).
45. Sallaud, C. *et al.* High throughput T-DNA insertion mutagenesis in rice: a first step towards *in silico* reverse genetics. *Plant J.* **39**, 450–464 (2004).
46. Hsing, Y. I. *et al.* A rice gene activation/knockout mutant resource for high throughput functional genomics. *Plant Mol. Biol.* **63**, 351–364 (2007).
47. Nakamura, H. *et al.* A genome-wide gain-of-function analysis of rice genes using the FOX-hunting system. *Plant Mol. Biol.* (2007).
48. Wu, Z. *et al.* A chlorophyll-deficient rice mutant with impaired chlorophyllide esterification in chlorophyll biosynthesis. *Plant Physiol.* **145**, 29–40 (2007).
49. Wu, J. L. *et al.* Chemical- and irradiation-induced mutants of *indica* rice IR64 for forward and reverse genetics. *Plant Mol. Biol.* **59**, 85–97 (2005).
50. McCallum, C. M., Comai, L., Greene, E. A. & Henikoff, S. Targeted screening for induced mutations. *Nature Biotechnol.* **18**, 455–457 (2000).
51. Comai, L. & Henikoff, S. TILLING: practical single-nucleotide mutation discovery. *Plant J.* **45**, 684–694 (2006).
This paper describes TILLING, which provides targeted inactivation of rice genes identified by sequence analysis.
52. Till, B. J. *et al.* Discovery of chemically induced mutations in rice by TILLING. *BMC Plant Biol.* **7**, 19 (2007).
53. Xu, K. *et al.* Sub1A is an ethylene-response-factor-like gene that confers submergence tolerance to rice. *Nature* **442**, 705–708 (2006).
54. Gong, J. M. *et al.* Microarray-based rapid cloning of an ion accumulation deletion mutant in *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA* **101**, 15404–15409 (2004).
55. Wang, S., Sim, T. B., Kim, Y. S. & Chang, Y. T. Tools for target identification and validation. *Curr. Opin. Chem. Biol.* **8**, 371–377 (2004).
56. Levy, J. *et al.* A putative Ca^{2+} and calmodulin-dependent protein kinase required for bacterial and fungal symbioses. *Science* **303**, 1361–1364 (2004).
57. Mitra, R. M. *et al.* A Ca^{2+}/calmodulin-dependent protein kinase required for symbiotic nodule development: gene identification by transcript-based cloning. *Proc. Natl Acad. Sci. USA* **101**, 4701–4705 (2004).
58. Wang, G. L. *et al.* Isolation and characterization of rice mutants compromised in Xa21-mediated resistance to *X. oryzae* pv. *oryzae*. *Theor. Appl. Genet.* **108**, 379–384 (2004).
59. Eckardt, N. A. Good things come in threes: a trio of triple kinases essential for cell division in *Arabidopsis*. *Plant Cell* **14**, 965–967 (2002).
This paper shows that RNA silencing is a useful method for the functional analysis of gene families in rice. Each of the seven members of the *OsRac* gene family was specifically suppressed by its respective inverted-repeat construct. In addition, the expression of all members of the gene family was suppressed with variable efficiencies.
60. Miki, D., Itoh, R. & Shimamoto, K. RNA silencing of single and multiple members in a gene family of rice. *Plant Physiol.* **138**, 1903–1913 (2005).
61. Schwab, R., Ossowski, S., Riester, M., Warthmann, N. & Weigel, D. Highly specific gene silencing by artificial microRNAs in *Arabidopsis*. *Plant Cell* **18**, 1121–1133 (2006).
62. Sunkar, R., Girke, T., Jain, P. K. & Zhu, J. K. Cloning and characterization of microRNAs from rice. *Plant Cell* **17**, 1397–1411 (2005).
63. Sheen, J. Signal transduction in maize and *Arabidopsis* mesophyll protoplasts. *Plant Physiol.* **127**, 1466–1475 (2001).
64. Bart, R., Chern, M., Park, C. J., Bartley, L. & Ronald, P. C. A novel system for gene silencing using siRNAs in rice leaf and stem-derived protoplasts. *Plant Methods* **2**, 13 (2006).
This paper describes a system for isolation, transformation and gene silencing of etiolated rice leaf and stem-derived protoplasts.
65. Chen, S. *et al.* A highly efficient transient protoplast system for analyzing defence gene expression and protein–protein interactions in rice. *Mol. Plant Pathol.* **7**, 417–427 (2006).
66. Kawasaki, T. *et al.* The small GTP-binding protein rac is a regulator of cell death in plants. *Proc. Natl Acad. Sci. USA* **96**, 10922–10926 (1999).
67. Isshiki, M., Tsumoto, A. & Shimamoto, K. The serine/arginine-rich protein family in rice plays important roles in constitutive and alternative splicing of pre-mRNA. *Plant Cell* **18**, 146–158 (2006).
68. AbuQamar, S. *et al.* Expression profiling and mutant analysis reveals complex regulatory networks involved in *Arabidopsis* response to *Botrytis* infection. *Plant J.* **48**, 28–44 (2006).
69. Brazhnik, P., de la Fuente, A. & Mendes, P. Gene networks: how to put the function in genomics. *Trends Biotechnol.* **20**, 467–472 (2002).
70. Voll, L. M. *et al.* The photorespiratory *Arabidopsis SHM1* mutant is deficient in SHM1. *Plant Physiol.* **140**, 59–66 (2006).
71. Jung, K. H. *et al.* Rice Undeveloped tapetum1 is a major regulator of early tapetum development. *Plant Cell* **17**, 2705–2722 (2005).

72. Jiao, Y., Ma, L., Strickland, E. & Deng, X. W. Conservation and divergence of light-regulated genome expression patterns during seedling development in rice and *Arabidopsis*. *Plant Cell* **17**, 3239–3256 (2005).
73. Su, N. *et al.* Distinct reorganization of the genome transcription associates with organogenesis of somatic embryo, shoots and roots in rice. *Plant Mol. Biol.* **63**, 337–349 (2007).
74. Li, M., Xu, W., Yang, W., Kong, Z. & Xue, Y. Genome-wide gene expression profiling reveals conserved and novel molecular functions of the stigma in rice (*Oryza sativa*, L.). *Plant Physiol.* **144**,1797–1812 (2007).
75. Walia, H. *et al.* Comparative transcriptional profiling of two contrasting rice genotypes under salinity stress during the vegetative growth stage. *Plant Physiol.* **139**, 822–835 (2005).
76. Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci. USA* **95**, 14863–14868 (1998).
77. Takai, R., Kaneda, T., Isogai, A., Takayama, S. & Che, F. S. A new method of defense response analysis using a transient expression system in rice protoplasts. *Biosci. Biotechnol. Biochem.* **71**, 590–593 (2007).
78. Causier, B. & Davies, B. Analysing protein–protein interactions with the yeast two-hybrid system. *Plant Mol. Biol.* **50**, 855–870 (2002).
79. Wang, Y. S. *et al.* Rice XA21 binding protein 3 is a ubiquitin ligase required for full Xa21-mediated disease resistance. *Plant Cell* **18**, 3635–3646 (2006).
80. Jiao, Y. & Deng, X. W. A genome-wide transcriptional activity survey of rice transposable element-related genes. *Genome Biol.* **8**, R28 (2007).
81. Dardick, C. & Ronald, P. Plant and animal pathogen recognition receptors signal through non-RD kinases. *PLoS Pathog.* **2**, e2 (2006).
This paper describes a phylogenomic database, the RKD, to facilitate functional analysis of the rice protein kinase gene family.
82. Dardick, C., Chen, J., Richter, T., Ouyang, S. & Ronald, P. The rice kinase database. A phylogenomic database for the rice kinome. *Plant Physiol.* **143**, 579–586 (2007).
83. Meinke, D. W. *et al.* A sequence-based map of *Arabidopsis* genes with mutant phenotypes. *Plant Physiol.* **131**, 409–418 (2003).
84. Kim, S. *et al.* Recombination and linkage disequilibrium in *Arabidopsis thaliana*. *Nature Genet.* **39**, 1151–1155 (2007).
85. McNally, K. L. *et al.* Sequencing multiple and diverse rice varieties. Connecting whole-genome variation with phenotypes. *Plant Physiol.* **141**, 26–31 (2006).
86. Song, W. Y. *et al.* A receptor kinase-like protein encoded by the rice disease resistance gene, Xa21. *Science* **270**, 1804–1806 (1995).
87. Li, X. *et al.* Control of tillering in rice. *Nature* **422**, 618–621 (2003).
88. Qu, S. *et al.* The broad-spectrum blast resistance gene *Pi9* encodes a nucleotide-binding site-leucine-rich repeat protein and is a member of a multigene family in rice. *Genetics* **172**, 1901–1914 (2006).
89. Zhou, B. *et al.* The eight amino-acid differences within three leucine-rich repeats between Pi2 and Piz-t resistance proteins determine the resistance specificity to *Magnaporthe grisea*. *Mol. Plant Microbe Interact.* **19**, 1216–1228 (2006).
90. Ueguchi-Tanaka, M. *et al.* Gibberellin insensitive dwarf1 encodes a soluble receptor for gibberellin. *Nature* **437**, 693–698 (2005).
91. Ikeda, A. *et al.* Slender rice, a constitutive gibberellin response mutant, is caused by a null mutation of the SLR1 gene, an ortholog of the height-regulating gene GAI/RGA/RHT/D8. *Plant Cell* **13**, 999–1010 (2001).
92. Sasaki, A. *et al.* Accumulation of phosphorylated repressor for gibberellin signaling in an F-box mutant. *Science* **299**, 1896–1898 (2003).
93. Sasaki, A. *et al.* Green revolution: a mutant gibberellin-synthesis gene in rice. *Nature* **416**, 701–702 (2002).
94. Ma, J. F. *et al.* A silicon transporter in rice. *Nature* **440**, 688–691 (2006).
95. Konishi, S. *et al.* An SNP caused loss of seed shattering during rice domestication. *Science* **312**, 1392–1396 (2006).
96. Mori, M. *et al.* Isolation and molecular characterization of a Spotted leaf 18 mutant by modified activation-tagging in rice. *Plant Mol. Biol.* **63**, 847–860 (2007).
97. Moon, S. *et al.* The rice *FON1* gene controls vegetative and reproductive development by regulating shoot apical meristem size. *Mol. Cells* **21**, 147–152 (2006).
98. Jeon, J. S. *et al.* Leafy hull sterile1 is a homeotic mutation in a rice MADS box gene affecting rice flower development. *Plant Cell* **12**, 871–884 (2000).
99. Chern, M., Fitzgerald, H. A., Canlas, P. E., Navarre, D. A. & Ronald, P. C. Overexpression of a rice NPR1 homolog leads to constitutive activation of defense response and hypersensitivity to light. *Mol. Plant Microbe Interact.* **18**, 511–520 (2005).
100. Jeon, J. S. *et al.* T-DNA insertional mutagenesis for functional genomics in rice. *Plant J.* **22**, 561–570 (2000).
101. An, G., Jeong, D. H., Jung, K. H. & Lee, S. Reverse genetic approaches for functional genomics of rice. *Plant Mol. Biol.* **59**, 111–123 (2005).
102. van Enckevort, L. J. *et al.* EU-OSTID: a collection of transposon insertional mutants for functional genomics in rice. *Plant Mol. Biol.* **59**, 99–110 (2005).
103. Jeong, D. H. *et al.* Generation of a flanking sequence-tag database for activation-tagging lines in *japonica* rice. *Plant J.* **45**, 123–132 (2006).
104. Miyao, A. *et al.* A large-scale collection of phenotypic data describing an insertional mutant population to facilitate functional analysis of rice genes. *Plant Mol. Biol.* **63**, 625–635 (2007).
This paper reports the phenotypes of 50,000 Tos*17* insertion lines in the M_2 generation which were observed in the field.

Acknowledgements

We thank B. C. Meyers, L. Bartley, C. Dardick, L. Comai, D. Neale, J. Schroeder, J. Leach, G. L. Wang, K. Shimamoto, V. Sundaresan and R. C. Buell for comments and discussions. We also thank S. Ouyang, Y. S. Lee and P. Cao for helping to generate tables and figures. This work was supported by National Institutes of Health grants 5R01GM055962-0 United States Department of Agriculture grant 2004-63560416640 and National Science Foundation grants DBI-0313887 to P. R., the 21st Century Frontier Program CG1111 and Biogreen 21 Program to G. A, Korea Research Foundation grant 2005-C00155 to K. H. J.

DATABASES

Entrez Gene: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene
HKT1
Entrez Genome Project: http://www.ncbi.nlm.nih.gov/sites/entrez?db=genomeprj
Arabidopsis thaliana | *Oryza sativa*
GRAMENE: http://www.gramene.org
LOC_Os03g47610 | *LOC_Os03g52840* | *Udt1*
NCBI gene expression omnibus:
http://www.ncbi.nlm.nih.gov/geo
GPL4105 | GPL4106

FURTHER INFORMATION

Pamela Ronald's homepage: http://indica.ucdavis.edu
Agilent Rice Oligo Microarray Kit:
http://www.chem.agilent.com/scripts/pds.asp?lPage=12133
CSIRO Ac/Ds in Australia: http://www.pi.csiro.au/fgrttpub
Current TIGR rice genome pseudomolecules release:
http://www.tigr.org/tdb/e2k1/osa1/pseudomolecules/info.shtml#feat
DNA Data Bank of Japan: http://www.ddbj.nig.ac.jp
European Molecular Biology Laboratory:
http://www.embl.org
Genoplante *Oryza* tag lines:
http://urgi.versailles.inra.fr/OryzaTagLine
GRAMENE pathway module:
http://www.gramene.org/pathway
International Rice Functional Genomics Consortium:
http://irfgc.irri.org
International Rice Genome Sequencing Project:
http://rgp.dna.affrc.go.jp/IRGSP/
Knowledge-based *Oryza* Molecular Biological Encyclopedia:
http://cdna01.dna.affrc.go.jp/cDNA
***Magnaporthe grisea–Oryza sativa* Interaction Database:**
http://www.mgosdb.org
NCBI Gene Expression Omnibus:
http://www.ncbi.nlm.nih.gov/geo
NCBI map viewer: http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi?taxid=4530
NIAS Tos*17* insertion mutant database:
http://tos.nias.affrc.go.jp
NSF Rice Oligonucleotide Array Project:
http://www.ricearray.org
***Oryza* Map Alignment Project:**
http://www.omap.org/index.html
OryGenesDB, France: http://orygenesdb.cirad.fr
POSTECH Laboratory:
http://www.postech.ac.kr/life/pfg
POSTECH rice T-DNA insertion sequence database:
http://an6.postech.ac.kr/pfg/index.php
Rice Annotation Project database:
http://rapdb.dna.affrc.go.jp
Rice Functional Genomic Express Database:
http://signal.salk.edu/cgi-bin/RiceGE
RiceGE: database sources, details and summary:
http://signal.salk.edu/RiceGE/RiceGE_Data_Source.html
Rice Kinase Database: http://rkd.ucdavis.edu
Rice Multi-Platform Microarray Search tool:
http://www.ricearray.org/matrix.search.shtml
Rice Mutant Database, Huazhong Agricultural University, China: http://rmd.ncpgr.cn/introduction.cgi?nickname=
Rice MPSS: http://mpss.udel.edu/rice
Rice Tilling Database:
http://tilling.ucdavis.edu/index.php/Main_Page
Shanghai T-DNA Insertion Population:
http://ship.plantsignal.cn/index.do
Taiwan Rice Insertional Mutants Database:
http://trim.sinica.edu.tw
TIGR multiexperiment viewer 4.0:
http://www.tm4.org/mev.html
TIGR Rice Genome Annotation:
http://www.tigr.org/tdb/e2k1/osa1
University of California, Davis Rice Functional Genomics Databases: http://www-plb.ucdavis.edu/Labs/sundar/Rice_Genomics.htm
Zhejiang University, China rice T-DNA tags:
http://www.genomics.zju.edu.cn/ricetdna.html

ALL LINKS ARE ACTIVE IN THE ONLINE PDF