Title
The rice kinase database. A phylogenomic database for the rice kinome.

Permalink
https://escholarship.org/uc/item/8748k0nn

Journal
Plant physiology, 143(2)

ISSN
0032-0889

Authors
Dardick, Christopher
Chen, Johann
Richter, Todd
et al.

Publication Date
2007-02-15

Peer reviewed

The Rice Kinase Database. A Phylogenomic Database for the Rice Kinome[1][OA]

Christopher Dardick, Johann Chen, Todd Richter, Shu Ouyang, and Pamela Ronald*

United States Department of Agriculture-Agricultural Research Service, Appalachian Fruit Research Station, Kearneysville, West Virginia, 25430 (C.D.); Department of Plant Pathology, University of California, Davis, California, 95616 (J.C., T.R., P.R.); and Institute for Genomic Research, Rockville, Maryland, 20853 (S.O.)

The rice (*Oryza sativa*) genome contains 1,429 protein kinases, the vast majority of which have unknown functions. We created a phylogenomic database (http://rkd.ucdavis.edu) to facilitate functional analysis of this large gene family. Sequence and genomic data, including gene expression data and protein-protein interaction maps, can be displayed for each selected kinase in the context of a phylogenetic tree allowing for comparative analysis both within and between large kinase subfamilies. Interaction maps are easily accessed through links and displayed using Cytoscape, an open source software platform. Chromosomal distribution of all rice kinases can also be explored via an interactive interface.

The presence of large gene families in plant and animal genomes, 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. Newer technologies such as RNAi provide enhanced capability to study gene families, as RNAi can be used to knock down multiple genes simultaneously. However, this technology does have practical limitations on the numbers of genes that can be simultaneously silenced and still requires rational selection of gene targets.

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/or multigene families. The scattered nature of genomic data across multiple databases creates additional challenges to data integration.

A new field of study that is at least in part resolving these limitations is phylogenomics. Phylogenomics represents a merger between phylogenetics and genomics and puts genomic data in a phylogenetic context (K. Solander, personal communication). 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 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. Thus, phylogenomic analyses can provide a more logical basis for rational selection of gene candidates for further detailed functional studies.

One family of genes for which redundancy poses enormous challenges is protein kinases. Kinases comprise a highly conserved family of enzymes that control diverse cellular processes and are key components of virtually all biological systems. The high degree of similarity found between even diverse protein kinases and the ability to generate robust phylogenetic groupings makes this gene family an excellent candidate for phylogenomic studies.

Kinases are found both as cytoplasmic proteins and as domains within larger membrane-bound receptors. Sequencing of the rice (*Oryza sativa*) genome has enabled the characterization of the entire complement of rice kinases or kinome. Remarkably, the rice kinome contains 40% more kinases than Arabidopsis (*Arabidopsis thaliana*) and is 3 times larger than the human kinome (Shiu and Bleecker, 2001; Shiu et al., 2004; Dardick and Ronald, 2006). The expansions in rice are

primarily attributed to large multigene families of the membrane-bound receptor type, nearly all of which have unknown functions (Fig. 1). Determining the functions of these genes presents a difficult challenge. Studies in Arabidopsis have shown that single gene knockouts in some receptor kinases produce little or no observable phenotypes, in part due to gene redundancy (Shpak et al., 2004). Phenotypic characterization of many kinase families can be accomplished only after crossing multiple knockouts into a single genetic background. For larger gene families and for those species that are not as genetically tractable as Arabidopsis, such approaches may not be practical. Therefore, alternative methods and associated bioinformatics tools are needed to deduce kinase function(s) and narrow the scope of genetic experiments.

Protein kinases have been classified into seven major phylogenetic groups (Manning et al., 2002). The rice kinome contains members of six of these: PKA, PKG, and PKC kinases (AGC), containing CDK, MAPK, GSK3, and CLK kinases (GMGC), calcium/calmodulin-dependent protein kinases (CAMK), casein kinase 1 (CK1), Tyr kinase like, mixed lineage kinases, transforming growth factor-β receptor kinases, and Raf kinases (TKL), and homologs of yeast (*Saccharomyces cerevisiae*) sterile 7, sterile 11, and sterile 20 kinases (STE). Like Arabidopsis, the rice kinome lacks obvious members of the Tyr kinases group. Seventy-five percent of all rice kinases (1,068) fall into the TKL group that includes the large interleukin-1 receptor-associated kinase (IRAK) family and includes both receptor and cytoplasmic kinases. Within the rice IRAK family, 69 subfamilies have been delineated based on phylogenetic analyses and organization of extracellular domains
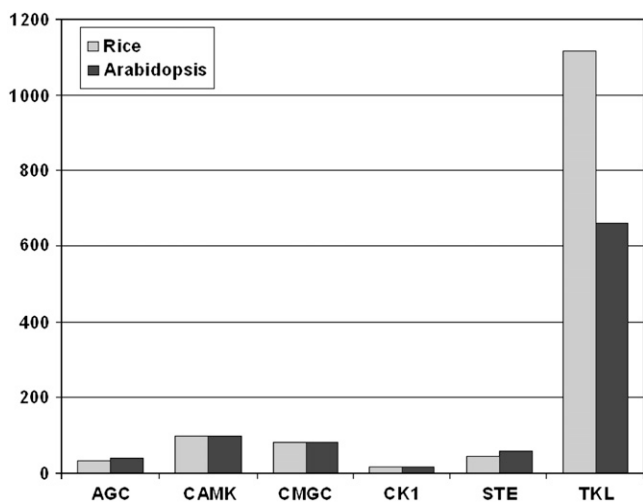
(Shiu et al., 2004; Dardick and Ronald, 2006). Only a small number of these kinases have known functions, including Xa21, Xa26 (Leu-rich repeat [LRR] XII), and Pi-2 d (SD-2b) that function in disease resistance and FON1 (LRR XI) that plays a role in determining floral organ number (Song et al., 1995; Sun et al., 2004; Suzaki et al., 2004; Chen et al., 2006).

Gene expression and proteomic data for rice genes and proteins is growing at an exponential rate. The release of two new microarray platforms, Affymetrix (http://www.affymetrix.com/products/arrays/specific/rice.affx) and the National Science Foundation (NSF)-funded 45 K array (http://www.ricearray.org), should further accelerate data deposition. Likewise, massively parallel signature sequencing (MPSS) data, a powerful method to accurately determine the representation of transcripts within mRNA or regulatory small RNA populations, is also becoming increasingly available. Rice MPSS data is rapidly growing and is currently available for multiple tissues as well as abiotic and biotic stress treatments (http://mpss.udel.edu/rice/; Nakano et al., 2006). An NSF-funded kinase proteomics project has identified interacting partners of 275 rice kinases through yeast two-hybrid (Y2H) screening and in vivo tandem affinity purification (TAP; Rohila et al., 2006; W. Song, unpublished data). This project includes the generation of 75 RNAi knockdown kinase mutants to complement the growing national and international collections of sequenced rice transposon and T-DNA insertional mutants (see http://indica.ucdavis.edu/Links/index.php?page=rgsites for a list of searchable populations). These populations will expedite the identification of mutant phenotypes for individual genes.

Here we report the creation of a publicly available rice kinase phylogenomic database. How the database was constructed and guidelines for its use are described as well as results from an initial global kinase expression analysis.

DATABASE CONSTRUCTION

Kinase Sequences and Phylogram

Representative kinases from six kinase groups (STE, TKL, CMGC, AGC, CK1, and CAMK) were used to perform reiterative TBLASTN searches against three databases: National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/), The Institute for Genomic Research (TIGR; release 2; http://www.tigr.org/tdb/e2k1/osa1/), and Knowledge-based Oryza Molecular biological Encyclopedia full-length cDNA database (http://cdna01.dna.affrc.go.jp/cDNA/; Kikuchi et al., 2003). All high-scoring hits were assembled against the TIGR rice cultivar Nipponbarre genome annotation, release 2, into a single nonredundant set of 1,585 kinases (the RKD kinase dataset has since been converted to TIGR release 3 Os designations). Each kinase sequence was subsequently verified using the KinG catalytic domain search tool

Figure 1. Distribution of rice kinases among the six kinase groups: AGC (PKA, PKG, and PKC kinases), CMGC (containing CDK, MAPK, GSK3, and CLK families), CAMK, CK1, TKL (includes mixed lineage kinases, TGF β receptor kinases, and Raf kinases), and STE (homologs of yeast sterile 7, sterile 11, and sterile 20 kinases). Number of genes is shown on the y axis. Subfamilies are listed along the x axis.

(http://hodgkin.mbu.iisc.ernet.in/~king/). A total of 77 genes lacking identifiable kinase domains or encoding only short kinase fragments were removed, resulting in a final set of 1,508 sequences. Of these 1,508, 79 were identified as duplications derived from alternative splice variants producing a total of 1,429 unique protein kinase encoding genes. His and atypical kinases were not included in this analysis and are thus not represented in the RKD.

Kinase domains identified by the KinG server (http://hodgkin.mbu.iisc.ernet.in/~king/) were used to create a single topological tree. All kinases were aligned using ClustalW (Krupa et al., 2004). Alignments were manually refined using BIOEDIT (IBIS Therapeutics), and a phylogram was constructed using ClustalW. Kinase group and family classifications were determined by comparing each rice kinase to the nearest Arabidopsis kinase via BLASTP. Kinases producing only low scoring hits were labeled as not classified. These classifications were verified and further refined using a second multikingdom phylogram that included the human, fly, worm, Arabidopsis, and rice kinomes (data not shown). For IRAK family kinases, BLAST searches were conducted against the Indica rice kinome previously assembled by Shiu et al.

(2004). Kinase subfamily classifications were then added to both the Japonica rice kinome tree and the multikingdom tree to verify phylogenetic groupings. Several IRAK subfamilies were further deconstructed based the presence or absence of conserved kinase residues (Dardick and Ronald, 2006). Finally, the rice kinome tree was manually colored for easy visualization of all kinase subfamilies. A fragment of the tree as shown in the RKD tree viewer is shown in Figure 2, and the complete tree can be viewed at http://rkd.ucdavis.edu.

Kinase Sequence Data

Sequence information including annotation, chromosome number, position of 5′ and 3′ ends, and source bacterial artificial chromosome clone were provided by TIGR. TIGR also provided information related to sequence and annotation quality, including sequencing status, availability of cDNA or expressed sequence tags, genes with matches to known transposable elements, and whether or not annotation of each sequence has been verified by Program to Assemble Spliced Alignments (http://www.tigr.org/software/genefinding.shtml). Kinase genomic, cDNA,
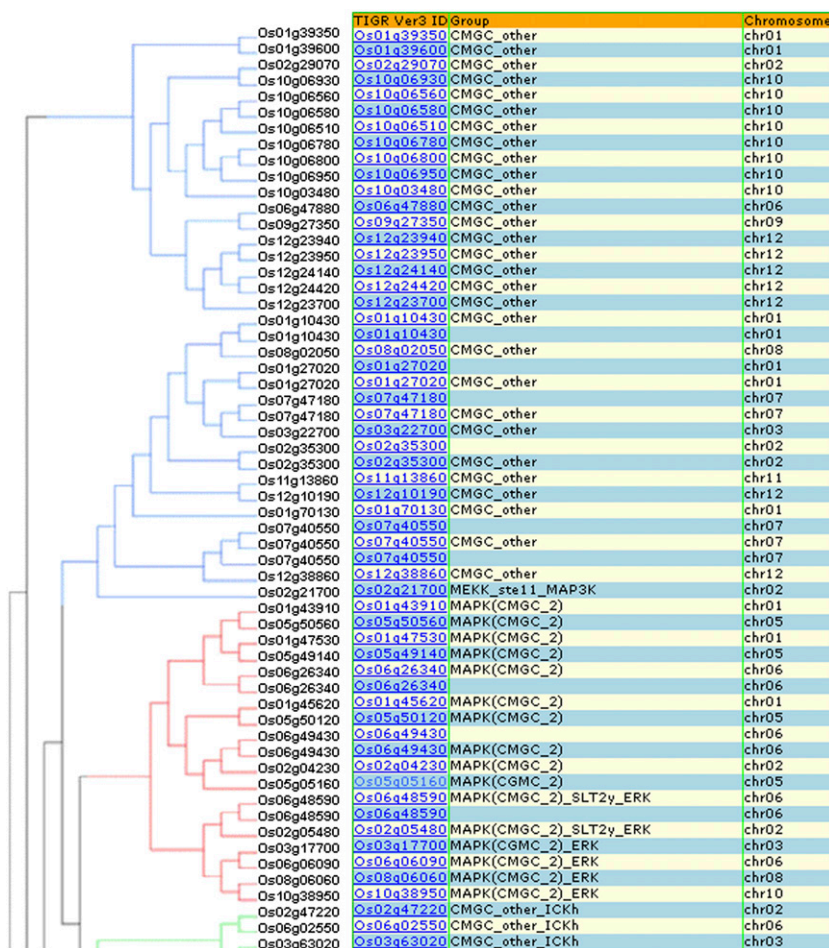

Figure 2. Screen shot of the top portion of the RKD phylogenetic tree. The TIGR model number (gene ID) and group are listed in spreadsheet format adjacent to the tree to allow for easy and flexible visualization.

and protein sequences in FASTA format are also listed and available for download.

Kinase Motifs and Conserved Domains

Conserved kinase motifs, including kinase subdomains I (G-loop), II, III, VI, VII, and VIII, were extracted from global multiple alignments of all rice kinase domains. These motifs contain highly conserved residues required for catalytic activity and can serve as a predictor of kinase function. Kinase topology, including predicted transmembrane domains and *N*-myristoylation sites, as well as predicted functional domains, were provided by Mike Gribskov and are also available in PlantsP (http://plantsp.genomics.purdue.edu/; Tchieu et al., 2003). LRRs, lectin, and PAN/Apple domains are listed separately due to the large number of kinases that contain them.

Chromosome Maps

Kinase chromosomal distributions were determined and visualized using GenomePixelizer (Kozik et al., 2002). Kinases were color coded according to family or subfamily. Colored lines that range from yellow (>60% nucleotide identity) to red (>90% nucleotide identity) are provided to visualize closely related kinases derived from recent duplication and translocation events.

T-DNA and Transposon Insertional Mutants

We searched all available transposon and T-DNA flanking sequences in the National Center for Biotechnology Information for matches to rice kinases using BLASTN. These consist primarily of deposits from two insertional sequencing projects (Miyao et al., 2003;

Sallaud et al., 2004). Sequence identity was further verified by manual inspection; however, due to the presence of large kinase gene families, some matches may represent closely related genes (>98% identity) rather than the target gene listed.

Kinase Expression Data

Expression data for each rice kinase was extracted from the rice MPSS database (http://mpss.udel.edu/rice/; Nakano et al., 2006). Data from 22 separate MPSS libraries were downloaded into the RKD. Some of the libraries include biological replicates for mature leaves (four) and roots (two). Values listed represent transcripts per million (TPM) and were derived from the sum of all 17 bp signatures in classes 1, 2, 5, and 7 (close to or within annotated open reading frames). Data was not available for 529 kinases, which were not represented in the MPSS database.

Nearest Arabidopsis Kinase

Potential Arabidopsis homologs of all rice kinases were identified using BLASTP (http://www.ncbi.nlm.nih.gov/). Arabidopsis identification (ID) numbers for the highest-scoring hit and the associated BLAST *E* value are indicated.

Protein-Protein Interaction Maps

The kinase interaction maps presented on RKD include combined data obtained from an NSF-funded high-throughput Y2H and TAP project. A total of 275 rice kinases were used as baits in the Y2H system to identify interacting proteins (X. Ding, unpublished data). Concurrently, the same 275 rice kinases were

Figure 3. Screen shot of RKD tree viewer format. Checking each box and clicking submit displays the selected data next to the tree.

MPSS Data

NLA	294
NLB	126
NLC	44
NDR	99
NIP	3
NCR	6
NYL	68
NSL	4
NCA	21
NLD	131
NDL	54
NCL	6
NRB	227
NRA	237
NME	34
NSR	198
NGD	37
NYR	116

Sequence Information	Sequence Quality
OS number: Os09g39650	TE-related:
Group: RLCK-IXb	cDNA: fl-cDNA
Chromosome: chr09	Finished/Unfinished: unfinished
5' end: 20841921	PASA Status: PASA_validated
3' end: 20846057	
Expression Data:	
BAC Name: OJ1003_C09	
TIGR TU Name:	
Alternative Splice Products:	
PlantsP ID:	

Conserved Kinase Motifs	Conserved Domains
G-loop: GEGGYG	Leucine Rich Repeat(LRR):
Domain II: VAIK	LECTIN(LEC):
Domain III: E	PAN/Apple:
Domain VI: HAD	All other domains:
Domain VII: DFG	
Domain VIII: DPV	

Knockouts	Topology
T-DNA Knockout:	Transmembrance domain(TM): TM
Tos 17 Knockout: dbjjAG021072	N-myristoylationsite(N-myr):

Interaction Map

Os06g31100 Os06g04640
Os07g37250 Os06g14080
Os10g40100 Os09g39650 Os02g42940
Os05g30950 Os04g45190
Os10g36350 Os08g06020

Interactor Annotation

Os02g42940	vesicle-associated membrane protein
Os04g45190	N/A
Os08g06020	vesicle-associated membrane protein
Os10g36350	N/A
Os05g30950	vesicle-associated membrane protein
Os10g40100	N/A
Os07g37250	expressed protein
Os06g31100	N/A
Os06g04640	N/A
Os06g14080	N/A

Interactor Raw Sequence

```
>Os02g42940        vesicle-associated membrane protein
gaattcggcacgaggCAAAATTCCAACGAGTTCCCCGTTTTGCTCCGCTTCCTCCTCTGC
TCCAGGAATAATATTTTGCTGGAGAATAGTCGGATGCAGGAGGCGAAAAAAGATTCGATC
TTTGGAGGCCTTGCCCTCTGGGGCGCCCGTCGGAGGCGGATCGCTTCCCGTTTGATCCTCG
GGAGTATTCGCCGGCGATGAGCTCCGACTCCAGGGAGCTGCTCGGGATCGATCCCCCGGA
GCTCATCTTCCCCTTCGAATTAAAGAAGCAAATCTCGTGTTCTCTGCACCTAACAAACAA
GACAGATGAATATGTCACGTTTAAGGTTAAGACAACTAGTCCAAAGAAGTACCTGCGTCC
GTCCAAATAAACGGCATTGTGGCTCCCCAGTCCACATCCAA
>Os04g45190        N/A
>Os08g06020        vesicle-associated membrane protein
gaattcggcacgaggCTCTCTCCTCTCCTCTCGTCGTTGCACCCACCTCCTCTCCCAAGA
TCTCACCTTTCCACCCCCTCCGCCCCCGCCTCGATCCGGCAATGGCTTCCCCCGGACCGG
CCACCGCCGGCGAGCTCCTCCGCATCGACCCCGTCGAACTCCGCTTCCCCTTCGAGTTCGA
AGAAGCAGATCTCGTGCTCGATGCAGCTATCGAATCTCAGCGACGACTACATCGCCTTCA
AGGTGAAAACTACAAGCCCAAAGAAGTATTCGGTGCGGCTAACACAGGGGTTGTCTTGC
CACGGTCTACCTGCGATGTCGTTGTGACAATGCAAGCACAGCCGGAAGCCCCGCCTGACA
TGCACTGCAAAGGATAAGTTCCTAGTCCAGAGCGTCATTT
>Os10g36350        N/A
>Os05g30950        vesicle-associated membrane protein
gaattcggcacgaggGCTTCCGCCCGACTCCTGATCAGCAGCCGCTCGCCACCAGCCCACCA
CCAGTGAGTGACGCGACGCGGACCGGCGGCGGCGGCGATGGGCAGCGTGGACTTCGGCCT
CGTCGAGATCCACCCCGGGGAGATCGCCGGTTCGAATTTGAGGTGAAGAAGAAAAGTTCATG
TTATGTCTGTCTAGTGAACAAATCAGAGGAATATGTTGCCATTCAAGGTCAAAAACAACT
TTCTCCAAAAAGGTATTGTGTTCGACCAAATGTAGGAGTCATTCTTCCAAGGCGCAACCA
TGTGTTTATACCAGTTATTA
>Os10g40100        N/A
>Os07g37250        expressed protein
gaattcggcacgaggGCTCTAAACATCAACAAAAGAAGTGTCGACAGGGATCTCGATGTT
TACCGGAACATACTCTCCAAATTGGTCCAGGCTAAGGAACTTCTCAAGGAATACGTGGAA
AGGGAAAAGAAGAAGAGAGAGGAAAGATCAGAGACCCCAAAATCGAATGAAGCTGTTACG
AAATTTGACGGGAGTCTCAATTCCATGAGGCATTAACTCAAGCCTTCATAAAGGGAGTGA
GTGCTACCGTGGATAACACGATTTTAACAGAGTTCTGGATCACGTTGAGAAACTATCAAGG
TCACCCTTTCACTGCACGGACACTTTGGCTGCTCTCGTGTGGTATAGTATTCTGCCCGTAT
>Os06g31100        N/A
>Os06g04640        N/A
>Os06g14080        N/A
```
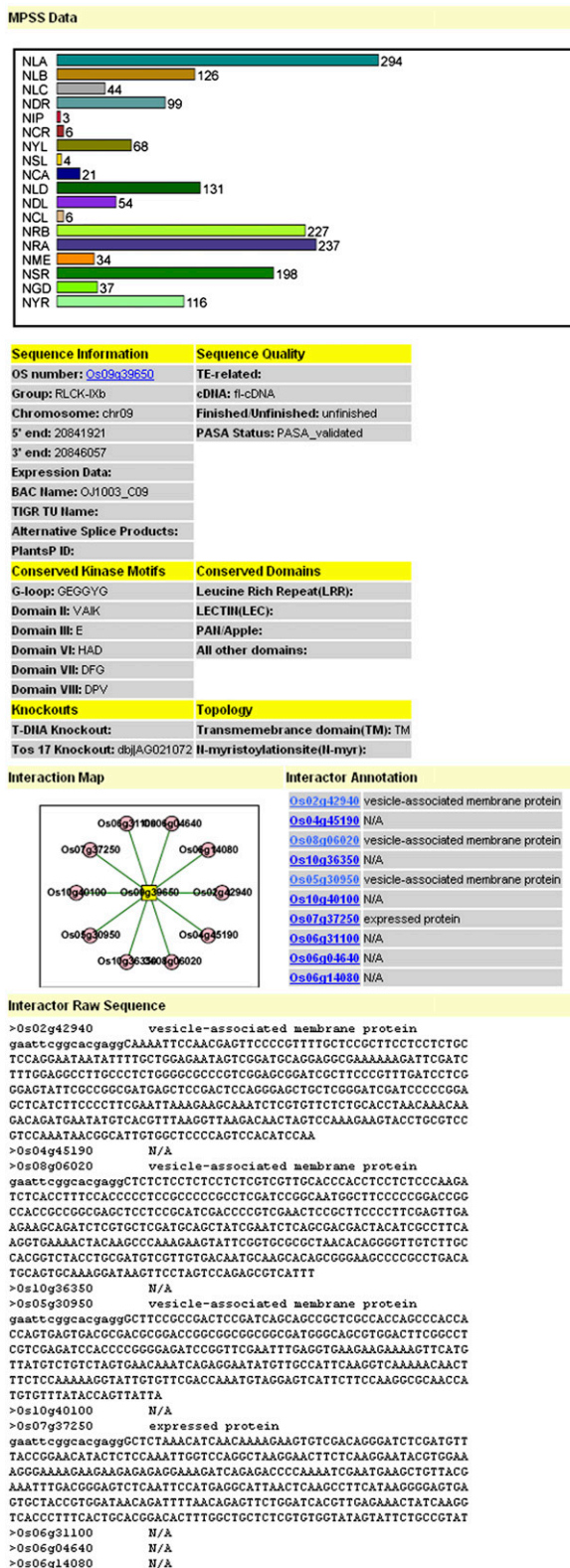
Figure 4. Screen shot of RKD summary page. Links to summary pages are provided from the TIGR model number of each kinase. Summary pages include all data in the RKD as well as interactive Cytoscape interaction maps. Nodes (circles or squares) represent proteins. The color red denotes proteins used as bait in the Y2H or kinases that have been TAP tagged. The color blue denotes proteins identified in the screens. Solid lines represent interactors identified from the Y2H screen, and dashed lines represent interactors identified from TAP-tag copurification. Arrows point from the bait to the target. Scrolling over the node links the proteins to the TIGR rice database.

TAP tagged and transformed into rice. Stable transgenic rice plants expressing the TAP tagged kinases were used to isolate in vivo copurifying proteins. The identities of copurifying proteins for the first 45 TAP tagged kinases were determined using mass spectroscopy (Rohila et al., 2006). Interaction maps were generated using the open source software platform, Cytoscape (http://www.cytoscape.org; Shannon et al., 2003). Cytoscape allows for the visualization and integration of multiple sources of data. Currently, the interaction map displays all kinase baits and identified interactors. A flat file of the interaction data that is compatible with Cytoscape is also provided by the link Raw Data. This text file can be downloaded and used with Cytoscape.

Because Y2H is an indirect screen in a heterologous system, additional evidence is needed to validate the biological relevance of these putative kinase interactors. Protein interactions identified using other methods (i.e. in vitro/in vivo coimmunoprecipitation, or TAP) or corroborative evidence from other biological systems adds validity to the physiological relevance. For example, we found that CK II α-subunit interacts with CK β-subunit (Rohila et al., 2006). This has previously been shown in yeast (Gietz et al., 1995). However, validation of Y2H interactions that do not have corroborating evidence from other systems is difficult. The interaction in yeast may be real but not of biological relevance. Added information is needed to ascertain biological relevance. For example, interacting proteins need to be expressed both temporally and spatially in the same cell types. RKD allows for the integration of MPSS data with rice kinase protein-protein interaction maps to aid in the interpretation of uncharacterized Y2H data. If transcripts from two interacting proteins are not present in the same specified tissue type, or physiological condition, caution should be taken when interpreting the interaction.

Using the RKD

Links to the Tree Viewer, Interactome, and Chromosome Distribution maps are indicated on the home page. In the Tree Viewer page, genomic and functional genomic fields can be selected by checking each box (Fig. 3). Pressing submit displays the selected data adjacent to the tree. Arabidopsis kinases most similar to each rice kinase can also be displayed by selecting nearest Arabidopsis kinase to enable cross species comparisons. The spreadsheet format allows all data or user-defined subsets of data to be readily transferred into any database, such as Excel, for further analysis. Clicking on the gene ID link brings up a summary window showing all of the available data for

that kinase, including an interactive Cytoscape protein-protein interaction map (Fig. 4). Links to the TIGR rice database and PlantsP can be displayed for easy navigation. These links provide simple navigation between all data display formats as well as complementary databases.

Kinases used in the interaction mapping study can be indicated on the phylogenetic tree by selecting the kinase interaction map icon, which provides a link to the interaction map displayed in the summary window (Fig. 4). Alternatively, the complete interaction maps can be viewed by clicking on the Interactome icon listed on the home page. The Interactome page displays the complete Cytoscape interaction map for all kinases. The YTH and TAP tagging studies are currently listed separately but will be integrated upon completion of the study.

Chromosome maps are color coded according to kinase subfamily and kinases are represented as colored squares (Fig. 5). The color key is provided. Mousing over each square generates a pop up showing the ID of each kinase. Clicking on the square navigates the

user back to the Tree Viewer page with the selected kinase highlighted in red.

Phylogenomic Analysis of Kinase Expression

A handful of plant receptor kinases are known to function as specific pathogen recognition receptors sometimes called disease resistance genes (for review, see Nürnberger and Kemmerling, 2006). Because there are hundreds of plant receptor kinases with diverse functions, identifying which ones serve as pathogen recognition receptors remains a difficult challenge. We previously identified two features of plant receptor kinases HRD that are predictive of those that function in pathogen perception (Dardick and Ronald, 2006). These include evidence of recent evolutionary expansion at the subfamily level and lack of a conserved kinase HRD motif (non-RD). A third and as of yet unexplored characteristic is related to gene expression. It is known that pathogen recognition receptors in the unrelated nucleotide-binding site (NBS)-LRR family tend to be expressed at relatively very low levels

Figure 5. Chromosome map of all rice kinases. Image was created using GenomePixelizer (Kozik et al., 2002). Kinases are represented by squares and color coded according to subfamily (complete legend is available at http://rkd.ucdavis.edu).
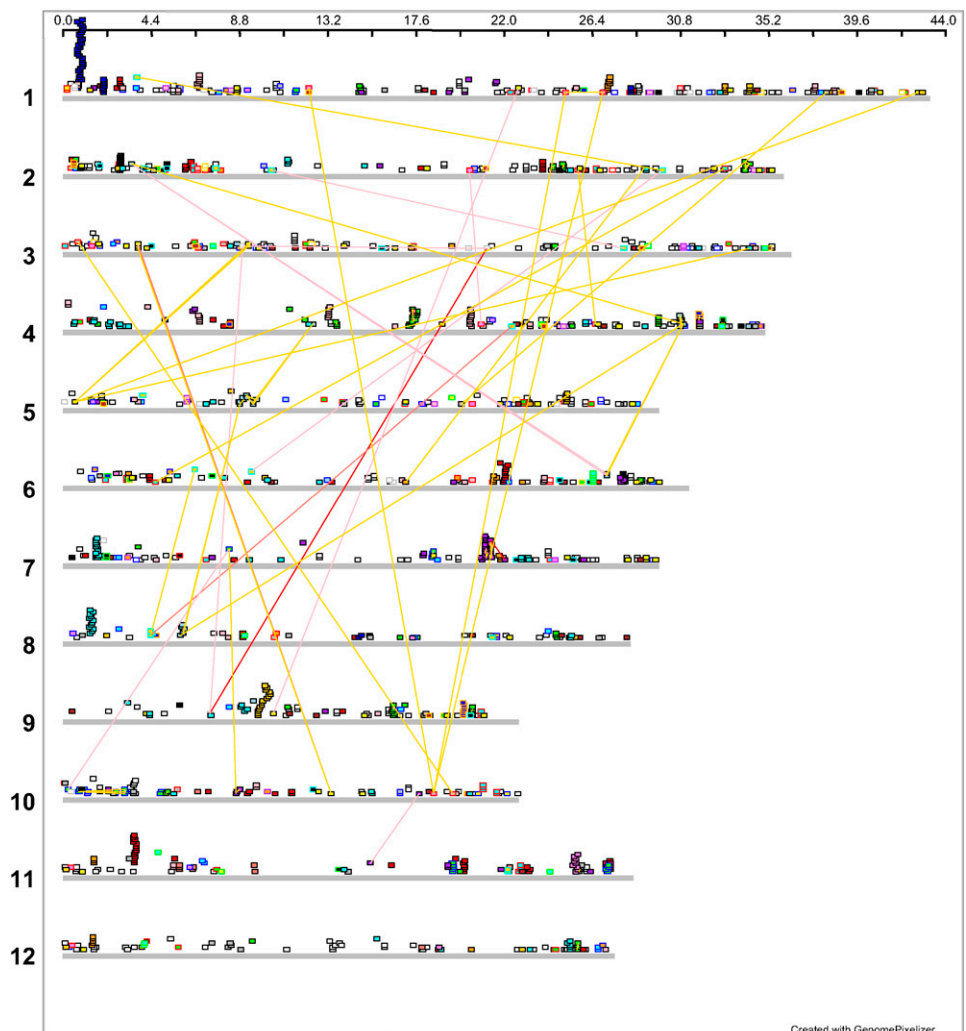
(Meyers et al., 2002). Therefore, we hypothesized that kinase subfamilies functioning as pathogen receptors should also show overall low expression levels compared with kinase subfamilies that function in other cellular processes such as metabolism or development.

To test this hypothesis and, at the same time, the utility of the RKD, we performed a global phylogenomic analysis of kinase expression using all available rice MPSS data (Nakano et al., 2006). Two additional gene families were included in this analysis to provide points of reference: NBS-LRRs and the cytochrome P450 family that is expanded but does not have functions in pathogen detection. NBS-LRRs and cytochrome P450s were identified from the TIGR rice database using existing annotation and were cross referenced to the rice MPSS database. A total of 175 P450s and 164 NBS-LRRs were identified in the MPSS database and used for this analysis. The MPSS dataset represents steady-state expression levels in a large number of rice tissues (mature leaf, young leaf, mature root, young root, stem, immature panicle, meristem, ovary/stigma, pollen, germinating seed [3 d], germinating seedlings [10 d], and callus). Cold and salt stress treatments are also included for leaf and root tissues.

For each kinase, NBS-LRR, or cytochrome P450 gene, the median expression level across all tissues was calculated. The values were normalized and indicated as TPM. Next, the median values for all kinases within each kinase group, family, or subfamily were averaged together. Similarly, median values for NBS-LRRs and P450s were also averaged together, respectively. The averages and SDs were plotted to assess the overall expression levels from each kinase clade (at the group, family, and subfamily levels; Fig. 6).

On the whole, IRAK kinases are expressed at lower levels than other kinase groups with the exception of the IRAK RLCK-VIII subfamily that shows the highest median expression levels of all kinases. IRAK RLCK-VIII includes homologs of tomato (*Lycopersicon esculentum*) Pti1, a known phosphorylation target of the Pto disease resistance gene (Zhou et al., 1995). As expected, the cytochrome P450 family showed significantly higher median expression than NBS-LRRs. Consistent with our hypothesis, non-RD IRAK subfamilies showed the lowest expression levels of all kinases and, in general, have expression levels similar to or lower than NBS-LRRs (Fig. 6). The low steady-state expression levels of these potential pathogen recognition receptors may be related to the need for precise signal regulation, because overexpression of disease resistance genes can lead to cell death and reduced stature (Mindrinos et al., 1994; Oldroyd and
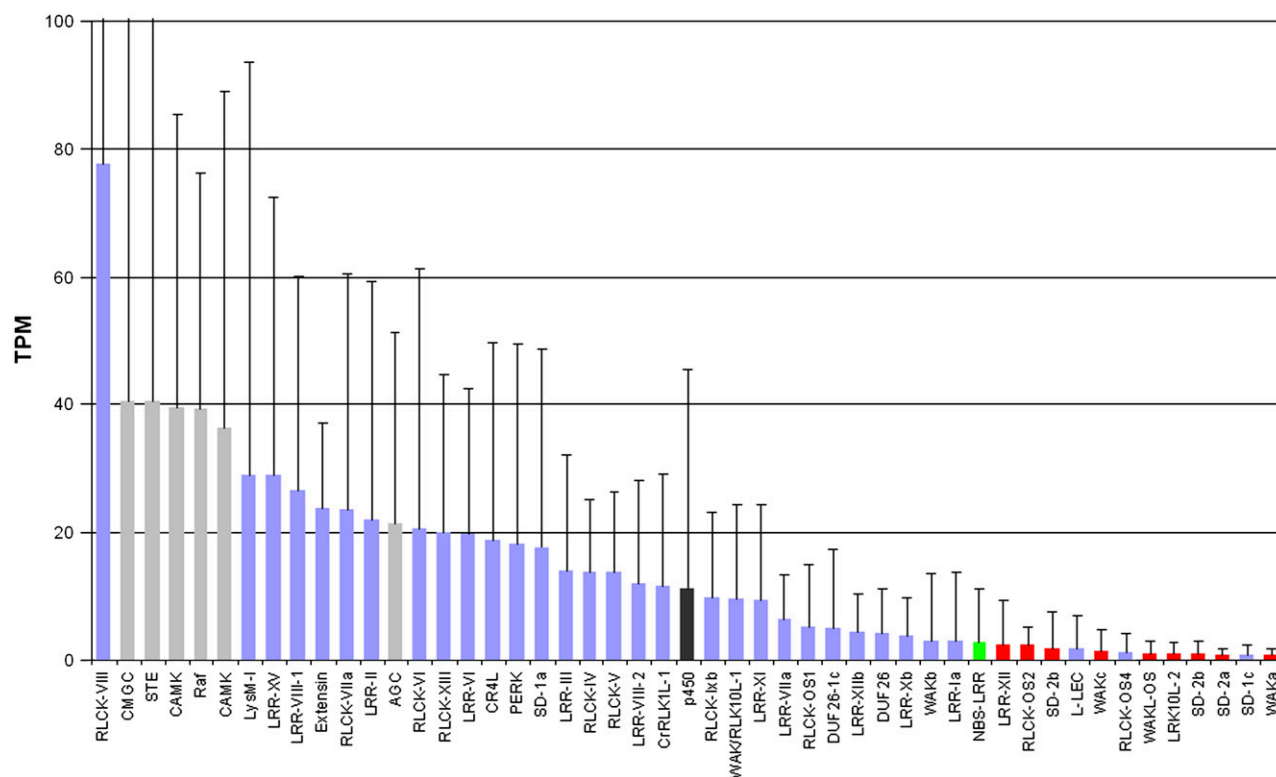

Figure 6. Median gene expression of kinase clades. Expression values along *y* axis represent average TPM of all kinases comprising each clade. Vertical bars represent SD. Clades labeled by kinase group or family are shown on the *x* axis. Group data is shown for AGC, CAMK, CMGC, CK1, and STE kinases (gray). The TKL group was further divided into the Raf kinases (gray) and IRAK subfamilies that contain more than five members (blue). Data for the NBS-LRR and cytochrome P450 families are shown in green and black, respectively. IRAK subfamilies predicted to function in pathogen recognition due to the presence of the non-RD motif are shown in red.

Staskawicz, 1998; Chern et al., 2005). It is important to note that pathogen treatments are not yet available in the rice MPSS database, and it will be interesting to compare future MPSS releases to assess whether or not these genes are up-regulated in response to pathogens.

DISCUSSION AND FUTURE DIRECTIONS

The RKD was created to provide a logical format to analyze diverse sets of genomic information in a phylogenetic context. User-selected genomic and functional genomic fields can be displayed on a phylogenetic tree with links to chromosomal and protein-protein interaction maps. Rather than analyzing 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 (Dardick and Ronald, 2006). This database has also proven useful for the rational selection of kinases to be used in a large-scale proteomic screen (Rohila et al., 2006). The ability to integrate and analyze growing functional genomic data sets in a logical and user-defined fashion will be essential to establishing a more global view of the role kinases play in signaling. We plan to add additional features to RKD including: links to PubMed citations, additional Y2H- and TAP-tagging data, data from new MPSS libraries, microarray expression data, and new T-DNA, transposon, and activation-tagged lines for both kinases and interacting proteins. Results of functional knockout and RNAi knockdown screens currently in progress will also be added. We anticipate this database will provide a useful service to the plant kinase community, and we hope that it will provide a template for the design of new phylogenomic databases.

ACKNOWLEDGMENTS

LITERATURE CITED

Chen X, Shang J, Chen D, Lei C, Zou Y, Zhai W, Liu G, Xu J, Ling Z, Cao G, et al (2006) A B-lectin receptor kinase gene conferring rice blast resistance. Plant J **46:** 794–804

Chern M, Fitzgerald HA, Canlas PE, Navarre DA, Ronald PC (2005) Overexpression of a rice NPR1 homolog leads to constitutive activation of defense response and hypersensitivity to light. Mol Plant Microbe Interact **18:** 511–520

Dardick CD, Ronald PC (2006) Plant and animal pathogen recognition receptors signal through non-RD kinases. PLoS Pathog **2:** e2

Gietz RD, Graham KC, Litchfield DW (1995) Interactions between the subunits of casein kinase II. J Biol Chem **270:** 13017–13021

Kikuchi S, Satoh K, Nagata T, Kawagashira N, Doi K, Kishimoto N, Yazaki J, Ishikawa M, Yamada H, Ooka H, et al (2003) Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice. Science **301:** 376–379

Kozik A, Kochetkova E, Michelmore RW (2002) GenomePixelizer: a visualization program for comparative genomics within and between species. Bioinformatics **18:** 335–336

Krupa A, Abhinandan KR, Srinivasan N (2004) KinG: a database of protein kinases in genomes. Nucleic Acids Res **32:** 153–155

Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S (2002) The protein kinase complement of the human genome. Science **298:** 1912–1934

Meyers BC, Morgante M, Michelmore RW (2002) TIR-X and TIR-NBS proteins: two new families related to disease resistance TIR-NBS-LRR proteins encoded in Arabidopsis and other plant genomes. Plant J **32:** 77–92

Mindrinos M, Katagiri F, Yu GL, Ausubel FM (1994) The *A. thaliana* disease resistance gene *RPS2* encodes a protein containing a nucleotide-binding site and leucine-rich repeats. Cell **78:** 1089–1099

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

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–D735

Nürnberger T, Kemmerling B (2006) Receptor protein kinases: pattern recognition receptors in plant immunity. Trends Plant Sci **11:** 519–522

Oldroyd GED, Staskawicz BJ (1998) Genetically engineered broad-spectrum disease resistance in tomato. Proc Natl Acad Sci USA **95:** 10300–10305

Rohila JS, Chen M, Chen S, Chen J, Cerny R, Dardick C, Canlas P, Xu X, Gribskov M, Kanrar S, et al (2006) Protein-protein interactions of tandem affinity purification-tagged protein kinases in rice. Plant J **46:** 1–13

Sallaud C, Gay C, Larmande P, Bes M, Piffanelli P, Piegu B, Droc G, Regad F, Bourgeois E, Meynard D, et al (2004) High throughput T-DNA insertion mutagenesis in rice: a first step towards in silico reverse genetics. Plant J **39:** 450–464

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res **13:** 2498–2504

Shiu S, Bleecker AB (2001) Receptor-like kinases from Arabidopsis form a monophyletic gene family related to animal receptor kinases. Proc Natl Acad Sci USA **98:** 10763–10768

Shiu SH, Karlowski WM, Pan R, Tzeng YH, Mayer KF, Li WH (2004) Comparative analysis of the receptor-like kinase family in Arabidopsis and rice. Plant Cell **16:** 1220–1234

Shpak ED, Berthiaume CT, Hill EJ, Torii KU (2004) Synergistic interaction of three ERECTA-family receptor-like kinases controls Arabidopsis organ growth and flower development by promoting cell proliferation. Development **131:** 1491–1501

Song WY, Wang GL, Chen LL, Kim HS, Pi LY, Holsten T, Gardner J, Wang B, Zhai WX, Zhu LH, et al (1995) A receptor kinase-like protein encoded by the rice disease resistance gene, Xa21. Science **270:** 1804–1806

Sun X, Cao Y, Yang Z, Xu C, Li X, Wang S, Zhang Q (2004) Xa26, a gene conferring resistance to Xanthomonas oryzae pv. oryzae in rice, encodes an LRR receptor kinase-like protein. Plant J **37:** 517–527

Suzaki T, Sato M, Ashikari M, Miyoshi M, Nagato Y, Hirano HY (2004) The gene FLORAL ORGAN NUMBER1 regulates floral meristem size in rice and encodes a leucine-rich repeat receptor kinase orthologous to Arabidopsis CLAVATA1. Development **131:** 5649–5657

Tchieu JH, Fana F, Fink JL, Harper J, Nair TM, Niedner RH, Smith DW, Steube K, Tam TM, Veretnik S, et al (2003) The PlantsP and PlantsT functional genomics databases. Nucleic Acids Res **31:** 342–344

Zhou J, Loh YT, Bressan RA, Martin GB (1995) The tomato gene Pti1 encodes a serine/threonine kinase that is phosphorylated by Pto and is involved in the hypersensitive response. Cell **83:** 925–935