Jung *et al. Rice* (2015) 8:26
DOI 10.1186/s12284-015-0060-7

Rice
a SpringerOpen Journal

REVIEW Open Access

Phylogenomics databases for facilitating functional genomics in rice

Ki-Hong Jung[1*], Peijian Cao[2], Rita Sharma[3], Rashmi Jain[4] and Pamela C Ronald[4,5*]

Abstract

The completion of whole genome sequence of rice (*Oryza sativa*) has significantly accelerated functional genomics studies. Prior to the release of the sequence, only a few genes were assigned a function each year. Since sequencing was completed in 2005, the rate has exponentially increased. As of 2014, 1,021 genes have been described and added to the collection at The Overview of functionally characterized Genes in Rice online database (OGRO). Despite this progress, that number is still very low compared with the total number of genes estimated in the rice genome. One limitation to progress is the presence of functional redundancy among members of the same rice gene family, which covers 51.6 % of all non-transposable element-encoding genes. There remain a significant portion or rice genes that are not functionally redundant, as reflected in the recovery of loss-of-function mutants. To more accurately analyze functional redundancy in the rice genome, we have developed a phylogenomics databases for six large gene families in rice, including those for glycosyltransferases, glycoside hydrolases, kinases, transcription factors, transporters, and cytochrome P450 monooxygenases. In this review, we introduce key features and applications of these databases. We expect that they will serve as a very useful guide in the post-genomics era of research.

Review

Functional redundancy remains a large obstacle in functional genomics studies

Progress in functional genomics studies can be significantly inhibited by functional redundancy existing within a genome. This redundancy can be identified through analysis of genome sequences and transcriptomic data. The whole-genome rice sequence, completed by the International Rice Genome Sequencing Project (IRGSP) consortium, indicates the presence of up to as many as 3,865 paralogous protein families in rice (IRGSP 2005). These include 21,998 proteins out of 42,653 total non-transposable element (non-TE)-related proteins predicted by the Michigan State University Rice Genome Annotation Project (MSU-RGAP; http://rice.plantbiology.msu.edu/) team (Lin et al. 2008). This suggests that a gene within the rice genome has a 51.6 % possibility of being functionally redundant.

Several plant gene family databases are available for such analyses such as GreenPhyl V4 (http://www.greenphyl.org/cgi-bin/index.cgi) (Rouard et al. 2011), Plant Gene Family Database (Sakai et al. 2013), and SALAD Database (http://salad.dna.affrc.go.jp/salad/en/) (Mihara et al. 2010). These databases provide tools for phylogenetic analysis and have been applied to determine the similarity of assigned functions of gene families. However, the presence of predicted genetic redundancy based on intra-family examinations does not always accurately predict functional redundancy. For example, a genome-wide survey of predicted light-responsive genes in rice and functional analysis of T-DNA insertional mutants revealed that four of the tested family members have defects associated with normal growth or chlorophyll biosynthesis (Jung et al. 2008). These genes are highly expressed at leaf tissue. These results suggest that such highly expressed members of a gene family are good targets for functional investigations.

To further explore this idea (Jung et al. 2015), we analyzed the phylogenetic relationship and expression patterns of members of 79 gene families with known function. Of these 79 gene families, 65 carry at least one member that is highly expressed. We found that the

* Correspondence: khjung2010@khu.ac.kr; pcronald@ucdavis.edu
[1]Graduate School of Biotechnology & Crop Biotech Institute, Kyung Hee University, Yongin 446-701, Republic of Korea
[4]Department of Plant Pathology and the Genome Center, University of California, Davis, California 95616, USA
Full list of author information is available at the end of the article

redundancy of these families was limited to two or three members of each family. This study confirmed that phylogenomics analysis integrating gene expression data within a phylogenetic context is an effective strategy to select genes for functional genomics studies.

Construction of phylogenomics databases for six large gene families in rice

In rice, phylogenomics databases for six large gene families, i.e., those for kinase, glycosyltransferases (GTs), glycoside hydrolases (GHs), transcription factors (TFs), transporters, and cytochrome p450 monooxygenases (P450s), have been constructed. A description of three of these databases (kinase, GT, and GH) has been reported (Cao et al. 2008; Dardick et al. 2007; Jung et al. 2010; Sharma et al. 2013). Although the other databases have not yet been fully described, some features of those databases are now publicly available. Figure 1 shows the homepage for the rice phylogenomics database s (http://ricephylogenomics.ucdavis.edu/) plus snapshots of Tree viewer webpages for all six databases.

Rice Kinase Database

The Rice Kinase Database (RKD) was created to host various functional genomics data within the context of a phylogenic tree for individual subfamilies. The RKD provides a platform to integrate disparate data sets on such a tree. Those sets include ten types of information (Table 1). Of these, elements that are key to functional genomics studies are the protein–protein interaction data and meta-expression data based on microarray data analyses. The former has 378 interactions revealed by Y2H as well as 254 distinct kinase interactors and 364 interactions via TAP assays. This information is unique to RKD. The latter provides information about 1,867 samples from 105 data series available from the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) or the plexDB (http://www.plexdb.org) (Barrett et al. 2011; Dash et al. 2012). Therefore, users can easily estimate the functional redundancy between closely linked family members for the selected kinase. Tree Viewer is the main web tool for integrated analysis; part or all of the listed data for a selected gene family can be seen in a phylogenetic

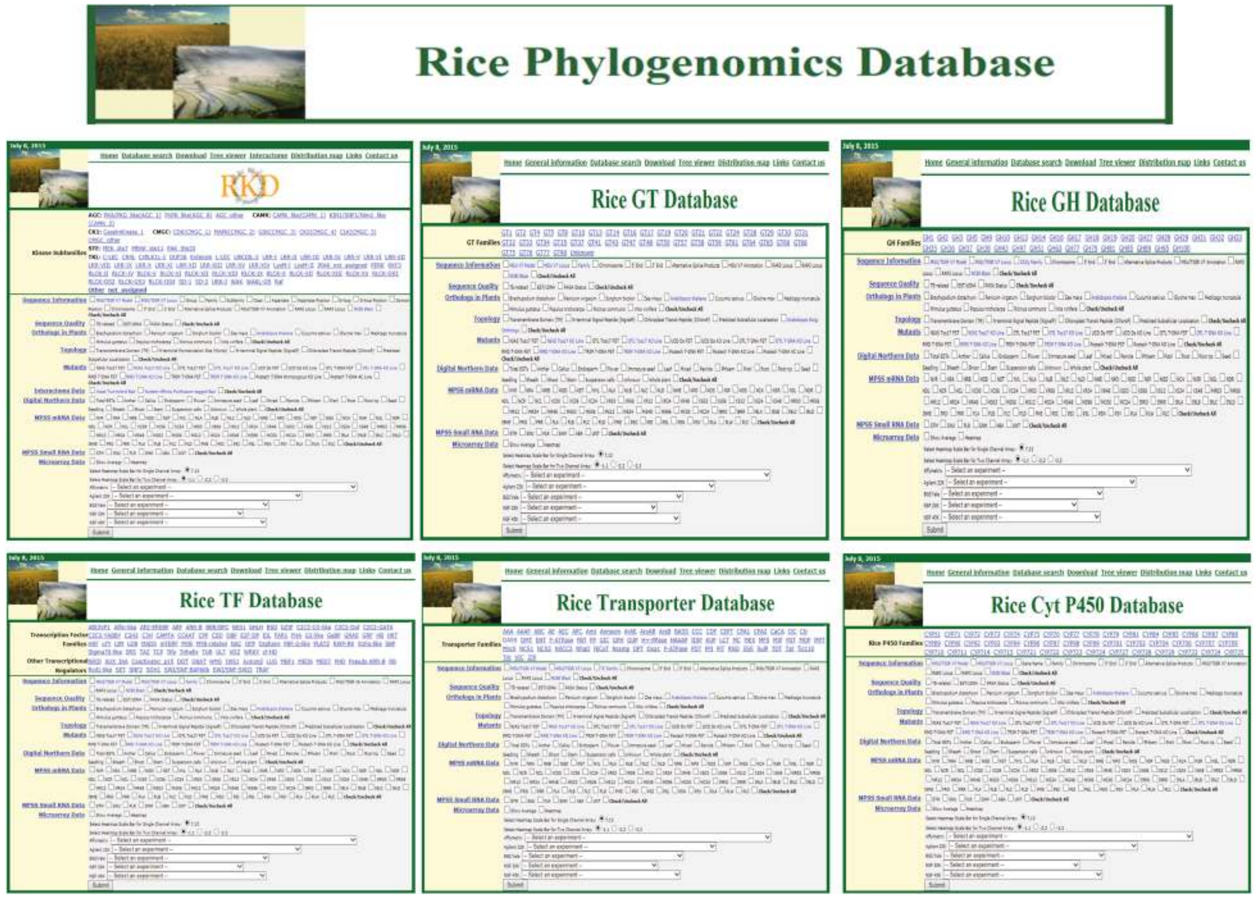


Fig. 1 Snapshots from homepage for rice phylogenomics databases and Tree viewer sites for six phylogenomics databases

Table 1 Summary of the types of data integrated in phylogenomics databases

Data type	Providing information or data	Reference
Sequence	locus IDs from MSU-RGAP and The Rice Annotation Project Database (RAP-DB; http://rapdb.dna.affrc.go.jp/), family and sub-family names, domain positions, NCBI blast result	(IRGSP 2005; Yuan et al. 2005)
Sequence Quality	TE-relatedness, existence of EST/cDNA, and Program to Assemble Spliced Alignments (PASA) status	
Orthologs in Plants	orthologs from 12 plant species, (i.e., *Brachypodium distachyon*, *Panicum virgatum*, *Sorghum bicolor*, *Zea mays*, *Arabidopsis thaliana*, *Cucumis sativus*, *Glycine max*, *Medicago truncatula*, *Mimulus guttatus*, *Populus trichocarpa*, *Ricinus communis*, and *Vitis vinifera*)	(Berglund et al. 2008)
Topology	Transmembrane Domain (TM), N-terminal Myristoylation Site (Myrist), N-terminal Signal Peptide (SignalP), Chloroplast Transit Peptide (ChloroP), and predicted Subcellular Localization	
Mutants	mutant lines and corresponding flanking sequence tags from eight institutes	(Chandran and Jung 2014)
Interactome Data	experimentally validated network of protein–protein interactions based on Yeast Two-Hybrid (Y2H) and Tandem Affinity Purification (TAP) methods	(Ding et al. 2009)
Digital Northern Data	normalized frequency of ESTs in selected tissues/organs	(Dardick et al. 2007; Jung et al. 2010)
MPSS mRNA Data	meta-expression data from 70 libraries	(Nakano et al. 2006)
MPSS Small RNA Data	meta-expression data from six libraries	(Nakano et al. 2006)
Microarray Data	meta-expression data from the six microarray platforms including Affymetrix, Agilent22K, Agilent44K, BGI/YALE60K, NSF20K, and NSF45K (http://ricephylogenomics.ucdavis.edu/description.shtml)	(Cao et al. 2012)

context (Fig. 2). The RKD also includes an interactive chromosomal map showing the positions of all rice kinases and interactive protein–protein interaction maps. Links are provided to navigate around those pages and also to MSU-RGAP. This format simplifies comparisons of closely-related kinases within subfamilies. The RKD has been updated to include rice genome annotations in MSU-RGAP *ver* 7. In that version, 1,467 kinase genes (loci) correspond to 1,934 transcripts (gene models).

Rice GT database

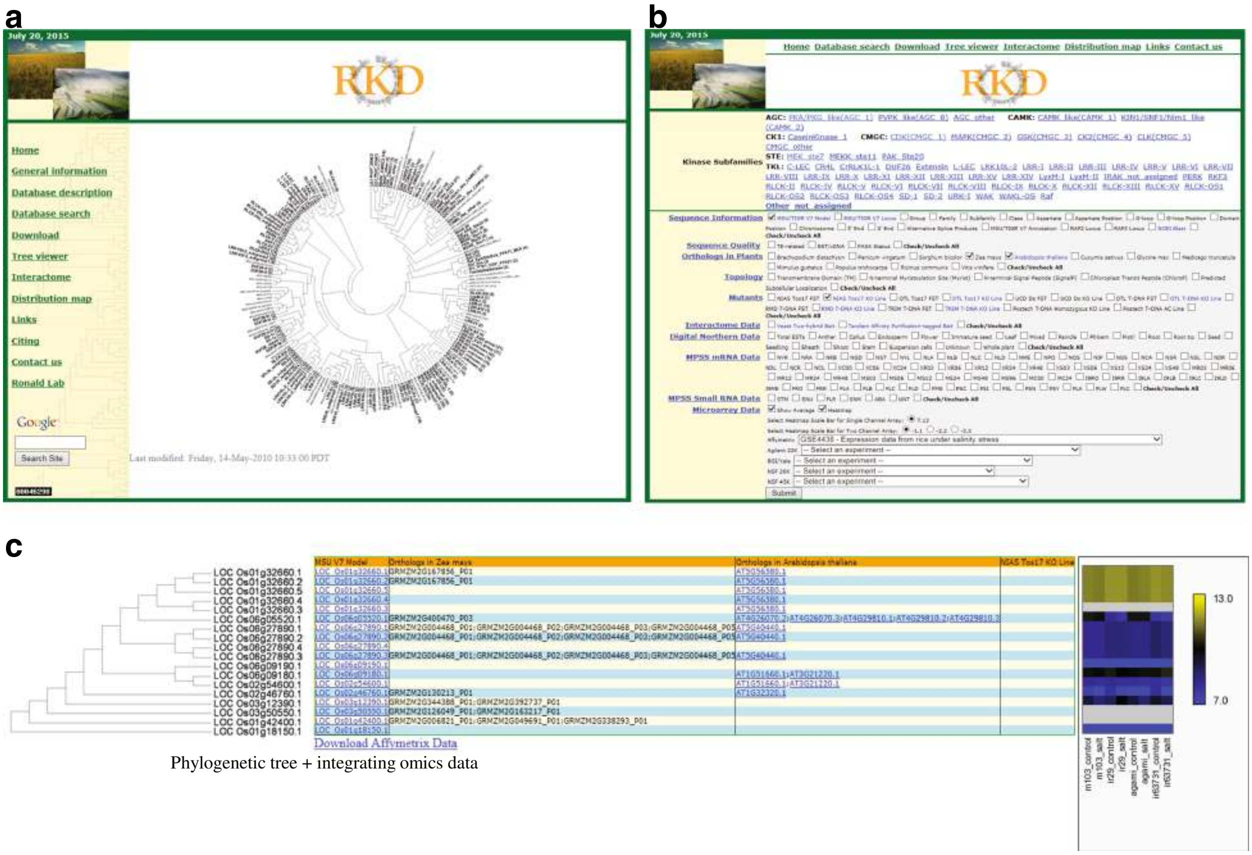
Glycosyltransferases constitute a large group of enzymes that form glycosidic bonds through the transfer of sugars from activated donor molecules to acceptor molecules. They are critical to the biosynthesis of plant cell walls. The Rice GT Database was created to integrate and host functional genomics information for all putative rice GTs (Cao et al. 2008). This database contains information about 609 potential GT genes (loci) that correspond to 769 transcripts (gene models). Those loci have been identified from the rice genome through similarity searches that utilized GT sequences available from the Carbohydrate Active enZymes (CAZy) database (http://www.cazy.org/) (Egelund et al. 2004). Based on domain compositions and sequence similarities, we have classified rice GTs into 41 CAZy families, including one unknown class. Following analysis with Inparanoid, we can suggest that 282 'rice-diverged' GTs have no orthologs in sequenced dicot species (e.g., *A. thaliana*, *P. trichocarpa*, *M. truncatula*, and *R. communis*) (Sonnhammer and Ostlund 2014). Similar to the RKD, we have developed a platform to display user-selected functional genomics data on a phylogenetic tree. These include all integrated data except interactome data (http://ricephylogenomics.ucdavis.edu/cellwalls/gt/).

Rice GH database

Glycoside hydrolases (GHs) catalyze the hydrolysis of glycosidic bonds in cell wall polymers and, along with GTs, are major contributors to plant cell architecture (Sharma et al. 2013). Several GHs have been identified from the rice genome based on sequence similarity searches that used GH sequences in the CAZy database. The rice genome encodes 437 GH genes corresponding to 614 gene models that have been classified into 34 families. Using the massive datasets available in public databases, we have created a phylogenomics database of rice GHs (http://ricephylogenomics.ucdavis.edu/cellwalls/gh/) that integrates multiple data types. The new sets incorporate structural features, orthologous relationships, mutant availability, and gene expression patterns for each GH family within a phylogenomics context (Sharma et al. 2013). After comparing them with dicot GHs, we believe that 138 GH genes are possibly monocot-diverged. By integrating and analyzing these phylogenetic and expression data, researchers should be able to identify potential targets for engineering cell wall structure and stress tolerance. Other features of the GH database are similar to those of the GT database.

Rice TF database

A transcription factor binding to specific DNA sequences controls the rate of transcription of genetic


a
July 20, 2015
RKD
Home
General information
Database description
Database search
Download
Tree viewer
Interactome
Distribution map
Links
Citing
Contact us
Ronald Lab
Google
Search Site
Last modified: Friday, 14-May-2010 10:33:00 PDT
b
July 20, 2015
Home Database search Download Tree viewer Interactome Distribution map Links Contact us
RKD
Kinase Subfamilies
AGC: PKA/PKG like(AGC 1) PVPK like(AGC 8) AGC other CAMK: CAMK like(CAMK 1) KIN1/SNF1/Nim1 like (CAMK 2)
CK1: CaseinKinase 1 CMGC: CDK(CMGC 1) MAPK(CMGC 2) GSK(CMGC 3) CK2(CMGC 4) CLK(CMGC 5) CMGC other
STE: MEK ste7 MEKK ste11 PAK Ste20
TKL: C-LEC CR4L CrRLK1L-1 DUF26 Extensin L-LEC LRK10L-2 LRR-I LRR-II LRR-III LRR-IV LRR-V LRR-VI LRR-VII LRR-VIII LRR-IX LRR-X LRR-XI LRR-XII LRR-XIII LRR-XV LRR-XIV LysM-I LysM-II IRAK not assigned PERK RKF3 RLCK-II RLCK-IV RLCK-V RLCK-VI RLCK-VII RLCK-VIII RLCK-IX RLCK-X RLCK-XII RLCK-XIII RLCK-XV RLCK-OS1 RLCK-OS2 RLCK-OS3 RLCK-OS4 SD-1 SD-2 URK-I WAK WAKL-OS Raf
Other not assigned
Sequence Information
Sequence Quality
Orthologs in Plants
Topology
Mutants
Interactome Data
Digital Northern Data
MPSS mRNA Data
MPSS Small RNA Data
Microarray Data
GSE4438 - Expression data from rice under salinity stress
-- Select an experiment --
Submit
c
LOC Os01g32660.1
LOC Os01g32660.2
LOC Os01g32660.5
LOC Os01g32660.4
LOC Os01g32660.3
LOC Os06g05520.1
LOC Os06g27890.1
LOC Os06g27890.2
LOC Os06g27890.4
LOC Os06g27890.3
LOC Os06g09190.1
LOC Os06g09180.1
LOC Os02g54600.1
LOC Os02g46760.1
LOC Os03g12390.1
LOC Os03g50550.1
LOC Os01g42400.1
LOC Os01g18150.1
13.0
7.0
m103_control
m103_salt
ir29_control
ir29_salt
agami_control
agami_salt
ir63731_control
ir63731_salt

MSU V7 Model	Orthologs in Zea mays	Orthologs in Arabidopsis thaliana	NIAS Tos17 KO Line
LOC_Os01g32660.1	GRMZM2G167856_P01	AT5G56380.1	
LOC_Os01g32660.2	GRMZM2G167856_P01	AT5G56380.1	
LOC_Os01g32660.5		AT5G56380.1	
LOC_Os01g32660.4		AT5G56380.1	
LOC_Os01g32660.3		AT5G56380.1	
LOC_Os06g05520.1	GRMZM2G400470_P03	AT4G26070.2;AT4G26070.3;AT4G29810.1;AT4G29810.2;AT4G29810.3	
LOC_Os06g27890.1	GRMZM2G004468_P01;GRMZM2G004468_P02;GRMZM2G004468_P03;GRMZM2G004468_P05	AT5G40440.1	
LOC_Os06g27890.2	GRMZM2G004468_P01;GRMZM2G004468_P02;GRMZM2G004468_P03;GRMZM2G004468_P05	AT5G40440.1	
LOC_Os06g27890.4			
LOC_Os06g27890.3	GRMZM2G004468_P01;GRMZM2G004468_P02;GRMZM2G004468_P03;GRMZM2G004468_P05	AT5G40440.1	
LOC_Os06g09190.1			
LOC_Os06g09180.1		AT1G51660.1;AT3G21220.1	
LOC_Os02g54600.1		AT1G51660.1;AT3G21220.1	
LOC_Os02g46760.1	GRMZM2G130213_P01	AT1G32320.1	
LOC_Os03g12390.1	GRMZM2G344388_P01;GRMZM2G392737_P01		
LOC_Os03g50550.1	GRMZM2G126049_P01;GRMZM2G163217_P01		
LOC_Os01g42400.1	GRMZM2G006821_P01;GRMZM2G049691_P01;GRMZM2G338293_P01		
LOC_Os01g18150.1			

Download Affymetrix Data

Phylogenetic tree + integrating omics data

Fig. 2 Snapshots from homepage for rice kinase database showing phylogenomics data for cyclin dependent kinase (CDK) subfamily using Tree viewer tool. At Rice kinase database (RKD) homepage (**a**), we clicked Tree viewer option and moved to RKD Tree viewer website (**b**). This website consists of two parts: Kinase subfamilies and integrating data sections. After then, we selected CDK CMGC_1) from kinase subfamilies, MSU/TIGR V7 Model from sequence information, Zea mays and Arabidopsis thaliana from orthologs in plants, transmembrane domain from topology, NIAS Tos17 KO line from mutants, and show average and heatmap of Affymetrix GSE4438 dataseries (**c**). The scale bar for single channel array data from 7 to 13 log2 transformed intensity value was applied. Yellow color indicates the high level of expression and blue one indicates the low level

information from DNA to messenger RNA (Todeschini et al. 2014). Rice TFs have been retrieved from the Plant Transcription Factor Database (http://plntfdb.bio.uni-potsdam.de/v3.0/) (Zhang et al. 2011). This Rice TF Database (http://ricephylogenomics.ucdavis.edu/tf/) hosts 2,385 genes corresponding to 3,119 models classified into 80 families. It integrates and provides functional genomics information for all putative rice TFs and other predicted transcriptional regulators. Like other databases, we have integrated multiple data types, such as structural features, orthologous relationships, mutant availability, and gene expression patterns for each TF family within a phylogenomics context. Other features are similar to those of the GT database.

Rice transporter database

A transporter is a membrane protein involved in the movement of ions or small molecules (Saier et al. 2014). Transporter proteins exist permanently within and span the membrane across which substances are transferred. The rice genome contains 1,211 potential transporter genes (loci) corresponding to 1754 gene models (Ren et al. 2007). These sequences have been retrieved from the Transporter Protein Analysis Database (TransportDB; http://www.membranetransport.org/), which was created to merge and provide functional genomics information for all putative rice transporters. Like for the other databases, we have integrated multiple data types that include structural features, orthologous relationships, mutant availability, and gene expression patterns for each transporter family (http://ricephylogenomics.ucdavis.edu/transporter/). Other features are similar to those of the GT database.

Rice Cytochrome P450 database

Cytochrome P450 monooxygenases belong to the superfamily of proteins containing a heme cofactor. They have roles in the terminal oxidation of electron transfer chains. The rice genome has 302 genes that encode

P450s corresponding to 341 transcripts. Rice P450s have been retrieved from the Cytochrome P450 Database (http://drnelson.uthsc.edu/CytochromeP450.html), and mapped onto the MSU-RGAP *ver* 6 genome annotation. The Rice P450 Database (http://ricephylogenomics.ucdavis.edu/p450/) was created to integrate and provide functional genomics information for all putative rice P450s. As with the other databases, we have integrated multiple data types for each P450 family. The other features are similar to those of the GT database.

Applying a phylogenomics approach to estimate functional redundancy within a gene family

To demonstrate the functionality of the phylogenomics databases, we obtained relevant information, including anatomical meta-expression data in a phylogenetic context, for the aquaporin gene family (Fig. 3) (Nguyen et al. 2013). The functions of five aquaporin gene members have been characterized via loss-of-function studies. Three of the aquaporin gene family members belong to the Nodulin-26 like Intrinsic Protein (NIP) subfamily and two are in the Plasma membrane Intrinsic Protein (PIP) subfamily. One gene, *Low silicon rice1* (*Lsi1*) (*LOC_Os02g51110*), in the NIP subfamily, has been demonstrated to play a role in Si uptake in roots. Anatomical meta-expression data indicate that NIPs show root-preferential patterns of expression that are well-matched with their biological functions. For example, Lsi6 is closely linked to Lsi1 and is involved in Si translocation from roots to shoots. Anatomical meta-expression data reveal that Lsi6 is highly expressed in the root, shoot, leaf, and palea/lemma tissues. Integrated meta-expression data confirms previous reports of distinct functional roles for *Lsi1* and *Lsi6*.

OsNIP3;1 (LOC_Os10g33924) is a boron transporter. P Plant growth cannot be sustained under boron-deficient conditions when expression of that gene is knocked down using RNAi. A phylogenetic tree has

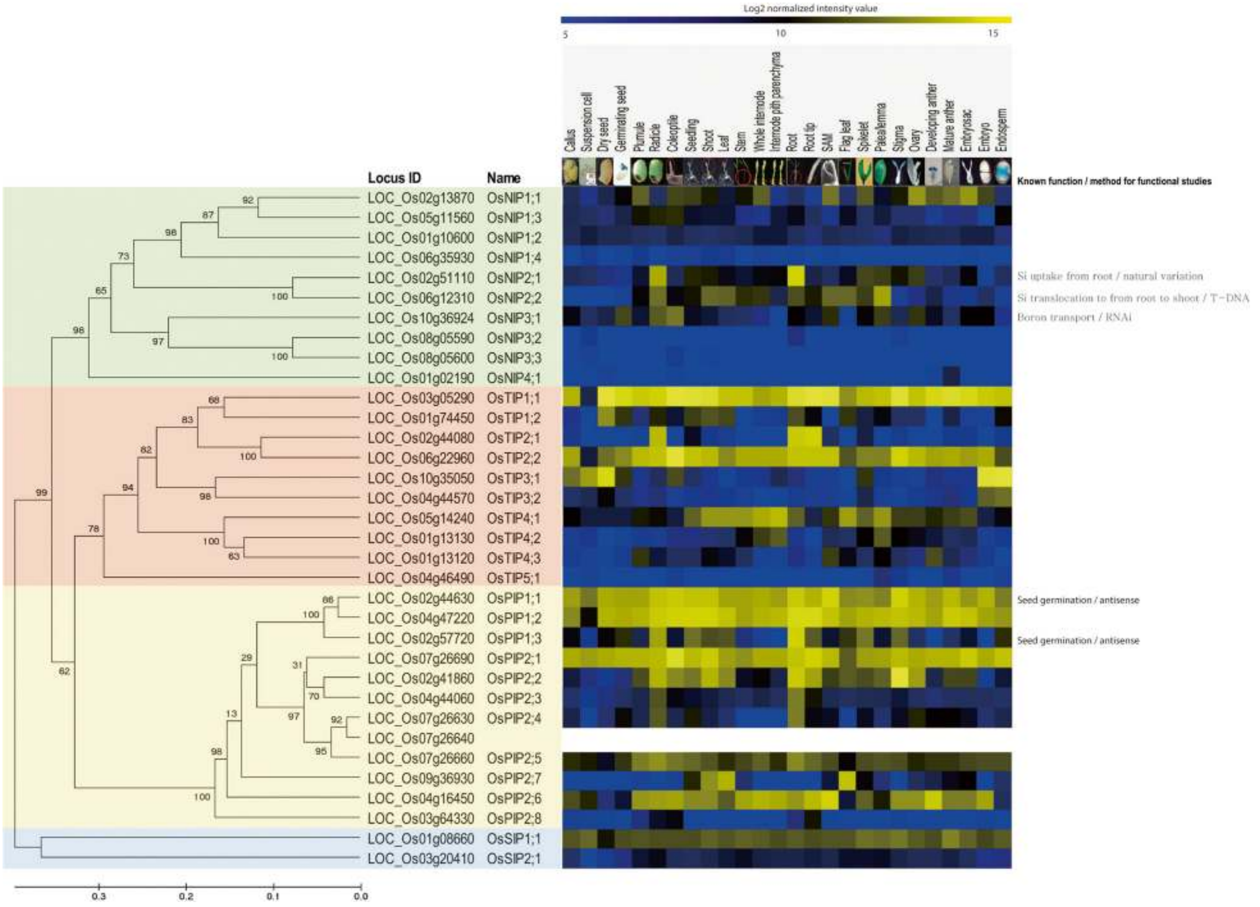


Fig. 3 Case study to evaluate the functionality of phylogenomics tool: Phylogenomics analysis of aquaporin family in rice. Anatomical expression data were integrated into phylogenetic tree context of aquaporin family. Anatomical meta-expression data were generated from ROAD with 34 rice aquaporin genes. Their protein classifications revealed 20 NIPs (green boxes), 12 PIPs (pale-yellow boxes), 10 tonoplast intrinsic proteins (pink boxes), and 8 small basic intrinsic proteins (light-blue boxes). For heatmaps, blue color denotes low expression, black medium level expression and yellow indicates high expression. Numbers indicate estimated evolutionary distances

indicated that OsNIP3;1, OsNIP3;2, and OsNIP3;3 cluster together. *OsNIP3;1* is the dominantly expressed gene family member (Fig. 3) (Liu et al. 2007). Both *OsPIP1;2* and *OsPIP1;1* are ubiquitously expressed suggesting functional redundancy between the two. Antisense suppression of *OsPIP1;1* causes partially defective phenotypes during seed germination, but overexpression of that gene does not stimulate the occurrence of a more normal phenotype. Even though OsPIP1;3 is closely related to OsPIP1;1 and OsPIP1;2, meta-expression data indicates that *OsPIP1;3* expression is highest in the radical and root. Rice plants that are silenced for *OsPIP1;3* are defective in seed germination whereas overexpression of *OsPIP1;3* enhances germination (Liu et al. 2007). These results demonstrate that *OsPIP1;3* has a major role in seed germination in contrast to the other genes in that clade. These observations confirm that phylogenomic analysis, that integrates global expression data with phylogenetic analysis, is a useful method for identifying distinct roles for closely related gene family members.

Conclusion and Prospect

Functional genomics studies of rice genes belonging to a single family can be facilitated by phylogenomic analysis that integrates diverse types of biological information in a single page view. Future prospects to advance analysis of gene function include coupling phylogenomic analyses with computational predictions of gene function. For example, we have recently generated a probabilistic functional gene network for rice, called RiceNet (Lee et al., 2011; Lee et al., 2015). We used the GH database phylogenomics database to identify 17 GH glycoside hydrolase gene family members (Sharma et al. 2013). We used these seventeen candidate GH genes to query RiceNet v1. We found that these nine genes are highly predicted to function in the same biological process as cellulose synthase and cellulose synthase-like genes of rice, suggesting a potential role for these nine *GH* genes in cell wall biosynthesis.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

All authors contributed to the collection of information and writing of this manuscript. All authors read and approved the final manuscript.

Acknowledgement

This work was carried out with the support of "Cooperative Research Program for Agriculture Science & Technology Development (Project title: Global identification and functional study of rice genes for enhancement of root development and nutrient use efficiency using genome information, Project No.PJ01100401)" Rural Development Administration, Republic of Korea, the Ramalingaswami Fellowship from the Department of Biotechnology, Government of India to RS, and funding from The Joint BioEnergy Institute, the Office of Science, Office of Biological and Environmental Research, U.S. Department of Energy under Contract No. DE-AC02-05CH11231 to PCR.

Author details

[1]Graduate School of Biotechnology & Crop Biotech Institute, Kyung Hee University, Yongin 446-701, Republic of Korea. [2]China Tobacco Gene Research Center, Zhengzhou Tobacco Research Institute, Zhengzhou 450001, China. [3]School of Life Sciences, Jawaharlal Nehru University, New Delhi 110067, India. [4]Department of Plant Pathology and the Genome Center, University of California, Davis, California 95616, USA. [5]The Joint Bioenergy Institute, Emeryville, California 95616, USA.

Received: 5 February 2015 Accepted: 14 July 2015
Published online: 30 July 2015

References

Barrett T, Troup DB, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Muertter RN, Holko M, Ayanbule O, Yefanov A, Soboleva A (2011) NCBI GEO: archive for functional genomics data sets–10 years on. Nucleic Acids Res 39:D1005–1010

Cao PJ, Bartley LE, Jung KH, Ronald PC (2008) Construction of a rice glycosyltransferase phylogenomic database and identification of rice-diverged glycosyltransferases. Mol Plant 1:858–877

Cao P, Jung KH, Choi D, Hwang D, Ronald PC (2012) The Rice Oligonucleotide Array Database: an atlas of rice gene expression. Rice 5:17

Chandran AKN, Jung KH (2014) Resources for systems biology in rice. J Plant Biol 57:80–92

Dardick C, Chen J, Richter T, Ouyang S, Ronald P (2007) The rice kinase database. A phylogenomic database for the rice kinome. Plant Physiol 143:579–586

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

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

Egelund J, Skjot M, Geshi N, Ulvskov P, Petersen BL (2004) A complementary bioinformatics approach to identify potential plant cell wall glycosyltransferase-encoding genes. Plant Physiol 136:2609–2620

IRGSP (2005) The map-based sequence of the rice genome. Nature 436:793–800

Jung KH, Lee J, Dardick C, Seo YS, Cao P, Canlas P, Phetsom J, Xu X, Ouyang S, An K, Cho YJ, Lee GC, Lee Y, An G, Ronald PC (2008) Identification and functional analysis of light-responsive unique genes and gene family members in rice. PLoS Genet 4, e1000164

Jung KH, Cao P, Seo YS, Dardick C, Ronald PC (2010) The Rice Kinase Phylogenomics Database: a guide for systematic analysis of the rice kinase super-family. Trends Plant Sci 15:595–599

Jung KH, Kim SR, Giong HK, Nguyen MX, Go HJ, An G (2015) Genome-wide identification and functional analysis of genes expressed ubiquitously in rice. Mol Plant. 8:276-289

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

Lee T, Oh T, Yang S, Shin J, Hwang S, Kim CY, Kim H, Shim H, Shim JE, Ronald PC, Lee I (2015) RiceNet v2: an improved network prioritization server for rice genes. Nucleic Acids Res 43:W122–127.

Lin H, Ouyang S, Egan A, Nobuta K, Haas BJ, Zhu W, Gu X, Silva JC, Meyers BC, Buell CR (2008) Characterization of paralogous protein families in rice. BMC Plant Biol 8:18

Liu HY, Yu X, Cui DY, Sun MH, Sun WN, Tang ZC, Kwak SS, Su WA (2007) The role of water channel proteins and nitric oxide signaling in rice seed germination. Cell Res 17:638–649

Mihara M, Itoh T, Izawa T (2010) SALAD database: a motif-based database of protein annotations for plant comparative genomics. Nucleic Acids Res 38:D835–842

Nakano M, Nobuta K, Vemaraju K, Tej SS, Skogen JW, Meyers BC (2006) Plant MPSS databases: signature-based transcriptional resources for analyses of mRNA and small RNA. Nucleic Acids Res 34:D731–735

Nguyen MX, Moon S, Jung KH (2013) Genome-wide expression analysis of rice aquaporin genes and development of a functional gene network mediated by aquaporin expression in roots. Planta 238:669–681

Ren Q, Chen K, Paulsen IT (2007) TransportDB: a comprehensive database resource for cytoplasmic membrane transport systems and outer membrane channels. Nucleic Acids Res 35:D274–279

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

Saier MH Jr, Reddy VS, Tamang DG, Vastermark A (2014) The transporter classification database. Nucleic Acids Res 42:D251–258

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

Sharma R, Cao P, Jung KH, Sharma MK, Ronald PC (2013) Construction of a rice glycoside hydrolase phylogenomic database and identification of targets for biofuel research. Front Plant Sci 4:330

Sonnhammer EL, Ostlund G (2014) InParanoid 8: orthology analysis between 273 proteomes, mostly eukaryotic. Nucleic Acids Res 43:D234–239

Todeschini AL, Georges A, Veitia RA (2014) Transcription factors: specific DNA binding and specific gene regulation. Trends Genet 30:211–219

Yuan Q, Ouyang S, Wang A, Zhu W, Maiti R, Lin H, Hamilton J, Haas B, Sultana R, Cheung F, Wortman J, Buell CR (2005) The institute for genomic research Osa1 rice genome annotation database. Plant Physiol 138:18–26

Zhang H, Jin J, Tang L, Zhao Y, Gu X, Gao G, Luo J (2011) PlantTFDB 2.0: update and improvement of the comprehensive plant transcription factor database. Nucleic Acids Res 39:D1114–1117