J. Plant Biol. (2014) 57:80-92
DOI 10.1007/s12374-014-0903-6

REVIEW ARTICLE

Resources for Systems Biology in Rice

Anil Kumar Nalini Chandran[1] and Ki-Hong Jung[1,2]*

[1]*Department of Plant Molecular Systems Biotechnology & Crop Biotech Institute, Kyung Hee University, Yongin 446-701, Korea*
[2]*Graduate School of Biotechnology, Kyung Hee University, Yongin 446-701, Korea*

Received: January 15, 2014 / Accepted: February 20, 2014

Abstract Systems biology is an upcoming trend in the field of functional genomics. Recently, there has been a significant improvement in the resources for systems biology in *Oryza sativa* (rice), a model crop. These resources include whole-genome sequencing/re-sequencing data, transcriptomes, protein-protein interactomes, reactomes, functional gene network tools, and gene indexed mutant populations. The integration of diverse omics data can lead to greater understanding of the functional genomics of rice. In this review, we address the development and current progress of the resources available for systems biology in rice: Genome browsers and databases for the orthology identification, transcriptome analysis, protein-protein interaction network and functional gene network analyses, a co-expression network, metabolic pathway analysis for promoter analysis, and gene indexed mutants.

Keywords: Genome browsers, Metabolic pathway analysis, Protein-protein interaction analysis, Rice, Systems biology, Transcriptome analysis

Introduction

Rice is one of the most significant crops, representing a staple food for humans. Functional identification of rice on a whole genome scale will be required to significantly improve the quality of rice, rice yield, and stress tolerance in response to changing climate (Jung et al. 2008). As of now, the functions of 756 genes in rice have been elucidated through genetic analyses (Yamamoto et al. 2012). These genes cover just 2% of non-transposable element genes in rice annotated by the rice genome annotation project (RGAP, http://rice.plantbiology.msu.edu/) (Kawahara et al. 2013). There might be two reasons to explain why most rice gene functions have not been elucidated: one reason is functional redundancy; and the other is the limited information on the identification of gene function. Recently, large collections of microarray data have been produced using several whole genome array platforms in rice (Jung et al. 2008). The meta-expression analysis is very useful to obtain the functional implication of the remaining genes. Web tools for transcriptome analysis in rice have been established and will facilitate the functional identification of gene(s) of interest based on expression patterns (Jung et al. 2011).

Protein-protein interaction information provides clues about the direct regulatory elements of the target protein, while whole genome-transcriptome data provide a larger number of putative downstream elements. Although it depends on the experimental scheme, it is generally very difficult to identify immediate downstream candidate genes for a gene of interest. Approaches adopted to compensate for the limitations of transcriptome analyses include protein-protein interaction analysis based on yeast two hybrid (Y2H) and Tandem Affinity Purification (TAP) immunoprecipitation systems, which provide the candidate proteins for the direct regulation of the target protein. In rice, interacting proteins for 116 kinases have been identified by Y2H or TAP immunoprecipitation systems (Ding et al. 2009).

Studying metabolic pathways is an effective way to understand the molecular function of a target gene. Integration of data from high-throughput analysis on the framework of global metabolic pathways resulted in the identification of a specific pathway or pathways associated with applied biological problems. Many reviews have focused on introducing specific areas of functional genomics such as genomics, proteomics, and transcriptomics. Therefore, unified information which bridges all of the rice-related resources is necessary.

Here we present such a source dealing with popular tools and databases for systems biology in rice. A description of the tools will cover their functionalities, search modes, and the backend dataset. Since rice genome sequencing has been completed, a number of tools and databases for functional

*Corresponding author; Ki-Hong Jung
Tel : +82-31-201-3474
E-mail : khjung2010@khu.ac.kr

genomics have been developed. However, some of them are not very functional due to outdated information and technical problems. Hence, we focus on sources which are regularly updated or frequently accessed by related scientific communities. A practical application or case study for each tool is beyond the focus of this review and will not be discussed.

Genome Browsers

The accumulation of diverse data from different experiments poses a challenge when retrieving the data as a reference for further studies. In rice research, after the completion of the genome sequence had reached the level of 'gold standard', the next major challenge was organizing the available information and data from subsequent experiments in a systematic manner. A unified platform for retrieving all primary and secondary data is time saving. Advances in the area of bioinformatics resulted in the development of species specific or general databases with functional capabilities including genome browsing and cross referencing data with multiple plant species, as well as other functional capabilities. For this purpose, genome browsers such as the Rice Genome Annotation Project (MSU-RGAP), Rice Annotation Project Database (RAP-DB), Rice Functional Genomic Express Database (RiceGE), and Gramene were developed. In addition to various candidates from plant kingdom, Gramene database (Youens-Clark et al. 2011) provides genomic information and annotation data for 15 *Oryza* species including ssp *indica* and *japonica*. A notable point in data organization among these databases is the extension of the dataset to other organisms, while maintaining the central focus on the rice genome. Gramene is a standard representative of this trend and reflects the recent advancements in comparative genomics studies.

Even though the purpose of the various genome browsers for rice research is the same, the nature of the dataset varies for each database. RGAP uses an updated version (Nipponbare release 7) of fundamental rice genome data, which includes information on gene loci, gene models, repetitive sequences, and exon and intron positions in (16,941) transposable elements (TE) and (39,045) non-TE regions (Kawahara et al. 2013). Apart from this primary level of information, various analysis tools for functional studies are also enabled, such as GO retrieval, BLAST search, tRNA search, domain and motif search, and putative function identification. In addition, the RGAP database provides gene co-expression analysis, aimed at identifying co-expressed genes for a selected candidate gene based on 15 array series obtained from biotic and abiotic stress experiments in rice. The Sequence Read Archive (SRA) is a public database maintaining primary next generation sequence (NGS) data across various experiments (Leinonen et al. 2011). Forty-eight RNA-Seq SRA libraries comprising diverse tissues/organs and stresses are mapped to the genome and the number of fragments per kilo base of exon per million fragments (FPKM) is assigned to individual genes. In addition, a batch data retrieval option is available at http://rice.plantbiology.msu.edu/downloads_gad.shtml. By using locus identifiers (IDs), users conveniently retrieve locus Genomic sequences, Gene model sequences, Protein sequences, Putative functions, and GOSlim assignments.

RAP-DB also has adopted the Rice Genome Sequencing Project (IRGSP) Os-Nipponbare-Reference-IRGSP-1.0 for annotation purposes (Sakai et al. 2013). Later on curtailed the sequencing errors using ~44X coverage NGS sequencing reads. RAP-DB provides annotation information for 37,869 loci based on transcripts, FLcDNAs, and protein sequence information of 180 monocot species obtained from various sources including the DNA Databank of Japan (DDBJ) (Kosuge et al. 2013) and the Triticeae Full-Length CDS Database (TriFLDB) (Mochida et al. 2009). A feature of RAP-DB is the locus identifier converter between RAP-DB and RGAP (http://rapdb.dna.affrc.go.jp/tools/converter). Lineage analysis as well as phylogenomic studies indicate *Oryza rufipogon* as the possible common ancestor of *Oryza sativa ssp. indica* and *japonica* (Yang et al. 2011). RAP-DB holds information on 15 *O. rufipogon* genes which were not detected in the *japonica* genome, representing a valuable set of genes for the analysis of rice domestication and species divergence (Sakai et al. 2013). The batch retrieval option retrieves diverse sequences, GO, domain, genomic description, and the chromosomal location of an input gene locus set. In the same search module, a marker-based search option is enabled, which retrieves the sequence between two specified markers. The BLAST Like Alignment Tool (BLAT) (Kent 2002) is a homology search tool that is functionally similar to the Basic Local Alignment Search Tool (BLAST), with high accuracy and speed. The RAP-DB homology search function relies on BLAST and BLAT searches. The Short-Read Assembly Browser (S-RAB) is a newly developed feature of the RAP-DB database that can be used to align the short read genome sequences of representative *japonica* and *indica* cultivars (downloaded from both DDBJ and European Nucleotide Archive ftp sites) to the reference rice genome. Using the S-RAB tool, it is easy to identify the chromosomal location associated with a Single Nucleotide – Polymorphism (SNP). Further increasing the number of short read datasets constitutes a powerful tool in identifying DNA polymorphism related to cultivars. Genome-wide expression profiles comprising 27 RNA-Seq datasets are incorporated into RAP-DB, as in RGAP.

The Rice GE database is another standard rice genome browser with data belonging to major categories such as genes, insertions, FLcDNA, EST, markers, homology, and a

Table 1. Summary of rice genome browsers

Databases	Rice genome	Total rice loci	Expression data platform	Resource link	Reference
MSU-RGAP[a]	Japonica	55,986	Affymetrix (548)[b] RNA-Seq (16)[b] DEG (32)[b]	http://rice.plantbiology.msu.edu/	Kawahara et al. 2013
RAP-DB[c]	Japonica	37,869	RNA-Seq (27)	http://rapdb.dna.affrc.go.jp/	Sakai et al. 2012
RiceGE	Indica, japonica	55,986 (japonica)	Affymetrix (155)	http://signal.salk.edu/cgi-bin/RiceGE	N.A[d]
Gramene	Multiple species genome	45,420	N.R[e]	http://www.gramene.org/	Monaco et al. 2013

[a]Indicates the rice genome annotation project managed by Michigan State University.
[b]In addition to the Affymetrix data, MSU-RGAP also incorporated 16 RNA-Seq and 32 Digital Gene Expression (DGE) libraries. The total number of microarray/RNA-Seq samples is indicated in parenthesis.
[c]Indicates the rice genome annotation database.
[d]Indicates that information is not available.
[e]Indicates that the Gramene database currently does not support the expression data.

Table 2. Summary of web databases for rice orthology analysis

Databases	Total plant species[a]	Rice protein sequences[b]	Resource link	Reference
GreenPhylDB	16	67,393	http://www.greenphyl.org/v2/cgi-bin/index.cgi	Rouard et al. 2010
Phytozome	45	49,061	http://www.phytozome.net/	Goodstein et al. 2011
Plaza	22	42,211	http://bioinformatics.psb.ugent.be/plaza/	Proost et al. 2009
InParanoid	85	67,393	http://inparanoid.sbc.su.se/cgi-bin/index.cgi	Ostlund 2010

[a]Indicates the number of plant species used for orthology analysis.
[b]Indicates the total number of rice protein transcripts used for orthology analysis.

Yale tiling array (http://signal.salk.edu/cgi-bin/RiceGE). In contrast to the RGAP database, the RiceGE genome browser supports both the *indica* and *japonica* genomes. The i-Sect toolbox is a collection of tools for manipulating and mining information based on the genomic data. Gene expression data from two platforms (GPL2025 and GPL477) comprising 10 series have been included and can be fetched by the gAtlas Tool. Data related to genes, T-DNAs, or cDNAs can be retrieved by the mGene tool. Insertional mutant information of a locus is linked to 11 T-DNA databases, and provides a strong platform for functional studies based on the loss of function approach. A homology comparison of rice cDNA with wheat, maize, barley, and brassica is also provided. One unique feature of the RiceGE database is the primer design option for genome mapping data or input sequence information using PrimerL and PrimerR tools.

Orthology Identification Databases

Comparative genomics approaches have been successfully applied to reveal species diversity as well as transferring economically important QTL knowledge from other crop species to rice (Rouard et al. 2011). Recently, a comparatively analyzed protein-protein interaction pattern under biotic stress showed high synteny between rice and wheat (Cantu et al. 2013). Identification of the correct orthologs is a crucial step in these studies. In this regard, the Greenphyl database enables us to identify orthologs and paralogs associated with 22 full genomes from algae to angiosperms (Rouard et al. 2010). Gene family clustering was done based on the highly annotated protein sequences from rice and Arabidopsis. Later, other genome sequences were allocated to the clusters. Examples of other comparative genomics databases dealing with rice orthologue data include Phytozome, PlantGDB, Plaza, InParanoid, and OrthoMCL DB. Various features of these databases are given in Table 2.

As mentioned earlier, a number of economically important QTLs in rice are found to be orthologs in other species, or verse visa. An orthologous relationship between rice and other species can be used in the development of markers in cereal crops. For example, genome-wide analysis revealed 827 conserved ortholog set pairs between rice and wheat and this identification contributing to a marker-based approach for QTL mapping (Quraishi et al. 2009). Interolog method for a protein-protein interaction network relies on orthologs from the other species (Matthews et al. 2001). That is, experimentally-proven interaction data for a pair or group of orthologous proteins in a given species are used to construct a predicted protein-protein interaction network in other species. Thus, the identification of more interologs from diverse species with a large scale protein-protein interaction

network will contribute to the expansion of the predicted protein-protein interaction network in rice. The qualified identification of orthologs from other species with more functional information represents a useful landmark to facilitate the functional study of rice genes.

Databases for Transcriptome Analysis

Using rice as a model for crop studies, multiple databases have been developed to store and analyze publicly available rice array data. These resources allow researchers to easily retrieve the transcriptome data for single genes or a group of genes of interest. For example, the Rice Oligonucleotide Array Database (ROAD) incorporated 1,867 microarray slides from 105 experiments and the associated tools analyze the co-expressed genes, meta-expression profiling, highly expressed genes, and GO and KEGG ontology (KO) enrichment. Meta-expression profiling analysis is provided for two categories namely 'Developmental stages' and 'Anatomy' for two independent platforms, Affymetrix and Agilent 44K. Further, the co-expression network option of co-expression analysis generates gene networks associated with queried gene(s) under different categories such as general (whole dataset), biotic-stressed samples, and abiotic-stressed samples (Cao et al. 2012). GO enrichment analysis of locus identified by expression profiling data related to early and prolonged heat stress showed that the chaperonine-mediated protein folding co-factor was over-represented (Jung et al. 2013). Genome-wide microarray data sets used in RiceXPro are based on RAP-DB curated transcripts and cDNA information from the Knowledge-based Oryza Molecular biological Encyclopedia (KOME) database (http://cdna01.dna.affrc.go.jp/cDNA/). There are three categories according to the features of the data sets: Field/Development, Plant hormone, and Cell- and Tissue-Type (Laser Microdissection). Of these features, the 'Field/Development' category deals with the expression profiling data of anatomical tissues across an entire growth period under natural field conditions, and by diurnal and circadian regulation. The 'Plant hormone' category provides expression profiling of roots and shoots under treatment with six major plant hormones (i.e., ABA, GA, Auxin, BR, cytokinin, and JA), comprising 143 data points corresponding to 13 data sets. No other databases have dealt with hormone responsive data. The 'Cell- and tissue-type' category was created by combining laser microdissection and microarray approaches, containing 38 samples. The global profile search module integrates the data from all of these categories and summarizes the gene response in a single search (Sato et al. 2011). The above mentioned databases hold data exclusively related to the *Oryza sativa*.

Alternatively, the Plant Expression Database (PlexDB) holds the expression data of 14 plants and nine pathogens with a unique feature of being able to accommodate uploaded user data for further comparative analysis (Dash et al. 2012). Among the listed pathogen resources, the rice blast fungus *M. grisea* is included, providing a source for rice pathogenic studies. Gene Oscilloscope is a data mining module in PlexDB for identifying the differential expression of a gene or a set of genes across all incorporated experiments based on coefficient of variation (CV) values. As a homology prediction module, the Model Genome Integrator retrieves a complete annotation set for the homologue of an input gene and provides a link to other genome browser databases. The meta-analysis reference database, Genevestigator, is designed for the analysis of gene expression in different contexts based on data from 2,081 Affymetrix arrays (Zimmermann et al. 2008). The graphical user interface presents grouped tool sets: conditional search tools, gene search tools, and similarity search tools. 'Conditional search tools' visualize the expression patterns of queried gene(s) from selected samples or experiments, anatomy, perturbations, and development. 'Gene search tools' search for genes that are the most stable in chosen tissues and conditions, specifically expressed genes in chosen tissue types, specifically expressed genes in chosen perturbations, and genes associated with a specific developmental profile. 'Similarity search tools' group genes according to similarity of expression patterns based on a selected dataset, and identify the conditions in which a user-specified expression signature occurs. Extensive cross-talk between different hormone signals is a reason for hormone signaling complexity. Expression profiles of hormones and related genes in various tissues can facilitate an integrative view of related networks. Uniformed Viewer for Integrated Omics (UniVIO: http://univio.psc.riken.jp/) stores comprehensive analysis of 43 hormone-related compounds as well as transcriptome data of 14 organ parts of a rice plant at the reproductive stage (Kudo et al. 2013). Therefore, UniVIO provides a combined heat map of hormone-metabolome and transcriptome data. In addition, this database includes transcriptome data obtained from seedling shoots of three gibberellin signaling mutants. Table 3 summarizes the features of the major resources for transcriptome analysis in rice.

The emergence of next generation sequencing technologies (NGS) has caused a paradigm shift in conventional genomics approaches and accelerated data deposition in publicly-available databases such as NCBI gene expression omnibus (GEO) and ArrayExpress (Parkinson et al. 2005). Various reviews have discussed the different NGS platforms and their features. RNA-Seq-based profiling was used to develop high resolution transcriptome data by not depending on fixed gene information, and overcame the limitations of microarray techniques in the identification of an exon splicing junction (Lu et al. 2010). Interestingly, it was found that 52% of the

Table 3. Summary of databases for rice transcriptome analysis

Databases	Platforms	Number of plant species	Number of rice array	Resource link
ROAD	Affymetrix, Agilent 22K, Agilent 44K, BGI/Yale, NSF20K, NSF45K	1	105[a]	http://www.ricearray.org/index.shtml
RiceXPro V3	Agilent 44K	1	753[b]	http://ricexpro.dna.affrc.go.jp/
PlexDB	NSF20K, NSF45K, Affymetrix	14	102[a]	http://www.plexdb.org/
Genevestigator	Affymetrix	9	2081[b]	https://www.genevestigator.com/gv/plant.jsp

[a]Indicates the number of microarray samples.
[b]Indicates the number of total experiments.

identified (15,708) novel transcriptional activation regions (nTARs) had no homologues in the annotated protein sequences of public genome browsers. NGS-based expression profiling of rice under different developmental stages as well as under perturbations is available from a few databases such as RGAP and RAP-DB. The amount of RNA-seq data in these databases is quite limited. In addition, manual analysis of RNA-seq data is particularly complex because it deals with millions of short reads and demands high performance computing, hindering the universalization of this technique for more researchers. Even though restrictions exist in RNA-seq analysis, this approach will take the place of databases for microarray data in the near future.

Databases for Protein-Protein Interaction Network and Functional Gene Network Analyses

Major biological responses of an organism either internally or to external environments are mediated by protein-protein interactions in a series of complex signaling cascades (Pawson and Nash 2000). These understandings have led to the development of different approaches for elucidating interactions between proteins. Current developments in plant protein interactomes relying on techniques such as yeast two-hybrid (Y2H), split-ubiquitin (sUbq), fluorescence resonance energy transfer (FRET), bioluminescence resonance energy transfer (BRET), bimolecular fluorescence complementation (BiFC), mass spectrometry, affinity purification (AP), and tandem affinity purification (TAP) are well understood (Morsy et al. 2008; Braun et al. 2013). Protein-protein interactive network predictions are based on the assumption that protein complexes among different species share high structural and functional conservation during evolution (Braun et al. 2013). That is, evolutionary conserved proteins are likely to share their interactions across species (Ho et al. 2012). Consistent with this assumption, despite more than 50 million years of divergence from their common ancestors, comparative analysis of protein-protein interactions in defense responses of rice and wheat revealed significant similarity (Cantu et al. 2013).

Even though a high resolution genome-wide interactome map has not yet been achieved in rice, many efforts have been made to develop sub-genome level interactome maps related to cell cycle, seed germination, and stress response (Cooper et al. 2003a; Cooper et al. 2003b). In other areas, we have to depend on the identification of interologs, which have determined protein-protein interactions from other species. The Rice Interactions Viewer (RIV) is a web tool for predicted protein-protein interactions based on a confidence value along with experimental data from the IntAct database (Kerrien et al. 2012). Given a list of gene IDs, the search can be done against all of the predicted and published protein-protein interactions. Subsequently, a network file in either SIF or CYS format for queried gene ID(s) is generated. The Predicted Rice Interactome Network (PRIN) is another web tool which can be used for protein-protein interaction analysis in rice (Gu et al. 2011). The strategy for the identification of orthologs in the PRIN database is in consensus with RIV, and the InParanoid algorithm was employed for the identification of orthologs. There are two options for searching related networks. The protein search module in PRIN provides gene ontology, subcellular localization, cross-reference, and protein interactions. The PPI module provides a predicted protein-protein interaction network as an input file carrying a list of queried genes.

The STRING database (http://string-db.org/) provides predicted and known protein-protein interaction analysis for 1,133 organisms including rice (Table 4). Direct (physical) and indirect (functional) associations consist of the interactions which are derived from four sources: genomic context, high-throughput experiments, conserved co-expression, and previous knowledge (Szklarczyk et al. 2011). Users use a single protein of interest, or a set of protein names to retrieve the associated interaction network. There are a total of eight options for the output images, adjustment of network complexity, and saving of output data. In addition, the STRING database provides a brief summary of the input protein and predicted functional partners. By using the save option, users can easily store the

Table 4. Summary of databases for predicted rice protein-protein interaction analysis

Databases	Total interactions	Organisms	Rice proteins	Method	Resource link
Rice interaction Viewer	37,472[a]	1[c]	4,567[e]	Interolog, Experimental data	http://bar.utoronto.ca/interactions/cgi-bin/rice_interactions_viewer.cgi
PRIN	76,585[a]	1[c]	5,049[e]	Interolog, Experimental data, GO annotation, Sub cellular location, gene expression data	http://bis.zju.edu.cn/prin/
STRING	336,561,678[b]	1,133[d]	5,214,234[f]	Experimental data, gene expression data	http://string-db.org/
DIPOS	14,614,067[a]	1[c]	27,746[e]	Ortholog relationship, Domain-domain interaction.	http://csb.shu.edu.cn/dipos/?id=5

[a]Indicates the total number of protein-protein interactions predicted in each database.
[b]Indicates the total number of protein-protein interactions predicted in STRING, including non-rice organisms.
[c]Indicates a specialized database for rice.
[d]Indicates the total number of species analyzed in STRING. The STRING database is not specific to rice.
[e]Indicates the total number of proteins which have interologs in each database.
[f]Indicates the total number of proteins analyzed in STRING, including non-rice organisms.

image data of the network, annotations of proteins in the network, and information related to the interaction of a pair of proteins in the network. The STRING database intends to provide convenient modules for searching and browsing the protein interaction data, as well as for inspecting the underlying evidence. One of the unique features in the STRING database is to analyze the enrichment of functional systems such as Gene Ontology, KEGG, Pfam, and InterPro (Franceschini et al. 2012). Resulting enrichments can be analyzed interactively, providing a visual highlight of the corresponding proteins in the network. In addition, the enrichments can be supported by their *P*-value through a Hypergeometric test (Rivals et al. 2007; Franceschini et al. 2012).

Considering the limitations in the availability of rice functional information, the RiceNet database (http://www.functionalnet.org/ricenet/) was constructed by integrating diverse gene to gene relationship data from multiple organisms to the initial rice framework using modified Bayesian integration (Lee et al. 2011). The Log Likelihood score (LLS) is a single value representation for an interacting gene pair, which is given based on diverse datasets. As a validation of this database, 15 genes with validated phenotypes were queried, and the tool retrieved 802 genes with predicted function in XA21-mediated immunity (Cantu et al. 2013). Candidates screened by LLS methods are experimentally confirmed and ultimately validated as positive and negative regulators of biotic stress response in rice.

Databases for Co-Expression Network

The similarity in gene expression of a group of genes across diverse experimental conditions provides a hint to their biological function. The Arabidopsis orthologue of rice *DREB1* involved in cold response was identified through co-expression analysis under chilling stress (Mao and Chen 2012). Meta-analysis of differentially-expressed genes (DEGs) between rice and Arabidopsis showed that DEGs in both model plants share a conserved expression under stress conditions (38.5% and 28.7%, respectively) (Shaik and Ramakrishna 2013). Co-expression studies are generally performed based on statistical methods such as the Pearson correlation coefficient, mutual ranking, Spearman correlation coefficient, and partial correlation coefficient (Ma et al. 2013). RiceFREND (http://ricefrend.dna.affrc.go.jp/) is a platform for the prediction of gene function in rice based on co-expression data which are obtained from the Agilent 44K array dataset available in the RiceXPro database (Sato et al., 2013). There are two options for a co-expression search: single guide gene and multiple guide genes. These options retrieve the co-expressed genes of a queried gene or a list of queried genes. The HyperTree option in the result table shows the relationship of co-expressed genes in a network model by integrating the mutual rank (MR) value according to the node hierarchy. In addition, users can obtain diverse information on target gene function such as gene description, GO/GOslim terms in three principle categories, KEGG pathway, transcription factor, gene symbol, WoLF PSORT-based subcellular location, and MSU ID. Of these, associated KEGG pathways can be indicated by different colors according to the KEGG pathway table. In addition to having a link to the KEGG pathways, the network can be viewed in the Cytoscape or Graphviz format by using the "other viewers" option. To guide the biological function or molecular regulation of co-expressed genes, analysis tools provide a GO enrichment test and cis-element analysis (Sato et al. 2013). Cis-element (motif) analysis is a common practice in systems biology to find out transcriptional regulation mechanisms associated with a developed network. Furthermore, this analysis has been successfully employed in the identification of motif-based sub-

Table 5. Summary of databases for rice co-expression analysis

Databases	Platform method	Number of rice genes	Organisms	Total number of experiments or microarray data for rice	Resource link	Reference
Ricefrend	Agilent Microarray	27,201	1	815[c]	http://ricefrend.dna.affrc.go.jp/	Sato et al. 2012
Oryzaexpress	Microarray	NA[a]	1	871[c]	http://bioinf.mind.meiji.ac.jp/Rice_network_public/script/	Hamada et al. 2011
ATTED-II	Affymetrix	20,625	7[b]	1,214[c]	http://atted.jp/	Obayashi. 2011
PLANEX	Affymetrix	57,382	8[b]	884[c]	http://planex.plantbioinformatics.org/	Yim et al. 2013
STARNET	Affymetrix	23,419	10[b]	148[d]	http://vanburenlab.medicine.tamhsc.edu/starnet2.html	Jupiter. 2009
Plant ArrayNet	Affymetrix	58,417	3[b]	183[c]	http://arraynet.mju.ac.kr/arraynet/	Ho-Lee et al. 2009
CoP	Affymetrix	NA[a]	8[b]	884[c]	http://webs2.kazusa.or.jp/kagiana/cop0911/	Ogata et al. 2010
PlaNet	Affymetrix	NA[a]	8[b]	156[c]	http://aranet.mpimp-golm.mpg.de/	Mutwil et al. 2011

[a]Indicates that information is not available.
[b]Indicates that the databases are not specialized for rice, and the number represents the total number of species analyzed.
[c]Indicates the total number of samples.
[d]Indicates the total number of experiments.

modules from the network (Ma et al. 2013). Genome-wide analysis of ABA responsive co-expressed genes in rice showed that their promoters are overrepresented by an ABA responsive element containing the CGMCACGTGB motif. This prediction was further confirmed by RT-PCR, indicating the significance of co-expression-based gene identification (Lenka et al. 2009).

The co-expressed genes in all of these databases represent the average view of all potential gene relationships. It is known that the extent of co-expression between genes fluctuates in different conditions. To address this issue, a 'condition-specific comparative view' has been enabled in the Atted II co-expression database. An updated version of the Atted II database includes rice co-expression data along with five other plant species, providing an opportunity for extensive comparative analysis with an Arabidopsis dataset (Obayashi et al. 2011). Compared with other similar databases, the PLANEX database applied a different approach for co-expression identification. That is, the Pearson correlation coefficient (PCC)-derived co-expression gene set was further subjected to principal component analysis (PCA), a popular method applied to microarray data for efficient identification of a gene set with changing expression over different experiments (Yim et al. 2013). The GO-Term Finder of the tool deals with GO analysis for a co-expressed gene set. The advantages of the PLANEX database are 1) the retrieval of co-expressed genes from eight species, and 2) further clustering and comparison of co-expressed genes. The available resources for co-expression analysis in rice are summarized in Table 5.

Database for Promoter Analysis

Various resources dealing with rice and other plant promoters include PLACE (Higo et al. 1998), Osiris (Morris et al. 2008), PlantCARE (Lescot et al. 2002), Grasslus (Yilmas et al. 2009), Plant Promoter DB (ppdb) (Yamamoto and Obokata 2008), and PlantPAN (Chang et al. 2008). These databases are summarized in Table 6. The Osiris resource integrates 92 experimentally-validated TF binding consensus sequences

Table 6. Summary of databases for promoter analysis in rice

Databases	Number of organisms	Number of rice gene promoters	Resource link	Reference link
Osiris	1	24,209[b]	http://www.bioinformatics2.wsu.edu/cgi-bin/Osiris/cgi/home.pl	Morris et al. 2008
Grassius	4[a]	56,278[b] (NA[c])	http://grassius.org/grasspromdb.html	Yilmas et al. 2009
PPDB	5[a]	600[b] (17,286[d])	http://ppdb.agr.gifu-u.ac.jp/ppdb/cgi-bin/index.cgi	Yamamoto and Obokata 2007
PlantPAN	3[a]	62,827[b]	http://plantpan.mbc.nctu.edu.tw/	Chang et al. 2008

[a]Indicates that the databases are not specialized for rice, and the number represents the total number of species analyzed.
[b]Indicates the total number of rice gene promoters serviced in each database.
[c]Indicates that information is not available.
[d]Indicates the total number of gene promoters serviced in the PPDB database, including non-rice organisms.

and 67 microarray datasets. Two kinds of promoter sequences are given. One is based on RGAP annotated data and the other corresponds to experimentally determined gene transcript data. The promoter sequence of an input gene can be comprehensively analyzed by the 'Visualization' search module. The 'Data mining' module allows the user to retrieve genes containing selected transcription factor (TF) binding sites or GO terms. By providing a consensus motif sequence query in the 'Custom motif' module, users retrieve all of the rice promoter sequences which contain the input sequence. The 'Analysis suite' deals with the positional aspects of TF binding sites in the promoter sequence. This module is linked to microarray datasets through which one can analyze the expression profile of genes of interest. Additionally, the correlation tool allows the user to analyze the relationship between transcription factor binding sites (TFBS) and gene expression.

PlantPAN (http://plantpan.mbc.nctu.edu.tw/index.php) is a plant promoter analysis navigator (Chang et al. 2008). PlantPAN has four modules for promoter analysis: Gene Group, Promoter analysis, Search, and Cross species. Of these, the 'Gene Group' module provides seven analysis steps to identify TFBSs conserved in a group of queried genes: 1) input gene group, 2) GO analysis, 3) promoter extracting, 4) Setting parameters, 5) Scanning TFBSs, 6) Co-occurrence analysis, and 7) Combinatorial analysis. The 'Promoter Analysis' module starts with a queried sequence name or queried promoter sequence in the FASTA format and can retrieve TFBSs, Tandem Repeats, CpNpG islands, and the miRNA target site in the input promoter. By providing gene_IDs, Locus, or Keyword, the 'Search' module provides transcription factor binding sites (TFBS), tandem repeats, and CpNpG islands in the input promoter sequence. There are three options for searching: Search by Gene, Search Promoter in Experimental Reference, and Search by transcription factor. The 'Search by Gene' option in the search module analyzes the GO terms, genomic DNA sequence, 5'UTR sequence, cross promoter analysis using paralogs and orthologs, promoter analysis with the option of selecting TFBSs, Tandem Repeat, CpNpG islands, and the upstream length from the transcription start site. The 'Search Promoter in Experimental Reference' option retrieves promoter sequences of genes in the experimentally-validated reference dataset. The 'Search by transcription factor' option provides information on selected TFs by linking the resources to PubMed and the PLACE TF database (Higo et al. 1999). There are three options in the 'Cross species' module: HOMOLOGENE SEARCH, BLAST SEARCH, and BLAST 2 SEQ SEARCH. The HOMOLOGENE SEARCH provides TFBSs shared with paralogs and orthologs. The BLAST SEARCH analyzes conserved regions between the input promoter sequence and its target promoter sequence. The BLAST 2 SEQ SEARCH provides conserved regions between two input promoter sequences.

The Grassius GrassPROMDB database is built based on experimentally validated cis-regulatory elements (CREs), and analyzes their sequence conservation with orthologus promoters of other organisms. The 'Browse promoter' option provides all of the promoters in a species, but this tool is not available for rice. The 'Search by gene ID' option, including queries for multiple gene IDs, identifies the number of experimentally-proven CREs and TFs associated with an input gene. In the 'Search by blast' option, an input sequence can be searched against a rice promoter dataset. The local distribution of short sequences (LDSS) is a method to identify the promoter constituents from the genomic region. Plant Promoter Database (PPDB) adopts this statistical method to identify the promoters and classifies cis-elements in promoters into different categories such as regulatory element groups (REG), TATA box, and Y patch. Thus, the PPDB database comprehensively describes the promoter information via a single gene or keyword query. The description includes an interactive view of the promoter structure, transcriptional start site (TSS), and REG.

Databases for Metabolic Pathway Analysis

Over the years, different methods have been adopted to provide pathway level information associated with a set of genes. Previously, over representation analysis (ORA) and functional class scoring approaches (FCS) were used, respectively. On the other hand, current approaches rely on pathway topologies (PT). Rather than simply listing out the genes involved in a pathway, PT-based resources focus on a comprehensive description of protein level interactions and cellular localization of the pathways with compound information (Khatri et al. 2012). The KEGG database is an effective platform for the systems biology approach allowing users to straightforwardly model, simulate, browse, and retrieve multi-omics data using this database. The KEGG pathway database is one of the largest curated metabolic pathway databases consisting of 123 rice pathways with 283,360 total references (Kanehisa et al. 2012). KEGG maintains five main databases: KEGG pathway, KEGG BRITE, KEGG Module, KEGG Mapper, and KEGG Atlas. KEGG pathway, as a collection of manually drawn pathway maps, presents the molecular interaction and reaction networks. Metabolic information is the main content, and it deals with 13 metabolism maps which include Global and overview maps, as well as 12 specialized metabolisms. KEGG BRITE, a collection of hierarchical classifications, provides a comprehensive understanding through integrating different types of relationships. KEGG module can be used to provide

annotation information and biological interpretations of sequenced genomes. There are four types of KEGG modules: pathway modules, structural complexes, functional sets, and signature modules. KEGG Mapper, a collection of tools for KEGG mapping, has three options: KEGG pathway mapping, BRITE mapping, and MODULE mapping. For each mapping tool, there are three or four sub-options: Search, Search & Color, Color or Join, and Reconstruct. Of them, the 'Search pathway' option in the KEGG Pathway Mapping tool retrieves the user-querying dataset consisting of genes, proteins, reactions, or other compounds in the KEGG pathway maps, and mapped objects are marked in red. Multiple coloring options help users to easily differentiate multiple reactions against the background pathway. BRITE mapping and Module mapping tools perform similar strategies with pathway mapping in terms of searching a user list against a respective database. KEGG Atlas is an advanced graphical interface used to search the KEGG Global and overview maps with capabilities of surveying KEGG modules. However, the KEGG Atlas is only available for eight pathways in the KEGG Global and overview maps.

Reactome is an open source, manually-curated database dealing with reactions, complexes, and pathways (Croft et al. 2011). Applications involve pathway editing, genome analysis, and systems biology. Twenty one species, including the human genome as a central standard, have been serviced, and some of the options are restricted to animal species. In total, 6,284 rice proteins and 1,130 reactions are integrated in this database. The 'Analyze data' option integrates and summarizes the dataset comprising genes/proteins/expression analysis into a single platform. The 'Compare species' option analyzes the extent of the similarity in pathways between human and user specified species. The rationale behind this comparison is the identification of orthologue proteins of a specific pathway in a species with a similar pathway in the human genome. Details of involved pathways and the physiochemical properties of a small molecule can be retrieved by the connectivity of the Reactome database to the Chemical Entities of Biological Interest (ChEBI) database. Furthermore, this resource is linked to a number of different chemical databases.

The MapMan application tool provides a hierarchical view of blocked functional classification for a gene list by assigning input candidates into different major functional categories (Thimm et al. 2004). The advantage of block representation rather than pathway representation is about the tentative assignment of genes whose function is not well characterized. The working principle of the tool is based on the SCAVENGER and IMAGEANNOTATOR modules. The SCAVENGER module works by functional classification of data retrieved from public resources into different levels, namely BIN and subBIN, generating a mapping file. The IMAGEANNOTATOR module is involved in importing the classification to different blocks and displays the integrated block classification data. Even though the tool is optimized to be applied to the Arabidopsis genome, bioinformatics approaches have been developed to broaden the target genomes by expanding the BIN size of the MapMan. For this extension, the KEGG orthology database was converted into the BIN structure of the MapMan, and mapping information of 47 species including rice and Arabidopsis is available at the MapManStore (http://mapman.gabipd.org/web/guest/mapmanstore) (Goffard and Weiller 2006). MapMan analysis of 589 early heat response genes in rice shows that diverse categories of genes are involved in heat stress (Jung et al. 2013). MapMan has been used to analyze metabolic pathways associated with transcriptome data in rice and non-rice species (Rotter et al. 2007; Degenkolbe et al. 2009; Usadel et al. 2009).

Another hierarchically-organized database, the Ricecyc pathway database, is available as a part of Gramene's organism specific databases. Ricecyc holds 2,103 enzymatic reactions and 87 transport reactions of 6,040 enzymes and 603 transporters, respectively (Dharmawardhana et al. 2013). Initial development was based on the known and predicted pathways from meta-cyc. Manual curation based on published literature is subsequently applied. The 'Comparative analysis' option enables interspecies comparison of different attributes in the database. For example, a user can compare a rice pathway with any of the nine species-specific pathways and general plantCyc pathways. Elements which are unique, common, or absent in this comparison are represented by a different color code, and this tool allows users to easily analyze the divergent points in the pathway among species. The 'pathway tools omics viewer' module functions by allowing the integration and visualization of the expression pattern of high throughput data such as transcriptome, proteome etc. into a cellular overview. Expression data can be mapped through the RGAP locus ID. The 'Omics validator' tool deals with the conversion of a user-provided microarray probe ID, gene name, or symbols to a gene ID (RGAP locus ID). For further analysis, the database can be downloaded in BioPAX and SBML formats. RiceCyc has functionalities of systems level analysis such as metabolic flux balance analysis (FBA). This tool was recently used for elucidating metabolic networks and has been exploited in understanding drought-related factors contributing to metabolic fluctuations (Lakshmanan et al. 2013).

Databases to Assess Gene Indexed Mutants

The development of rice as a representative model organism for crop species is driven by its high quality genome and

comparatively good annotation among other crop species. International efforts, which have been initiated by the International Rice Research Institute (IRRI) and organized by the International Rice Functional Genomics Consortium (IRFGC), have developed a huge collection of indexed rice mutant genes to facilitate further functional analyses. This progress is a key stepping stone towards the IRFGC's goal of elucidating the function of the entire predicted rice genome by 2020 (Zhang et al. 2008). These resources are distinguished according to the type of mutagen such as T-DNA or transposons, rice variety, mutated loci per genome, and seed availability. Insertional elements, gene entrapment, activation tagging, and chemical and physical mutagenesis have been employed for the development of a gene indexed mutant population (Krishnan et al. 2009). Details of different methods adopted for mutation and the features of such mutant resources have been reviewed (Hirochika et al. 2004; Krishnan et al. 2009; Chang et al. 2012). Among the different mutagenesis approaches, T-DNA insertion mutagenesis has emerged as a popular strategy and mutations covering more than half of the rice genome has been developed by the collective international effort (Jung and An 2013). The advantage of this approach is in the rapid identification of mutated genes based on the T-DNA tag sequences. The deposition of isolated Flanking Sequence Tags (FSTs) for the rice research community provides additional advantages for systems biology (Chang et al. 2012). The irregular insertion position of T-DNA in the chromosome poses a challenge in isolating the FSTs. However, NGS methods have been used in the identification of insertion sites caused by the irregular insertion of the T-DNA sequence (Chang et al. 2012; Polko et al. 2012).

Table 7. Summary of databases for pathway analysis in rice

Databases	Number of organism	Number of rice proteins	Number of rice pathways	Resource link	Reference
KEGG	2,959	35,681	278,001[b]	http://www.genome.jp/kegg/	Kanehisa et al. 2012
Reactome	21	6,284	687	http://www.reactome.org/	Croft et al. 2010
Ricecyc	1	6,040	311	http://pathway.gramene.org/gramene/ricecyc.shtml	Dharmawardhana et al. 2013
Mapman	47	N.A[a]	N.A[a]	http://mapman.gabipd.org/web/guest	Thimm et al. 2004
PANTHER	82	15,989	628	http://www.pantherdb.org/pathway/	Mi et al. 2012

[a]Indicates that information is not available.
[b]KEGG does not provide the information on the total number of pathways in rice. Hence, the total number of pathways, including non-rice organisms, is provided.

Table 8. Summary of databases for rice gene indexed mutant population

Databases	Genotype[a]	Number of FSTs	Number of mapped FSTs	Number of mapped FSTs to genic region	Number of mapped genes to genic region	Resource link
RMD	Zhonghua 11 Zhonghua 15 Nipponbare	33,197	31,892	16,219	6,641	http://rmd.ncpgr.cn/
POSTECH-RISD	Dongjin Hwayoung	107,171	105,739	59,707	20,889	http://onlinelibrary.wiley.com/doi/10.1046/j.1365-313x.2000.00767.x/abstract(**)
TRIM	Tainung 67	11,799	111,35	6,029	3,735	http://trim.sinica.edu.tw/
SHIP T-DNA	Zhonghua 11	12,614	9,385	4,808	1,340	http://ship.plantsignal.cn/home.do
CIRAD	Nipponbare	29,262	27,870	17,709	5,633	http://oryzatagline.cirad.fr/
CSIRO	Nipponbare	611	585	398	287	http://www.csiro.au/pi
EU-OSTID	Nipponbare	1,315	1,290	814	672	http://orygenesdb.cirad.fr/
UCD	Nipponbare	17,730	16,825	7,556	3,103	http://onlinelibrary.wiley.com/doi/10.1111/j.1365-313X.2005.02570.x/abstract (**)
NIAS	Nipponbare	18,024	17,939	14,554	3,604	https://tos.nias.affrc.go.jp/
GSNU	Dongjin Byeo	1,072	1,050	732	484	http://onlinelibrary.wiley.com/doi/10.1111/j.1365-313X.2004.02116.x/abstract(**)

[a] indicates the variety of rice cultivar used for developing mutant population resource.
** indicates that the website is not available or working. Hence publication link is provided.
Resource abbreviations and their expanded forms are given below.
Rice Mutant Database (RMD); Rice T-DNA Insertion Sequence Database (POSTECH RISD); Taiwan Rice Insertion Mutant (TRIM); Shanghai T-DNA Insertion Population (SHIP T-DNA); CIRAD/Genoplante oryza tag lines (CIRAD); CSIRO Ac/Ds (CSIRO); OrygenesDB database and from the EU (EU-OSTID); University of California, Davis (UCD); National Institute of Agrobiological Sciences (NIAS); Gyeongsang National Univ. (GSNU).

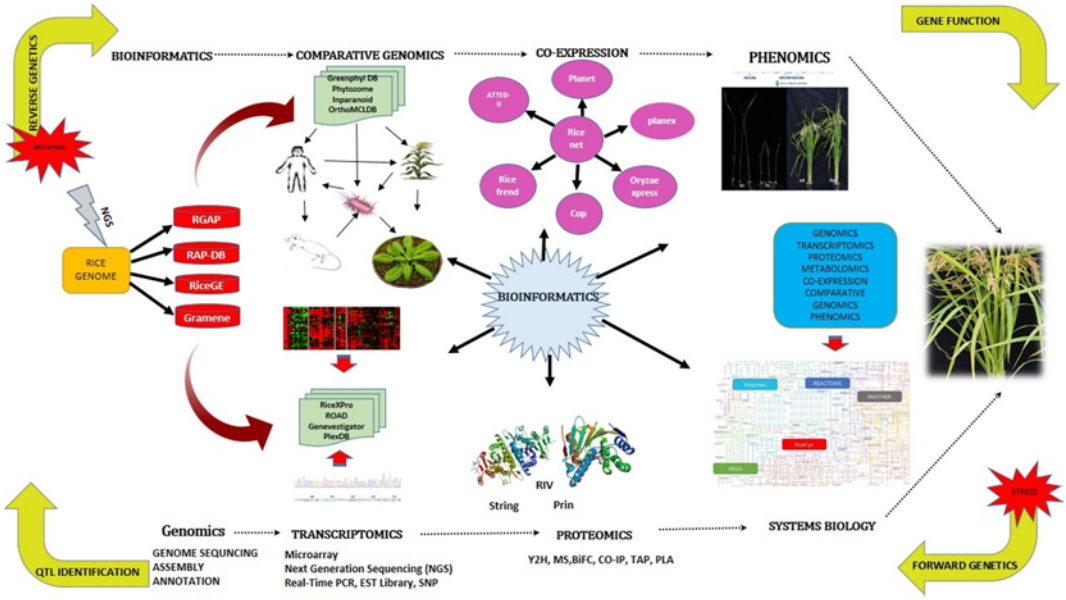

Fig. 1. Forward and reverse genetic strategies are employed for functional analysis of rice genes. Application of NGS methods results in a high quality rice genome sequence. This achievement resulted in tremendous development in areas such as transcriptome analysis, comparative genomics, proteomics, and co-expression analysis. Newly emerged interdisciplinary approaches, namely Bioinformatics and Systems biology, integrated the information from these resources and facilitated the functional characterization of rice genes. "ho" indicates homozygous progenies with the insertion of mutagen; "wt", the wild type segregants.

Beyond the level of a genome browser, RiceGE also represents a standard database for mutant information. In total, 11 mutant databases have been linked to the RiceGE database. Of these mutant resources, PFG-FST in Korea provides the largest number of indexed mutants in rice through 106,100 FSTs mapped to an RGAP v6 annotation (Jung et al. 2008). A unique feature of this database is the parallel visualization of mutant information from various sources along the genome browser. Visualization of this kind is helpful in the comparative analysis of T-DNA and FSTs for a particular region of the genome. Similarly with RiceGE, 11 mutant resources have been integrated in the OryGenesDB with a total of 245,508 FST sequences (Droc et al. 2006). Out of the 11 integrated resources for RiceGE and OryGenesDB, eight resources are common. Individual mutant resources and their features are summarized in Table 8.

Conclusion and Future Perspectives

The combined effort of research groups across the world has accelerated the rice functional annotation by employing a number of conventional as well as new generation techniques. An overview of the resources for systems biology in rice is shown in Fig. 1. Rice plants with improved stress tolerance coupled with enhanced grain yield are in high demand by the growing population. The large collection of mutant databases provides valuable resources for quick identification of key genes associated with new agronomic traits. The error rate in the rice genome is expected to be curtailed by genome re-sequencing using next generation sequencing technology. Proteomics-based interactive functional networks coupled with co-expression analysis enables biologists to effectively pick up significant elements or reactions for their studies. One centralized resource with high quality data that combines bioinformatics prediction with experimental validation is required. In addition, rather than focusing on an in-depth analysis of a single gene, systems level analyses of many genes can provide important insights.

Acknowledgements

This work was supported in part by the Next Generation Bio-green 21 Program in the Republic of Korea (No. PJ009110 and No. PJ008079 to KHJ), and a Kyung Hee University grant (No. 20100615 to KHJ).

References

Braun P, Aubourg S, Van Leene J, De Jaeger G, Lurin C (2013) Plant protein interactomes. Annu Rev Plant Biol 64:161–187

Cao P, Jung KH, Choi D, Hwang D, Zhu J, Ronald PC (2012) The rice oligonucleotide array database: an atlas of rice gene expression. Rice 5:17

Chang WC, Lee TY, Huang HD, Huang HY, Pan RL (2008) PlantPAN: Plant promoter analysis navigator, for identifying combinatorial cis-regulatory elements with distance constraint in plant gene

groups. BMC Genomics 9:561
Chang Y, Long T, Wu C (2012) Effort and contribution of T-DNA Insertion mutant library for rice functional genomics research in China: review and perspective. J Integr Plant Biol 54:953–966
Croft D, O'Kelly G, Wu G, Haw R, Gillespie M, Matthews L, Caudy M, Garapati P, Gopinath G, Jassal B, Jupe S, Kalatskaya I, Mahajan S, May B, Ndegwa N, Schmidt E, Shamovsky V, Yung C, Birney E, Hermjakob H, D'Eustachio P, Stein L (2011) Reactome: a database of reactions, pathways and biological processes. Nucleic Acids Res 39:D691–697
Dash S, Van Hemert J, Hong L, Wise RP, Dickerson JA (2012) PLEXdb: gene expression resources for plants and plant pathogens. Nucleic Acids Res 40:D1194–1201
Degenkolbe T, Do PT, Zuther E, Repsilber D, Walther D, Hincha DK, Köhl KI (2009) Expression profiling of rice cultivars differing in their tolerance to long-term drought stress. Plant Mol Biol 69:133–153
Dharmawardhana P, Ren L, Amarasinghe V, Monaco M, Thomason J, Ravenscroft D, McCouch S, Ware D, Jaiswal P (2013) A genome scale metabolic network for rice and accompanying analysis of tryptophan, auxin and serotonin biosynthesis regulation under biotic stress. Rice 6:15
Ding X, Richter T, Chen M, Fujii H, Seo YS, Xie M, Zheng X, Kanrar S, Stevenson RA, Dardick C, Li Y, Jiang H, Zhang Y, Yu F, Bartley LE, Chern M, Bart R, Chen X, Zhu L, Farmerie WG, Gribskov M, Zhu JK, Fromm ME, Ronald PC, Song WY (2009) A rice kinase-protein interaction map. Plant Physiol 149:1478–1492
Goffard N, Weiller G (2006) Extending MapMan: application to legume genome arrays. Bioinformatics 22:2958–2959
Gu H, Zhu P, Jiao Y, Meng Y, Chen M (2011) PRIN: a predicted rice interactome network. BMC Bioinformatics 12:161
Jung KH, Ko HJ, Nguyen MX, Kim SR, Ronald P, An G (2013) Genome-wide identification and analysis of early heat stress responsive genes in rice. J Plant Biol 55:458–468
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
Jung KH, Jeon JS, An G (2011) Web tools for rice transcriptome analyses. J Plant Biol 54:65–80
Kent WJ (2002) BLAT---The BLAST-like alignment tool. Genome Res 12:656–664
Kosuge T, Mashima J, Kodama Y, Fujisawa T, Kaminuma E, Ogasawara O, Okubo K, Takagi T, Nakamura Y (2013) DDBJ progress report: a new submission system for leading to a correct annotation. Nucleic Acids Res 42:D44–49
Krishnan A, Guiderdoni E, An G, Hsing YI, Han CD, Lee MC, Yu SM, Upadhyaya N, Ramachandran S, Zhang Q, Sundaresan V, Hirochika H, Leung H, Pereira A (2009) Mutant resources in rice for functional genomics of the grasses. Plant Physiol 149:165–170
Kudo T, Akiyama K, Kojima M, Makita N, Sakurai T, Sakakibara H (2013) UniVIO: a multiple omics database with hormonome and transcriptome data from rice. Plant Cell Physiol 54:e9
Lakshmanan M, Zhang Z, Mohanty B, Kwon JY, Choi HY, Nam HJ, Kim DI, Lee DY (2013) Elucidating rice cell metabolism under flooding and drought stresses using flux-based modeling and analysis. Plant Physiol 162:2140–2150
Lee I, Seo YS, Coltrane D, Hwang S, Oh T, Marcotte EM, Ronald PC (2011) Genetic dissection of the biotic stress response using a genome-scale gene network for rice. Proc Natl Acad Sci USA 108:18548–18553
Leinonen R, Sugawara H, Shumway M, International Nucleotide Sequence Database C (2011) The sequence read archive. Nucleic Acids Res 39:D19–21
Lenka SK, Lohia B, Kumar A, Chinnusamy V, Bansal KC (2009) Genome-wide targeted prediction of ABA responsive genes in rice based on over-represented cis-motif in co-expressed genes. Plant Mol Biol 69:261–271
Lu T, Lu G, Fan D, Zhu C, Li W, Zhao Q, Feng Q, Zhao Y, Guo Y, Li W, Huang X, Han B (2010) Function annotation of the rice transcriptome at single-nucleotide resolution by RNA-seq. Genome Res 20:1238–1249
Mao D, Chen C (2012) Colinearity and similar expression pattern of rice DREB1s reveal their functional conservation in the cold-responsive pathway. PLoS One 7:e47275
Matthews LR, Vaglio P, Reboul J, Ge H, Davis BP, Garrels J, Vincent S, Vidal M (2001) Identification of potential interaction networks using sequence-based searches for conserved protein-protein interactions or "interologs". Genome Res 11:2120–2126
Mochida K, Yoshida T, Sakurai T, Ogihara Y, Shinozaki K (2009) TriFLDB: a database of clustered full-length coding sequences from Triticeae with applications to comparative grass genomics. Plant Physiol 150:1135–1146
Morris RT, O'Connor TR, Wyrick JJ (2008) Osiris: an integrated promoter database for *Oryza sativa* L. Bioinformatics 24:2915–2917
Morsy M, Gouthu S, Orchard S, Thorneycroft D, Harper JF, Mittler R, Cushman JC (2008) Charting plant interactomes: possibilities and challenges. Trends Plant Sci 13:183–191
Obayashi T, Nishida K, Kasahara K, Kinoshita K (2011) ATTED-II updates: condition-specific gene coexpression to extend coexpression analyses and applications to a broad range of flowering plants. Plant Cell Physiol 52:213–219
Quraishi UM, Abrouk M, Bolot S, Pont C, Throude M, Guilhot N, Confolent C, Bortolini F, Praud S, Murigneux A, Charmet G, Salse J (2009) Genomics in cereals: from genome-wide conserved orthologous set (COS) sequences to candidate genes for trait dissection. Funct Integr Genomics 9:473–484
Rotter A, Usadel B, Baebler S, Stitt M, Gruden K (2007) Adaptation of the MapMan ontology to biotic stress responses: application in solanaceous species. Plant Methods 3:10
Rouard M, Guignon V, Aluome C, Laporte MA, Droc G, Walde C, Zmasek CM, Perin C, Conte MG (2011) GreenPhylDB v2.0: comparative and functional genomics in plants. Nucleic Acids Res 39:D1095–1102
Sakai H, Lee SS, Tanaka T, Numa H, Kim J, Kawahara Y, Wakimoto H, Yang CC, Iwamoto M, Abe T, Yamada Y, Muto A, Inokuchi H, Ikemura T, Matsumoto T, Sasaki T, Itoh T (2013) Rice annotation project database (RAP-DB): an integrative and interactive database for rice genomics. Plant Cell Physiol 54:e6
Sato Y, Antonio BA, Namiki N, Takehisa H, Minami H, Kamatsuki K, Sugimoto K, Shimizu Y, Hirochika H, Nagamura Y (2011) RiceXPro: a platform for monitoring gene expression in japonica rice grown under natural field conditions. Nucleic Acids Res 39:D1141–1148
Sato Y, Takehisa H, Kamatsuki K, Minami H, Namiki N, Ikawa H, Ohyanagi H, Sugimoto K, Antonio BA, Nagamura Y (2013) RiceXPro version 3.0: expanding the informatics resource for rice transcriptome. Nucleic Acids Res 41:D1206–1213
Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, Doerks T, Stark M, Muller J, Bork P, Jensen LJ, von Mering C (2011) The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Res 39:D561–568
Usadel B, Poree F, Nagel A, Lohse M, Czedik-Eysenberg A, Stitt M (2009) A guide to using MapMan to visualize and compare Omics data in plants: a case study in the crop species, Maize. Plant Cell Environ 32:1211–1229
Yamamoto E, Yonemaru J, Yamamoto T, Yano M (2012) OGRO: The Overview of functionally characterized genes in rice online database. Rice 5:26

Yamamoto YY, Obokata J (2008) ppdb: a plant promoter database. Nucleic Acids Res 36:D977–981

Yim WC, Yu Y, Song K, Jang CS, Lee BM (2013) PLANEX: the plant co-expression database. BMC Plant Biol 13:83

Youens-Clark K, Buckler E, Casstevens T, Chen C, Declerck G, Derwent P, Dharmawardhana P, Jaiswal P, Kersey P, Karthikeyan AS, Lu J, McCouch SR, Ren L, Spooner W, Stein JC, Thomason J, Wei S, Ware D (2011) Gramene database in 2010: updates and extensions. Nucleic Acids Res 39:D1085–1094

Zhang Q, Li J, Xue Y, Han B, Deng XW (2008) Rice 2020: a call for an international coordinated effort in rice functional genomics. Mol Plant 1:715–719

Zimmermann P, Laule O, Schmitz J, Hruz T, Bleuler S, Gruissem W (2008) Genevestigator transcriptome meta-analysis and biomarker search using rice and barley gene expression databases. Mol Plant 1:851–857