Title

The Rice Kinase Phylogenomics Database: a guide for systematic analysis of the rice kinase super-family

Journal

Trends in Plant Science, 15(11)

ISSN

13601385

Authors

Jung, Ki-Hong
Cao, Peijian
Seo, Young-Su
et al.

Publication Date

2010-11-01

DOI

10.1016/j.tplants.2010.08.004

Peer reviewed

Techniques & Applications

The Rice Kinase Phylogenomics Database: a guide for systematic analysis of the rice kinase super-family

Ki-Hong Jung[1,2,3], Peijian Cao[1,2], Young-Su Seo[1], Christopher Dardick[4] and Pamela C. Ronald[1,2,3]

[1] Department of Plant Pathology, University of California, Davis, CA 95616, USA
[2] The Joint Bioenergy Institute, Emeryville, CA 94710, USA
[3] Department of Plant Molecular Systems Biotechnology & Crop Biotech Institute, Kyung Hee University, Yongin, 446-701, Korea
[4] Appalachian Fruit Research Station, USDA-ARS, Kearneysville, WV 25430, USA

Determination of gene function is particularly problematic when studying large-gene families because redundancy limits the ability to assess the contributions of individual genes experimentally. Phylogenomics is a phylogenetic approach used in comparative genomics to predict the biological functions of members of large gene-families by assessing the similarity among gene products. In this report, we describe the application of the Rice for elucidating functions of individual members of this gene family.

Biological information raises the usage of phylogenomics

Despite the importance of rice (*Oryza sativa*) and its emergence as a model species for genetic studies, assignment of gene function has progressed slowly [1–3]. A major obstacle is functional redundancy, which is due to gene duplication and therefore metabolic redundancy [4,5]. Approximately 50% of all non-transposable element-related rice genes are potentially subject to functional redundancy [6,7]. For example, the Rice Genome Annotation Project (RGAP, http://rice.plantbiology.msu.edu/) recently identified 3842 rice paralogous gene families comprising 20 729 protein sequences [6,7].

To help to address these obstacles we recently developed two phylogenomics databases for 1508 rice kinases (Rice Kinase Database, RKD; http://rkd.ucdavis.edu) [8] and 769 glycosyltransferases (Rice Glycosyltransferase Database, RGTD; http://phylomics.ucdavis.edu/cellwalls/gt/) [9]. These databases present diverse data in a phylogenetic context, including gene annotations, orthologous gene predictions information about gene indexed mutants as well as transcriptome data from ESTs, massively parallel signature sequencing (MPSS), and microarray analyses. Table S1 describes other databases that are useful for phylogenomic approaches in plant species (see supplementary material online). Information in these databases is mostly dependent on various analyses with predicted protein sequences and therefore biological information for the databases is still limited.

The RKD integrates protein–protein interaction data based on high-throughput yeast two hybrid (Y2H) and *in vivo* tandem affinity purification (TAP) pull-down screens. In this report, we describe how the RKD can be used to elucidate the functions of individual members of the large rice kinase gene-family.

Application of phylogenomics databases

Here we describe two applications of the RKD: (i) identification of mitogen-associated protein kinase (MAPK-MAPKK-MAPKKK) signaling cascades that are coexpressed in response to a broad range of stress responses, and (ii) prediction of five functional interactions of a light-inducible kinase by integrating gene expression patterns with the protein–protein interaction map.

Identification of eight MAPK genes that are predicted to function in the same signaling cascade

A MAPK signaling cascade consists of a MAP kinase, MAP kinase kinase (MKK or MAP2K), and MAP kinase kinase kinase (MKKK or MAP3K). In general, the MAP3K phosphorylates a serine or threonine residue on a MAP2K, which sequentially activates a MAPK, the last protein in the cascade. The activated MAPK leads to the phosphorylation of downstream transcription factors that regulate various responses such as stress signaling, pathogen response, and hormone signaling [10]. The function of most MAPKs (MAP3K-MAP2K-MAPK) in rice is not known, with the exception of *MCK1*, *OsMSRMK2*, *OsMAPK4*, *OsMAPK5*, *OsMAPK6*, *OsBWMK1* and *OsWJUMK1* [10–13].

To identify the functions of the unknown MAPKs, we integrated diverse gene expression data from various developmental organs or tissues. We also included the following datasets, which compare cytokinin-treated leaves with dimethylsulfoxide-treated leaves and abiotic stress-treated leaves with untreated leaves and pathogen-treated leaves with mock-treated leaves (Figure 1a) [7].

On the basis of phylogenetic analysis, MAPKs are classified into four subgroups (Figure 1). Clustering analysis with developmental expression patterns of all rice MAPKs identified a set of MAPK components that appear to function in the same signaling cascade as the one shown in Figure 1a. Further support for this putative signaling cascade comes from protein–protein interaction data indicating that MAPK *Os03g17700* and MAP2K *Os02g54600* physically interact [14]. Therefore, we predict that MAP2K Os02g54600 is located upstream of the MAPK Os03g17700 and regulates various downstream stress responses.

Corresponding author: Ronald, P.C. (pcronald@ucdavis.edu).

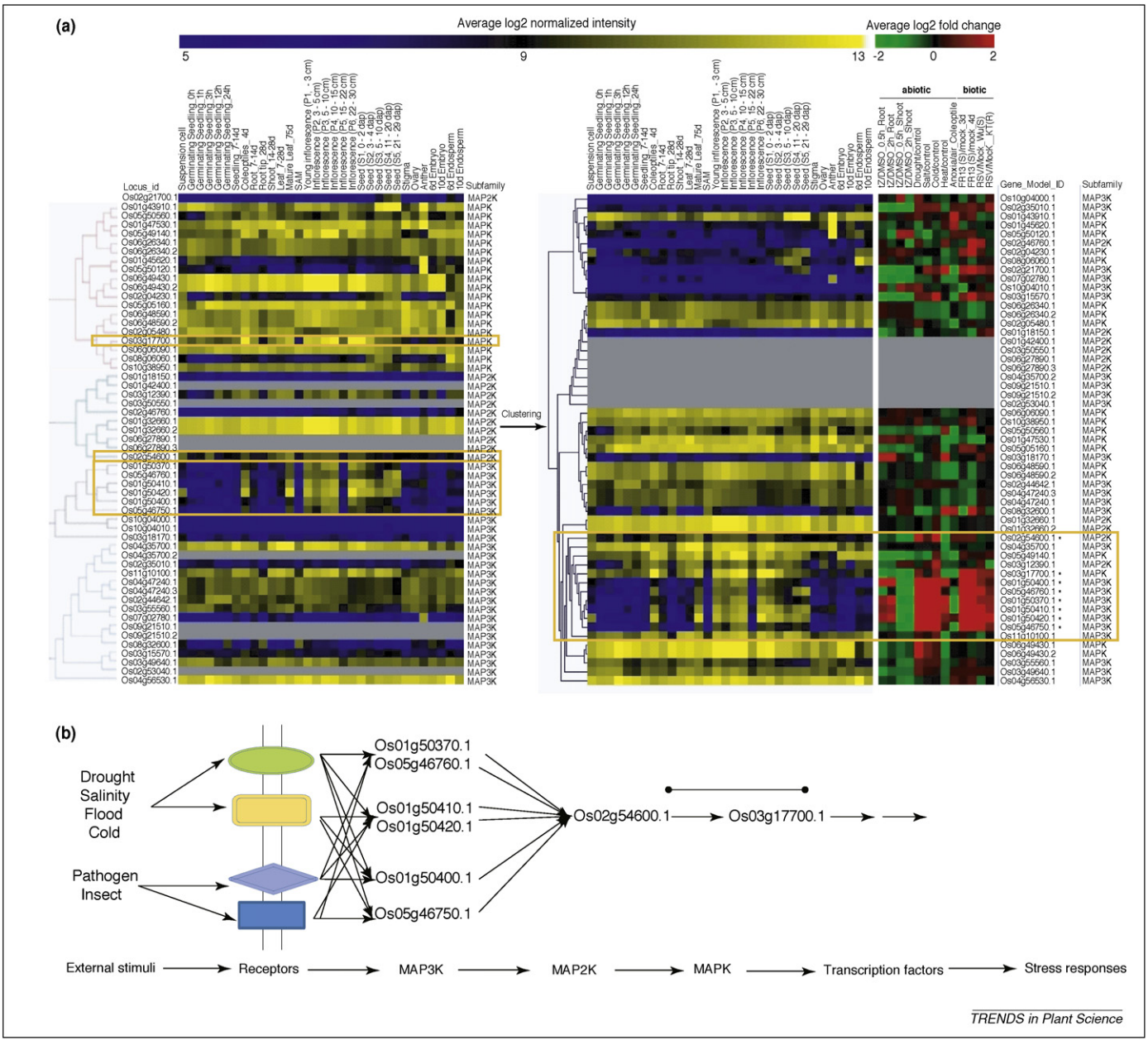

Figure 1. Identification of eight MAPK genes predicted to function in the same signaling cascade. **(a)** The left panel shows phylogenomics data for each MAPK, MAP2K, and MAP3K subfamilies and the right panel shows hierarchical clustering (HCL) analysis of normalized Affymetrix gene expression data for all subfamilies in 32 types of anatomically- or development-related tissues or organs. Corresponding Affymetrix differential expression data following exposure to various biotic or abiotic stresses is also shown. Asterisks (*) indicate eight genes grouped by HCL analysis, using normalized expression levels, including one MAPK, one MA2PK, and six MA3PKs. These genes are similarly regulated under various biotic or abiotic stresses. **(b)** A proposed regulatory model for the six MAPKs, MAP2Ks and MAP3Ks identified in **(a)**. Biotic or abiotic stresses are perceived by receptors represented by four color-filled boxes. Stress perception is transferred from the receptors to downstream responses via a MAPKs cascade. Lines with two dark-filled circles indicate the protein–protein interactions confirmed by yeast two hybrid screening assays.

Similarly, six *MAP3K*s (*Os01g50370*, *Os05g46760*, *Os01g50400*, *Os01g50410*, *Os01g50420* and *Os05g46750*) are predicted to be upstream of the *MAPK* and *MAP2K* genes in examined developmental stages and in response to various stress challenges. Notably, the six *MAP3K* genes show similar expression patterns in diverse developmental organs or tissues: suspension cells, germinating seedlings, leaves or shoots, roots, coleoptiles, flag leaves, developing inflorescences, and developing seeds (Figure 1a). The observed differential expression patterns in response to biotic or abiotic stresses are also similar: drought, salt, cold, fungal and viral pathogens induce the expression of these genes, whereas cytokinin (*trans-Zeatin*, *tZ*) treatment of leaf tissue represses transcript levels (Figure 1a).

These data suggest that these six *MAP3K* genes play a key role in a broad range of stress responses, including responses to biotic/abiotic stresses. Rice might have evolved an enhanced response to these stresses by retaining multiple gene family members that perform critical functions. The application of phylogenomics data to all MAPK subfamilies effectively revealed components in the same signaling pathway (Figure 1b). A parallel approach can be used to predict other gene sets that are co-expressed and therefore predicted to be commonly regulated.

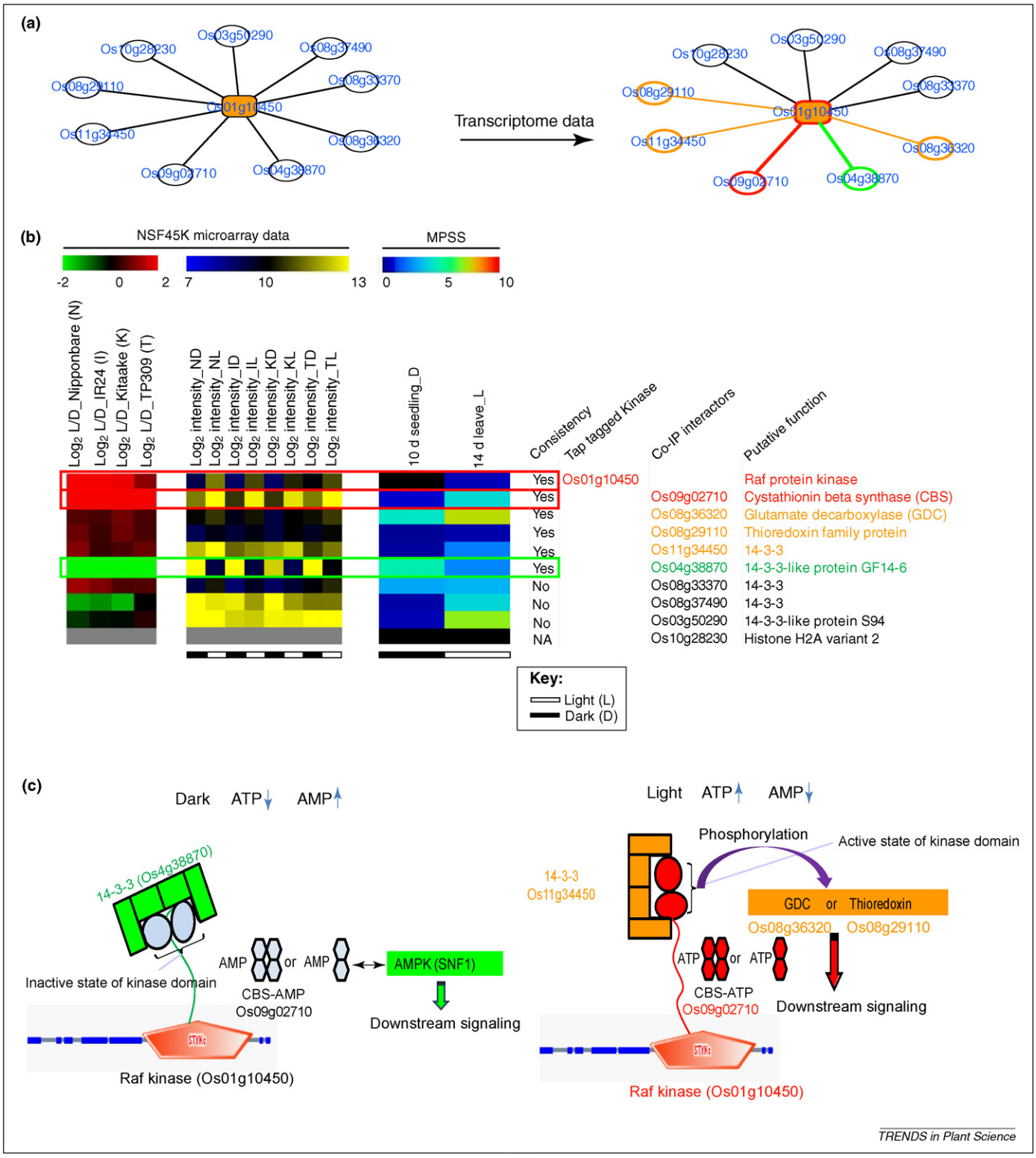

Figure 2. Identification of biologically functional interacting regulators of a target kinase by using integrated protein–protein interaction maps and transcriptomics data. **(a)** The process of integrating transcriptome data onto the interaction map of a TAP-tagged kinase (Os01g10450) to generate a functional model. The interactomes on the left and the right represent models before and after integration of gene expression data into the interaction map of Os01g10450, respectively. **(b)** Microarray and MPSS data derived from comparisons of light- and dark-grown seedlings were integrated into the interaction map of a TAP-tagged kinase (Os01g10450). Light- and dark-grown seedlings of four rice varieties: Nipponbare (N), IR24 (I), Kitaake (K) and Taipei 309 (T). Log_2 ratios of light (L) over dark (D) were generated for the four varieties (left) [5,7]. Log_2-transformed normalized NSF45K intensity under light and dark conditions and the log_2-transformed MPSS signature under light and dark conditions are shown (middle). **(c)** Model for the function of Os01g10450, encoding a Raf kinase, in response to light or dark. In the dark, binding of the regulatory adapter protein 14-3-3 with the Raf kinase stabilizes the basal inactive Raf-1 conformation. Because the AMP:ATP ratio is high at night, excess AMP might preferentially bind an AMP-associated kinase (AMPK, e.g. SNF1) instead of the CBS domains in Os09g02710. During the day, CBS-bound ATP likely forms and accumulates during photosynthesis because tandem pairs of CBS domains act as sensors of cellular energy status. Thereafter, binding of CBS-bound ATP (Os09g02710) to a Raf kinase might contribute to maintaining the active state of the kinase. In turn, the activated Raf kinase is supposed to phosphorylate components in the downstream signaling pathway. Red letters, lines, circles, boxes, rectangles and arrows indicate upregulation in the light or downregulation in the dark; light yellow symbols indicate slight upregulation in the light or downregulation in the dark; and green symbols indicate downregulation in the light or upregulation in the dark. Blue and yellow bars indicate the intensity of gene expression based on microarray data. Black and red bars indicate the frequency of MPSS signatures. TAP-tagged kinase and nine co-immunoprecipitated interactors are represented as TIGR locus identifiers (Locus Ids). IP, immunoprecipitated.

Prediction of biological function by integrating protein–protein interaction maps with phylogenomics data

We previously identified interacting proteins for 116 kinases using the Y2H system and identified co-immunoprecipitated proteins for 83 TAP-tagged kinases using mass spectroscopy [14,15].

The biological relevance of these putative kinase interactors can be further enhanced by integrating transcriptomics data. We identified nine interactors (http://rkd.ucdavis.edu/TAPbaitprey.php?id=Os01g10450) that co-immunoprecipitated with one of the TAP-tagged kinases (Os01g10450, Raf kinase). To these data, we incorporated gene expression data derived from the NSF45K light *versus* dark dataset for each of the Os01g10450-associated interactors (Figure 2a, b).

The combined analysis allowed us to identify interaction partner(s) co-regulated with the Raf kinase in response to light. For example, *Os09g02710*, which encodes a cystathionine beta-synthase (CBS), shows expression patterns that are most similar to that of *Os01g10450*. *Os08g36320* encoding glutamate decarboxylase (GDC), *Os08g29110* encoding a thioredoxin family protein, and *Os11g34450* encoding a 14-3-3 protein are positively associated with *Os01g10450*. We also found that *Os04g38870*, encoding a 14-3-3 protein, is significantly downregulated in the light. These results suggest that five out of nine interactors are positioned in the light-responsive regulatory pathway positively linked to *Os01g10450*. Expression patterns of the five genes are similar in both microarray (NSF45K light *versus* dark array) and MPSS analyses of 10-d-old seedlings grown in the dark (10 d seedling_D) and 14-day-old seedlings grown in light (14 d leaf_L) (Figure 2b) [5].

Based on these results we present a model for Raf kinase function (Figure 2C). In the dark, binding of the regulatory adapter protein 14-3-3 with the Raf kinase stabilizes the basal inactive Raf-1 conformation, as previously shown for yeast Raf-1 [16]. Because the AMP:ATP ratio is high at night, excess AMP might preferentially bind an AMP-associated kinase (AMPK, e.g. SNF1), which accumulates during times of energy deficiency. Even if the CBS domains from both AMPK and Os09g02710 can interact with both ATP and AMP, the CBS domain of AMPK preferentially binds to AMP because the *AMPK* gene is much more stimulated at night than *Os09g02710*. During the day, CBS-bound ATP probably forms and accumulates by photosynthesis because tandem pairs of CBS domains act as sensors of cellular energy status [17]. Thereafter, binding of CBS-bound ATP (Os09g02710) to a Raf kinase might contribute to maintaining the active state of the kinase. In yeast, a light-induced 14-3-3 protein also functions as cofactor to stabilize an activated Raf kinase by binding to a phosphorylated serine residue on the kinase [18]. Similarly, a light-stimulated 14-3-3 protein (Os11g34450) can interact with the Raf kinase (Os01g10450) and play a role in maintaining the active state of the kinase (Figure 2b), thereby enabling this active enzyme to phosphorylate light-stimulated GDC or thioredoxin and turn on related downstream signaling pathways (Figure 2c). Roles of different 14-3-3 proteins in activation or inactivation of Raf kinase were also reported in a previous study of a Raf kinase in yeast [18]. These predictions can be further validated with available genetic materials (knockout or RNAi or activation tagging lines) to study the functions of these genes.

Perspectives

The RKD provides a practical framework for exploring diverse and high-throughput expression data in a phylogenetic context. Recently, we used similar methods to dissect functions of 12 AP2 or ERFs transcription factors identified from genome-wide transcriptome analysis to compare the submergence tolerance line M202(Sub1) with the intolerant control M202 [19]. Because rice is a valuable model for candidate bioenergy grass crops, integration of orthologous gene information from other recently sequenced grasses such as sorghum (*Sorghum bicolor*), *Brachypodium*, maize (*Zea mays*), foxtail millet (*Setaria italica*), and switch grass (*Panicum virgatum*) into the rice RKD should facilitate the prediction of gene function in these species [8,9].

Acknowledgments

We thank Dr. Blake Meyers and Dr. Kan Nobuta for providing MPSS signatures data, Dr. Gynheung An for providing the information about T-DNA insertional lines developed in Kyung Hee University, Korea and Dr. Harkamal Walia for useful discussions. This work was supported by a Department of Energy grant to the Joint Bioenergy Institute (Grant no. DE-AC02-05CH11231) and a National Science Foundation Plant Genome Research Program (Grant no. DBI0313887) to PCR, and a Young Scientist Program through the National Research Foundation of Korea (Grant no. 2010-0981) to KHJ.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.tplants.2010.08.004.

References

1 Shimamoto, K. and Kyozuka, J. (2002) Rice as a model for comparative genomics of plants. *Annu. Rev. Plant Biol.* 53, 399–419
2 Jung, K.H. *et al.* (2008) Towards a better bowl of rice: assigning function to tens of thousands of rice genes. *Nat. Rev. Genet.* 9, 91–101
3 IRGSP (2005) The map-based sequence of the rice genome. *Nature* 436, 793–800
4 Pasek, S. *et al.* (2006) The role of domain redundancy in genetic robustness against null mutations. *J. Mol. Biol.* 362, 184–191
5 Jung, K.H. *et al.* (2008) Identification and functional analysis of light-responsive unique genes and gene family members in rice. *PLoS Genet.* 4, e1000164
6 Lin, H. *et al.* (2008) Characterization of paralogous protein families in rice. *BMC Plant Biol.* 8, 18
7 Jung, K.H. *et al.* (2008) Refinement of light-responsive transcript lists using rice oligonucleotide arrays: evaluation of gene-redundancy. *PLoS ONE* 3, e3337
8 Dardick, C. *et al.* (2007) The rice kinase database. A phylogenomic database for the rice kinome. *Plant Physiol.* 143, 579–586
9 Cao, P.J. *et al.* (2008) Construction of a rice glycosyltransferase phylogenomic database and identification of rice-diverged glycosyltransferases. *Mol. Plant* 1, 858–877
10 Reyna, N.S. and Yang, Y. (2006) Molecular analysis of the rice MAP kinase gene family in relation to *Magnaporthe grisea* infection. *Mol. Plant Microbe Interact.* 19, 530–540
11 Koo, S.C. *et al.* (2009) OsBWMK1 mediates SA-dependent defense responses by activating the transcription factor OsWRKY33. *Biochem. Biophys. Res. Commun.* 387, 365–370
12 Agrawal, G.K. *et al.* (2003) Novel rice MAP kinases OsMSRMK3 and OsWJUMK1 involved in encountering diverse environmental stresses and developmental regulation. *Biochem. Biophys. Res. Commun.* 300, 775–783

13 Fu, S.F. *et al.* (2002) Transcriptional regulation of a rice mitogen-activated protein kinase gene, OsMAPK4, in response to environmental stresses. *Plant Cell Physiol.* 43, 958–963
14 Ding, X. *et al.* (2009) A rice kinase-protein interaction map. *Plant Physiol.* 149, 1478–1492
15 Rohila, J.S. *et al.* (2006) Protein–protein interactions of tandem affinity purification-tagged protein kinases in rice. *Plant J.* 46, 1–13
16 Balan, V. *et al.* (2006) Identification of novel in vivo Raf-1 phosphorylation sites mediating positive feedback Raf-1 regulation by extracellular signal-regulated kinase. *Mol. Biol. Cell* 17, 1141–1153
17 Scott, J.W. *et al.* (2004) CBS domains form energy-sensing modules whose binding of adenosine ligands is disrupted by disease mutations. *J. Clin. Invest.* 113, 274–284
18 Tzivion, G. *et al.* (1998) A dimeric 14-3-3 protein is an essential cofactor for Raf kinase activity. *Nature* 394, 88–92
19 Jung, K.H. *et al.* (2010) The submergence tolerance regulator Sub1A mediates stress-responsive expression of AP2/ERF transcription factors. *Plant Physiol.* 152, 1674–1692

1360-1385/$ – see front matter. Published by Elsevier Ltd.
doi:10.1016/j.tplants.2010.08.004 Trends in Plant Science, November 2010, Vol. 15, No. 11